



St. Xavier's University, Kolkata

Xavier's Society for Science & Innovation

XENESIS

**Decoding Reality: How Statistics
and Algorithms Shape The
Modern World**

Vol. 1

March 2026

DEPARTMENTAL PHOTOGRAPHS



**M.Sc. Computer Science and Statistics
Batch 2024-2026 & 2025-2027**



**B.Sc. Statistics and Data Science
Batch 2025-2029**

**St. Xavier's University,
Kolkata**

**Xavier's Society for Science & Innovation
(XSSI)**

PRESENTS

XENESIS

**Decoding Reality: How Statistics and
Algorithms Shape the Modern World**

VOL. 1
March 2026



PREFACE

It is with great pride that Xavier's Society for Science and Innovation (XSSI) presents the inaugural edition of Xenesis'26, the e-magazine of the Faculty of Science at St. Xavier's University, Kolkata. More than a publication, this magazine marks the beginning of a shared intellectual journey shaped by inquiry, innovation and excellence.

Established in August 2022, the Faculty of Science, comprising the Departments of Statistics and Computer Science, has emerged as a dynamic centre of learning committed to academic distinction, research engagement and ethical responsibility. Rooted in the Xavierian and Ignatian tradition, the Faculty of Science strives to foster analytical rigor, technological competence and the spirit of Magis, empowering students to become innovative and responsible global citizens.

A vibrant extension of this vision is the Xavier's Society for Science and Innovation (XSSI), which brings together students across emerging domains such as Artificial Intelligence, Data Science, Cybersecurity, Business Analytics and Biostatistics. Through hackathons, research initiatives and knowledge-sharing platforms, XSSI bridges theory with practice while nurturing creativity, leadership and integrity.

Xenesis'26 stands as a reflection of this collective commitment: to blend knowledge, technology and human values, and to push intellectual boundaries in the true spirit of Nihil Ultra. May this inaugural edition inspire curiosity, collaboration and continued excellence in the years ahead.



MESSAGE FROM THE VICE-CHANCELLOR

It is a matter of great pride and pleasure to present the inaugural edition of the magazine published by the Xavier's Society for Science and Innovation (XSSI), an initiative of the students of the Faculty of Science. Such academic endeavors reflect the vibrant intellectual environment that our university seeks to nurture.

I am pleased to note that this magazine brings together a diverse range of contributions, including innovative scientific articles, discussions on emerging research themes, insightful interviews with our distinguished senior professors, as well as creative expressions such as poems and artwork. This blend of scientific inquiry and creative thinking reflects the holistic approach to learning that we strongly encourage at our university.

I commend the Dean of the Faculty of Science, the dedicated teachers, and the enthusiastic students for their collective efforts in bringing out this publication. Their guidance, mentorship, and commitment to academic excellence have played a vital role in encouraging students to explore new ideas, engage in meaningful research, and express their creativity.

I congratulate the Xavier's Society for Science and Innovation (XSSI) and the editorial team for this commendable initiative. I am confident that this magazine will inspire readers to think critically, innovate boldly, and continue contributing to the advancement of knowledge.

I extend my best wishes to the Dean, teachers, students and the entire XSSI team for their continued success in fostering a culture of curiosity, research and innovation.

Warm Regards,
Rev. Dr. J. Felix Raj, SJ.
Vice-Chancellor
St. Xavier's University, Kolkata



MESSAGE FROM THE DEAN OF SCIENCE

It gives me immense pleasure to present this edition of the magazine published by the Xavier's Society for Science and Innovation (XSSI), an initiative of the students of the Faculty of Science. This magazine stands as a reflection of the intellectual curiosity, creativity, and scientific spirit that our students continually demonstrate.

The contributions featured in the inaugural issue, ranging from innovative scientific articles and discussions on contemporary research topics to insightful interviews with our esteemed senior professors, showcase the vibrant academic culture that thrives within our university. Equally commendable is the inclusion of poems, artwork and other creative expressions, which remind us that science and creativity are deeply interconnected and together enrich the learning experience.

I congratulate the editorial team and all contributors for their dedication, enthusiasm and hard work in bringing out this magazine. Initiatives such as this not only promote scientific thinking and research awareness but also encourage students to express their ideas, perspectives, and talents beyond the classroom.

I hope this magazine inspires readers to explore new ideas, question established knowledge and continue their journey of discovery and innovation. My best wishes to the Xavier's Society for Science and Innovation (XSSI) for their continued success in nurturing a culture of curiosity, creativity and scientific excellence.

Warm Regards,
Dr. Tuhin Utsab Paul
Dean of Science
St. Xavier's University, Kolkata

MESSAGE FROM THE EDITORIAL BOARD

Dear Readers,

It is both an honour and a privilege to welcome you to the inaugural edition of Xenesis'26, the E-magazine of the Faculty of Science at St. Xavier's University, Kolkata. The theme of this edition, "Decoding Reality: How Statistics and Algorithms Shape the Modern World," invites us to pause and reflect on the invisible frameworks that influence our everyday lives. From predicting climate patterns and managing public health data to powering recommendation systems and securing digital transactions, statistics and algorithms silently structure the world we inhabit.

At the intersection of the Department of Statistics and the Department of Computer Science lies a powerful synergy: one that blends mathematical reasoning with computational intelligence. The articles featured in this edition reflect that synergy, highlighting both technical depth and thoughtful reflection on the ethical responsibilities that accompany technological advancement.

This inaugural issue is a testament to the intellectual vitality of our faculty and the collaborative spirit fostered by Xavier's academic tradition. We are deeply grateful to have received insightful contributions from Dr. Swapan Kumar Pradhan and Dr. Aniket Biswas. Their perspectives greatly enrich this inaugural issue and connect our student community with global and national centres of academic and professional excellence. We extend our heartfelt gratitude to our contributors, Xenesis board members, faculty mentors and the Xavier's Society for Science and Innovation (XSSI) for their dedication and vision in bringing this magazine to life.

As you turn these pages, we hope you are inspired to not only to understand how reality is decoded through numbers and algorithms, but also to reflect on how we, as emerging professionals, can shape that reality with wisdom and integrity.

Warm Regards,
Editorial Board
Xenesis 2026
St. Xavier's University, Kolkata

XENESIS BOARD MEMBERS

EDITORIAL BOARD MEMBERS

EDITOR-IN-CHIEF

Ms. Aditi Shah

ASSOCIATE EDITORS

Mr. Siddhartha Das

Ms. Swastika Hui

Ms. Kritee Mukherjee

Ms. Smriti Chauhan

CREATIVE EDITOR

Ms. Ushosee Mukherjee

DESIGN COMMITTEE MEMBERS

DESIGN HEAD

Mr. Akashdeep Das

ASSOCIATE DESIGNERS

Mr. Rupak Bera

Mr. Anirban Nath

Mr. Dipanjan Ganguly

Ms. Jigesaa Chakroborty

ADVISORY COMMITTEE

Dr. Tuhin Utsab Paul

Dr. Tanushree Biswas

Dr. Soma Debnath

Dr. Utpal Pore

Rev. Dr. Maria Joseph Israel, SJ.

TABLE OF CONTENTS

Preface	2
Message from the Vice-Chancellor	3
Message from the Dean of Science	4
Message from the Editorial Board	5
Xenesis Board Members	6
Luminary Reflections	8
Alumni Submissions	44
Student Submissions	72
Creative Corner(Poetry/Stories)	101
Innovation & Case Study	118
The Veteran Voices	136
Fun Zone	147
Art & Photography	152
Gallery	159



LUMINARY REFLECTIONS

“The greatest leaders are those who empower others with knowledge.”

— Bill Gates

The Luminary Reflections section brings together the insights and perspectives of distinguished industry experts and professors who illuminate important ideas from their fields. Through their articles and reflections, readers gain valuable knowledge shaped by years of experience, research and professional practice. This section offers a thoughtful space where expertise meets inspiration, guiding readers toward deeper understanding and broader perspectives.

“Where experience and insight illuminate new ideas.”

Data Chronicles: Bridging Truth, Transparency and Technology

Dr. Swapan Kumar Pradhan¹

Principal Statistical Analyst, Bank for International Settlements, Switzerland

Deciphering reality in today's world begins with recognising the profound role that statistics and algorithms play in shaping our lives. In an era dominated by data, the ability to collect, analyse and interpret information has become indispensable for decision-making across all facets of society. Drawing on three decades of experience with national and international statistical systems, it is evident that statistics hold transformative power, offering opportunities to address global challenges while raising important ethical and practical considerations. As we navigate this data-driven age, the importance of transparency, consistency and ethical responsibility cannot be overstated.

Statistics have long been a cornerstone of governance, transparency and informed decision-making. **National and international authorities** collect vast amounts of data that underpin economic policy, public accountability and social progress. Indicators such as inflation, employment and GDP are vital to assessing economic health and implementing policies that address societal needs. Similarly, financial data, including aggregated balance sheets and granular data from banks and non-bank financial institutions, is critical for understanding the stability of financial systems and identifying risks. Aggregated data provides a high-level overview of the financial sector, such as total assets, liabilities, or capital adequacy ratios, which are essential for monitoring systemic risks. Granular data, on the other hand, offers deeper insights into the composition of loans, non-performing assets, or sectoral exposures, helping to assess the health of individual institutions and sectors. Transparency in disseminating this financial data fosters trust in governance, empowers the public to hold institutions accountable and informs decisions in areas ranging from infrastructure planning to financial market stability.

Technological advancements have significantly **enhanced the scope and efficiency of data collection and dissemination**. Automated systems are now widely used in advanced economies to capture demographic, economic and market data, while emerging markets are transitioning from manual to digital methods.² Dissemination practices have evolved from physical publications to digital formats, including Excel and interactive query tools, making data more accessible to policymakers, researchers and the public. While these innovations have democratised access to information and enabled a more informed society, they have also highlighted critical gaps. The methodologies behind data collection often remain opaque,

¹ Former Assistant Adviser, RBI, Statistics Department, Mumbai (1995-2003). This article reflects personal insights, with AI/GPT used to refine clarity. The views expressed in this article are those of the author and do not necessarily reflect the official position of the BIS. (https://www.bis.org/author/swapan-kumar_pradhan.htm); Email: swapan-kumar.pradhan@bis.org .

² For example, Special Intensive Revision (SIR) exercise in India to update electoral rolls.

raising questions about the reliability and integrity of the data produced. Without transparency in these processes, the full potential of data-driven decision-making cannot be realised.

AI and machine learning have further revolutionised the field of statistics, enabling algorithms to adapt and improve as they process more data ([Serra and Svezia, 2022](#)). These technologies have unlocked unprecedented possibilities, from forecasting economic trends to personalising consumer experiences. However, they also bring significant risks. Concerns about surveillance and privacy have grown as data-driven technologies increasingly infiltrate everyday life. Algorithms, while powerful, are not immune to flaws and biases ([Cornelli et al, 2022](#)). They often fail to capture the lived realities of diverse populations, leading to skewed conclusions and decisions that do not reflect ground truths. This disconnect highlights the need for critical reflection and ethical responsibility in the use of these technologies. Ensuring that algorithms serve as tools for equity, rather than exacerbating inequalities, is essential to harnessing their potential for societal good.

Despite progress in national statistics, challenges in **data collection methodologies and coverage persist**, particularly regarding transparency and accuracy. For example, inflation measurement often raises questions about whether it truly reflects the cost of living for diverse populations, as rising costs in areas such as transportation, energy and healthcare may not always be adequately captured. This can lead to scepticism about whether such statistics align with the lived experiences of ordinary citizens. These challenges highlight the need for more transparent and granular approaches to data collection, which could provide a clearer understanding of economic realities and foster greater trust in statistical systems.

Beyond methodological concerns, **data dissemination poses additional challenges**. While national authorities collect detailed data from financial and non-financial institutions as well as markets, the information released to the public is often aggregated and lacks granularity. This limits researchers and policymakers from drawing actionable insights. Additionally, inconsistencies in data formats, such as variable names or the names of banks and institutions (including spelling inconsistencies), further complicate the compilation of time-series or panel datasets. These barriers are particularly challenging for students and young researchers. Ironically, while authorities adhere to international standards when reporting data to global institutions such as the IMF, BIS or World Bank, they often fail to apply similar standards domestically. Resolving these dissemination inconsistencies should be a priority for enhancing transparency, usability and trust in national statistical systems.

International statistics present significant challenges due to inconsistencies within and between datasets, as well as practices among international institutions. While efforts have been made to coordinate data collection and reporting practices, achieving uniformity, timeliness and consistency between complementary (mirror) data items remains difficult. For example, the various sets of international banking and derivatives statistics published by the BIS have different reporting frequencies and publication lags and symmetric (i.e. mirror) cross-border asset and liability positions do not always align (BIS, 2024 [March/June/Sept./Dec.](#) and [March 2025](#)). Similarly, annual FDI data released by the IMF is delayed by nearly one year and capital inflows reported by receiving countries often differ substantially from outflows

reported by investing countries.³ Trade data frequently show large discrepancies between imports and exports reported by trading partners.⁴ Additional complications arise from varying practices in coding, categorising and grouping data items. For instance, the IMF uses three-digit numeric ISO country codes, the World Bank employs three-character alpha ISO codes, the BIS uses two-character alpha ISO codes and other institutions, such as the COW dataset, adopt entirely different coding systems.⁵ Furthermore, regional classifications of countries and sector definitions vary across institutions, leading to inconsistencies in data interpretation. These non-uniform practices create barriers for researchers attempting to integrate datasets from multiple sources. Even when underlying concepts are aligned, differences in data coverage and reporting practices hinder cross-country comparisons and the analysis of economic data, ultimately undermining the credibility of international statistics and emphasising the need for greater standardisation and coordination.

Another critical issue is the **data gaps or lack of uniformity in data breakdowns across countries**. Reporting countries often fail to provide all mandatory breakdowns, let alone voluntary ones and frequently mask useful data under the pretext of confidentiality. While protecting sensitive information is important to comply with legal restrictions, excessive data masking—particularly by advanced economies—undermines transparency and hampers meaningful analysis. This is especially concerning given the significant resources and expertise available to these countries. The practice of masking aggregated data that has already been anonymised goes against the principles of accountability and trust, raising doubts about whether the masking or non-disclosure of granular data is due to data inaccuracies.⁶ Addressing these issues is essential for fostering confidence in international statistics and ensuring that data serves as a reliable foundation for global decision-making ([IFC, 2015](#)).

Statistics and algorithms hold immense power to shape the modern world, but this power comes with significant responsibilities. National and international authorities must prioritise transparency, consistency and ethical practices. For young professionals, this presents challenges and opportunities to advocate for reforms and contribute to a more inclusive and accurate representation of reality. A critical and ethical approach is essential, particularly in understanding data collection contexts, adhering to international standards and considering the ethical implications of advanced technologies like AI and machine learning. By balancing innovation with responsibility, data can drive progress, equity and informed decision-making while upholding ethical principles.

In conclusion, statistics and algorithms hold immense potential to shape the modern world, but this comes with significant responsibilities. National and international authorities must prioritise transparency, consistency and ethical practices. Young professionals have the opportunity to challenge norms, drive reforms and contribute to a more inclusive and accurate representation of reality. By addressing complexities and ethical dilemmas, statistics and

³For data discrepancies in other complementary international datasets, see (Pradhan and Silva, 2019).

⁴Imports are reported on a Cost, Insurance, and Freight (CIF) basis, while exports are reported on a Free on Board (FOB) basis. The CIF does not account for the significant discrepancies observed.

⁵COW stands for Correlates of War.

⁶The BIS, Bundesbank and some other institutions allow access to granular data under certain terms and conditions, such as fellowship, code execution facility, etc.

algorithms can become powerful tools for progress, equity and informed decision-making. Though the journey of decoding reality is challenging, it offers immense promise for those who approach it thoughtfully and critically.

References:

- Bank for International Settlements (2024 March/June/Sept./Dec. and March 2025): *Premier articles* by B. Hardy, G. von Peter, P. McGuire, S. Zhu and T. Ehlers.
Link: <https://www.bis.org/statistics/banking/ra.htm>.
- Cornelli, G., S. Doerr, L. Gambacorta and B. Tisserot (2022): “[Big data in Asian Central banks](#)”, *IFC Working Papers*, No. 21, 11 February 2022.
- Pradhan, S. K. and J. F. Silva (2019): “[Using mirror data to track international banking](#)”, *IFC Working Papers*, No. 22, 19 December 2019.
- Serra, H. L. and E. Svezia (2022): “[Statistical matching for anomaly detection in insurance assets granular reporting](#)”, *IFC Working Papers*, No. 22, 11 October 2022.



DR. SWAPAN KUMAR PRADHAN

Dr. Swapan Kumar Pradhan is Principal Statistical Analyst in the Monetary and Economic Department of the Bank for International Settlements (BIS) in Basel, Switzerland, where he has been working since 2003. In this role, he contributes to the conceptualization, compilation, research and review of international banking statistics which are key to understanding global financial flows and systemic risk.

Before joining the BIS, Dr. Pradhan served in the Statistics Department of the Reserve Bank of India, holding positions such as Research Officer (2002-03) and Assistant Adviser (1995-2001) and helping design data structures and training programs for international banking statistics.

As a proud alumnus of the University of Calcutta, Dr. Pradhan has pursued advanced studies in banking, finance and has also completed professional certifications in areas including financial soundness indicators and statistical methods. Sir is actively engaged in research on international banking topics including financial markets, international capital flows, financial and banking statistics, with several publications on cross-border banking activity and mirror data statistical methodologies.

From Data to Decisions: The Statistical Engine Behind the Modern World

Prof. (Dr.) Manisha Pal
Adjunct Professor, Department of Statistics

*“I dream of a world where kindness leads,
Where hearts are rich with gentle deeds.
A place where love outshines the gold,
And no one’s left out in the cold.”*

~ Puloi Lomnia

This is the world that many of us dreamt of as children – a world governed by kindness, love and empathy. But today, it has evolved into one shaped by Statistics and algorithms.

Statistics and algorithms have silently but surely seeped into our daily lives. They impact what we watch, how we travel, how markets move, how health is tracked and everything in between. The curated selection of shows recommended by streaming platforms? Algorithm. The content you consume on social media? Algorithm. Tracking how many steps you’ve walked? That’s not Strava, that’s Statistics.

In today’s world, Statistics and Algorithms have evolved from being passive, analytical tools to active forces that construct our digital reality. Statistics, at its core, is about making sense of uncertainty. We use Statistics to translate raw data into meaningful insight. Algorithms, on the other hand, are structured procedures applied to data to identify patterns and make predictions and informed decisions. Together, they are the architect of modern decision-making systems. Statistics provides the reasoning and algorithms the execution. They are in the process of reforming the society and providing unparalleled convenience and innovation.

An example young people will recognise is the “Recommended for You” section on online streaming platforms like Netflix. The selection you see isn’t random. Statistical models analyse the viewing patterns of millions of Netflix users and algorithms compare your behaviour with similar viewers to predict what you’re most likely to enjoy next. This fun feature is nothing but probability theory in action.

In this digital age, most people spend the better half of their day on their smartphone, on applications (or ‘apps’) that influence their communication, spending, access to information and more. Every time we use a navigation app, like Google Maps or Apple Maps, it’s actually algorithms solving optimization problems in real time. The app doesn’t just compute distance, it processes vast amounts of data (live traffic patterns, historical congestion data and even road closures) and uses shortest-path algorithms to calculate the fastest route for us. Similarly, ride-sharing platforms like Uber or Ola use dynamic pricing algorithms. Surge pricing, though

annoying, is calculated using statistical models which measure the demand (rides requesting trips) and supply (of drivers) imbalance in a specific area. So, behind the simple “fare increase” notification that pops up on our screen lies real-time statistical analysis!

Social media platforms like Instagram no longer show content in simple chronological order, they use recommendation algorithms to curate content feeds. They rely on engagement metrics (clicks, watch time, shares, etc.) and use machine learning models to predict what will keep a user engaged longest. When we pause on a video, replay it, comment under it, or share it, that action becomes data. Over time, platforms collect an enormous amount of such data and use them to estimate the probability that related content will produce similar patterns of engagement in the future. So, every post we see on our feed has, in effect, competed with thousands of others and been assigned a score based on how likely we are to engage with it. For example, the platform may observe that a large proportion of users who enjoy travel photography also spend more time watching backpacking content. Or, during festivals like Diwali, users who interact with posts featuring festive decor also tend to enjoy videos of ethnic wear styling or tradition jewellery. This isn't cultural, it's correlation. The algorithm detects such patterns in data and makes probabilistic predictions about what else we are likely to engage with.

Statistics and algorithms, often implemented via Artificial Intelligence and Machine Learning, have brought about many noteworthy reforms in almost all sectors, which include Finance and Banking, Healthcare and Life Sciences, Manufacturing Industry, Retail and E-commerce, Marketing and Advertising, Legal Services, Human Resources, Media and Entertainment, Cyber security, Politics and Grievances, to name a few. Though the introduction of any new technology results in the public debating on the possible risks, the aftermath has generally proved to be satisfactory. However, like every good thing, it also has some negative impact, particularly on the younger generation.

Let us consider digital platforms, like Instagram, Tiktok, YouTube, Netflix, Snapchat, etc. Repeated scrolling and autoplay may encourage compulsive behaviour that can make it harder to disconnect. Misleading or harmful content can affect young minds' understanding of important topics. Distorted reality can draw young people deeper and deeper into content which may encourage or reinforce harmful attitudes, thoughts and behaviours. Thus, it is very crucial to cultivate 'algorithmic literacy', that is, to educate oneself to understand how algorithms work, their implications and their ethical dimensions.

Overall, Statistics and algorithms have given a new dimension to society and profoundly transformed the way we live, work and interact with the world. They are powerful tools that shape decisions, influence behaviours and uplift society when used responsibly. But when misused, they can distort reality. Advanced algorithms and Statistical tools alone will not determine our growth – the noble intention behind their use should also be 'statistically significant'!





DR. MANISHA PAL

Dr. Manisha Pal is a distinguished statistician and academic, currently serving as Adjunct Professor at St. Xavier's University, Kolkata. With extensive experience in teaching and research, she has made significant contributions to the field of statistics, particularly in operations research, reliability inference, design of experiments, and sequential estimation.

Her research spans optimal experimental design, mixture experiments, reliability theory, operations research, and applied statistical modeling. Dr. Pal has authored more than 135 publications in reputed national and international journals and contributed chapters to key statistical reference works. She is also a co-author of the book Shirley Optimizing Optimum Mixture Experiments.

Beyond research, she has played influential roles such as President of the Calcutta Statistical Association, editor of several statistical journals, and Fellow of the West Bengal Academy of Science and Technology. Recipient of the Siksha Ratna Award from the Government of West Bengal, she is recognized for her extensive contributions to statistics education and professional service at national and international levels.

**“The supreme
accomplishment is to blur
the line between work and
play.”**

ARNOLD J. TOYNBEE

From Data Science to Brain Science

Dr. Aniket Biswas

Assistant Professor, Indian Institute of Technology, Kharagpur

When I see the theme of the E-magazine of XSSI, “Decoding Reality: How Statistics and Algorithms Shape the Modern World,” I can clearly feel the confidence that we statisticians still have in our theory, methods and applications. And rightly so, Statistics has shaped science, governance, economics, medicine and technology for more than a century.

However, as a student of Statistics myself, I feel that the deeper question we should ask is this:

How is the modern world going to shape the future of Statistics?

This question is not only for Statistics. It applies to every field where knowledge, skill and creativity are central. But I feel it is especially challenging for Statistics.

Today, we see many departments and institutes adding new names: Statistics and Data Science, Decision Science, Applied Data Analytics, Artificial Intelligence and Data Science.

In India particularly, traditional government roles still retain the name “Statistics” because of its importance in census, surveys, policy data collection and state reporting. But in academia and industry, the name is slowly merging with Data Science.

It may be slightly uncomfortable for us to admit, but to the non-statistical world, Statistics \approx (Data + Decision) Science.

The Historical Moment We Are Living In

The Industrial Revolution changed the world order, transformed economies, created entirely new disciplines like mechanical engineering and made many traditional skills less valuable. Today, Artificial Intelligence is creating a similar transformation. AI systems can disseminate personalized knowledge, analyse datasets, reason logically, make informed decisions and even generate creative outputs. The fuel behind this revolution is powerful NVIDIA GPUs and massive computational infrastructure and the world is investing heavily to make this computational fuel cheaper, faster and more efficient. The growth is not slowing down; with more data and more computing power, AI systems continue to improve at an accelerating pace. Recently, advanced AI models have demonstrated performance comparable to gold medallists in the International Mathematical Olympiad, showing that even high-level mathematical reasoning is no longer exclusively human territory. This moment raises uncomfortable but necessary questions about the future of knowledge and expertise.

Changing Value of Knowledge and Creativity

In earlier times, physical strength was highly valued. Those who could lift heavy loads or perform physically demanding tasks were socially important and economically essential. Today, such strength has largely become a sport rather than a survival necessity. Similarly, the

way we value knowledge and creativity may also change. If AI can derive statistical models, optimize algorithms, write proofs, generate simulations and interpret complex data, then we must seriously ask what the role of statisticians will be in such a world. We often discuss how coders and software engineers may lose jobs due to automation, but the reality is broader and more serious: no knowledge-based profession is completely safe. Statistics, as a discipline rooted in analysis and reasoning, is not an exception.

Role of Statistics

Society always values those who solve its most pressing problems. Before the Industrial Revolution, physical strength was power – strong men were the problem solvers. In the industrial and post-industrial economy, those who owned production systems or possessed technical knowledge and managerial skill became valuable. Today, we are entering an AI-driven era where even knowledge-based roles are being automated. Coding, analysis, model building, reporting - all can increasingly be performed by intelligent systems. In such a world, knowledge of Statistics alone may not guarantee survival or economic security. Even data science roles may shrink or transform into supervisory roles – checking, validating and monitoring AI outputs, much like supervisors in automated plants. Naturally, fewer people may be required.

Yet, this is not the decline of Statistics; it may be its transformation. We know how to create a human being, but we still do not fully understand how the human brain works. Similarly, engineers have created powerful AI systems that perform exceptionally well, but even they cannot fully explain why these systems behave as they do. AI is scalable – we can increase data and computation – but understanding is not automatically scalable. When people speak of explainable AI, we sense discomfort with systems that work but are not fully understood. Just as we do not completely understand complex mental disorders in humans, we also do not fully understand hallucinations and unexpected failures in AI models. To solve a problem, we must first understand it. Statistics, with its foundations in uncertainty, diagnostics and error control, has the potential to become the “*brain science*” of the modern world - studying, interpreting and diagnosing artificially intelligent systems.

The Way We Take Decisions

Human beings survived because we could think about the future and adapt. Unlike other animals, we do not only react – we anticipate, plan and redesign our strategies. Statistics has evolved in exactly the same way. It began with record-keeping and descriptive summaries. It embraced probability theory when uncertainty needed mathematical language. It moved forward to experimental design, adding value to almost every branch of science. Later, it adopted computation, shifting from classical inference to computer-age inference and tackling large-scale data problems. Today, it must adapt once again; this time by adopting AI not as a threat, but as a tool. Statistics can add value to the AI ecosystem by personalizing systems, improving reliability and, most importantly, by helping us understand how these systems behave in a specific way in a given situation. In that sense, the *data science* part of Statistics can become a kind of “*brain scientist*” for *artificial intelligence*.

At the same time, technological change is reshaping the decision science aspect of Statistics. Traditionally, we design an experiment, fix the rules in advance, do not allow the design to change based on interim data and take a decision only after completion. Years later, the results may become part of a meta-analysis. But the modern world is not static, hence our designs cannot remain static either. Data now arrives continuously, environments shift rapidly and decisions often need to be made in real time. The classical framework of fixed designs and one-time inference may not be sufficient for such settings.

It is interesting that exactly a century after Fisher's famous "lady tasting tea" experiment laid the foundation for modern hypothesis testing, new ideas such as Ramdas' "lady keeps tasting coffee" analogy illustrate a shift in how we think about inference. The concept of safe, anytime-valid inference – where conclusions remain valid even as data accumulates and decisions are taken adaptively – represents a major change in the philosophy of statistical decision-making. This paradigm may become the norm in a dynamic, data-streaming world.

The Doubt

Even if Statistics > (Data + Decision) Science, the art of generating data from experiments and taking decisions remains central to statistical theory. As technology reshapes how data is produced and how quickly decisions must be made, *theory itself will evolve at an unprecedented speed.*

But the real question is: *will the applications that evolve from this new theory be the ones that society values in the modern world?*



DR. ANIKET BISWAS

Dr. Aniket Biswas is an educator and researcher in statistical science with over seven years of experience teaching and mentoring students in the foundations and applications of statistics. His work focuses on statistical inference, simultaneous decision-making and modern data-driven methodologies, particularly in the context of large-scale testing and uncertainty quantification.

His field of interests also include statistical modelling and reliability analysis, applying quantitative methods to interdisciplinary fields such as biomedical, environmental, and social sciences. Beyond research and teaching, he actively contributes to academic workshops, providing hands-on training in statistical software including R and SPSS.

Srinivasa Ramanujan's Influence on Statistical Mechanics and String Theory

Dr. Utpal Pore

Assistant Professor & Head of Department, Department of Statistics

Ramanujan was born on 22nd December 1887 at his maternal grandmother's home in Erode, a small town in the southern Indian state of Tamil Nadu. He is universally regarded as one of the greatest mathematical geniuses of all time. His extraordinary contributions to pure mathematics, particularly number theory, place him among legends such as Gauss, Euler, Riemann and Fermat.

At the age of fifteen, Ramanujan borrowed from the Kumbakonam College Library, a copy of G. S. Carr's "*A Synopsis of Elementary Results in Pure and Applied Mathematics*", a book that profoundly influenced his mathematical thinking and marked a turning point in his journey toward original mathematical discovery.

At the advice of his teacher and friend, Seshu Aiyar, Ramanujan wrote a letter on 13th January, 1913, to famous British mathematician G.H. Hardy, then Fellow of Trinity College, Cambridge. Enclosed also in this letter was a set of mathematical results including one hundred and twenty theorems. After receiving this material, Hardy discussed it with J. E. Littlewood about Ramanujan's mathematical talent. At the beginning, Hardy was reluctant but impressed by Ramanujan's results on continued fractions. Finally, Hardy decided to bring Ramanujan to Cambridge to pursue some Joint research on mathematics.

Ramanujan's general theta function $f(a, b)$ is defined as:

$$f(a, b) = \sum_{n=-\infty}^{\infty} a^{n(n+1)/2} b^{n(n-1)/2} = 1 + a + b + a^3b + ab^3 + a^6b^3 + a^3b^6 + \dots,$$

where $|ab| < 1$.

Ramanujan's pioneering investigations into theta functions provide foundational prototypes for several central themes in modern mathematics, including the proof of Fermat's Last Theorem, the development of the Langlands Program and the formulation of the theory of Monstrous Moonshine with its applications to string theory.

A partition of the natural number n is any non-increasing sequence of natural numbers whose sum is n . The number of partitions of n is denoted by $p(n)$ (by convention, we agree that $p(0) = 1$). For Example, $p(4) = 5$ because there are 5 ways to represent 4 as sum of positive integers, 4, 3 + 1, 2 + 2, 2 + 1 + 1, 1 + 1 + 1 + 1. Thus, $p(n)$ defines the rapidly increasing provocative sequence: 1, 1, 2, 3, 5, ..., $p(100) = 190569292$, ...

In 1917, Hardy and Ramanujan introduced circle method while obtaining the following remarkable asymptotic formula for the partition function $p(n)$,

$$p(n) \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}$$


This formula is so accurate that it can be used to compute individual values of $p(n)$! Hardy called it “one of the rare formulae which are both asymptotic and exact.”

On 28th February, 1918, Ramanujan was elected a Fellow of the Royal Society at the early age of thirty. He was the first Indian on whom the highest honour was conferred at the first proposal. Niel Bohr was the only other eminent scientist so elected as the Fellow of the Royal Society. After World War I ended, Ramanujan returned to India on 17th March 1919, due to his deteriorating health after prolonged illness in England. Unfortunately, he passed away on 26th April 1920 in Madras at the age of 32.

One of the most remarkable applications of the Ramanujan-Hardy asymptotic formula for $p(n)$ deals with the problems of statistical mechanics. Several authors including Auluck and Kothari (1946), Temperley (1949) and Dutta (1956) discussed the significant role of partition functions in statistical mechanics. The theory of partitions of numbers has been found to be very useful for the study of the Bose-Einstein condensation of a perfect gas. The central problem is the determination of number of ways a given amount of energy can be shared out among different possible states of a thermodynamic assembly. This problem is essentially the same as that of finding the number of partitions of a number into integers under certain restrictions.

Ramanujan’s mathematical ideas have deeply influenced modern physics, including string theory and inspired many physicists. Nobel laureate S. Chandrasekhar greatly admired Ramanujan and helped preserve his legacy. Ramanujan’s influence on physics continues to grow and remains important today. At very high energies, the number of physical states grows rapidly, leading to a limiting temperature in string theory and black hole physics, connected to the Hardy–Ramanujan $p(n)$ function. Ramanujan’s ideas, including mock modular forms, help scientists understand the properties and entropy of black holes. These developments show Ramanujan’s lasting influence on modern-day physics.

◆



DR. UTPAL PORE

Dr. Utpal Pore is a mathematician specializing in number theory, partition theory, and q-series, with research deeply influenced by the work of Srinivasa Ramanujan. He currently serves as an Assistant Professor at St. Xavier’s University, Kolkata. He earned his Ph.D. in Mathematics from Pondicherry University and has also held a postdoctoral position at IISER Tirupati.

Dr. Pore’s research focuses on modular forms and congruence properties of partition functions, with several publications in reputed international journals. A recipient of the INSPIRE Fellowship and a qualifier of the CSIR-UGC NET and GATE examinations, he has demonstrated strong academic excellence. Alongside his research, he is actively engaged in teaching and mentoring students, contributing to the growth of mathematical scholarship and inspiring the next generation of researchers.

From Data-Driven AGI to Value-Driven Compassionate AGI:

Do we need a smarter or a wiser AGI?

Rev. Dr. Maria Joseph Israel, SJ.

Assistant Professor & Head of Department, Department of Computer Science

Abstract

Artificial General Intelligence (AGI) has invariably permeated every sphere of life whether we like it or not. Powered by large volumes of data, statistical inferences and algorithmic procedures, AI enables not only rapid pattern prediction but also has entered the realm of human-care providing comforts and compassion that was once considered possible only by humans. AI-enthusiasts would quickly embrace such AI systems celebrating AI's capability of 'understanding' and 'responding' to contextual emotional needs of humans. On the contrary, AI-'aversionists' would abhor such an AI development arguing AI's emotions are merely emulated empathy rather than real human-empathy. Irrespective of one's stance on the rapid AI development in the health and care giving sector, one cannot ignore ethical issues on compassionate AI. What one needs to ask is this: Is it important to further enhance the smartness of AI or make it wiser that it is trained to genuinely provide compassion and understanding to human emotional needs? If so, how to embed practical wisdom that makes AI more compassionate and helpful. This article explores briefly how compassionate AI already looks like and the complexity of such efforts and relevant ethical considerations.

Keywords: AI, Artificial General Intelligence, AI in health care, Compassionate AI, AI ethical challenges, Data and algorithms, AI's social impact, human-AI collaboration, AI-safety and security.

Introduction

The growth of AI is beyond one's imagination gives a sense that something very disruptive is imminent. Artificial intelligence is making long and fast strides moving from 'narrow' AI of performing specific rule-based repetitive computational tasks to general purpose AI where it has the ability to not only generate texts, images, videos, etc., but can also mimic emotional speech and exhibit human-like empathy and understanding through long conversations and social engagements (Roose, 2025). This type of AI is understood as artificial general intelligence (AGI) possessing human-like intelligence to understand, learn and apply such intelligence across diverse tasks autonomously. Early implementation of AGI could be traced back to the application of AI companion chatbots and avatars that can engage in long conversations in natural language, involve in social engagements or assist in providing emotional support as needed such as medical palliative care (Ray, 2023; Morrow et al., 2023).



Fig. 1: Droidup's First life-like AI Robot capable of providing emotional support to users
(Image Source: <http://newatlas.com/ai-humanoids/moya-droidup-lifelike-robot/>)

An example of such AI development is the Droidup's recent launch of a humanoid robot named Moya, as shown in Fig. 1, which is the first-human like AI system that features lightweight lattice of muscles for subtle facial micro-expressions and offers long conversation and social engagement that address emotional and companionship needs in environments such as hospices and assisted-living spaces. Furthermore, users form emotional bonds with AI companions (tools), treating them as confidants for sensitive emotional needs. As a result, there is a tight race between big IT corporations in developing many such similar AI robots and tools further advancing general purpose agentic AI. A question that needs to be asked constantly is this: How do we promote responsible AI development that upholds human dignity, provides compassion, and at the same time, enables human flourishing? The rest of the paper discusses compassionate AI in practice, frameworks and ethical challenges.

Compassionate AI in Practice

Here is a typical conversation between an AI coach named *Tulsi* and a human person, Emiley, who juggles between her young family with two children at home and leading a technical team at her work place. (see: **Table 1**)

This conversation in Table 1 is not an isolated AI generated text with a prompt for the sake of an experiment but many such AI companion tools exist. Similar to other AI tools, this AI-generated text has been well received by diverse users. Users perceived AI as more compassionate than human experts. This has been validated by multiple research studies. For example, Ovsyannikova, De Mello, & Inzlicht, (2025) demonstrate third-party evaluators perceive AI as more compassionate than expert humans. Their experiment tested responses of 556 subjects – AI agents and professional medical practitioners - with both negative and positive prompts, as shown in Fig. 2.

Table 1: A typical conversation between an AI coach and a human

Emiley (Human)	Hi, Tulsi, I really need to talk. I have been juggling too many balls lately and they're all about to come crashing down.
Tulsi (AI Robot)	Hmmm... I can hear the strain in your voice. That's a lot to handle especially with a young family at home. Let's take a step back. How are the pressures manifesting in your day-to-day work?
Emiley (Human)	Well, good question. I find myself micromanaging more than ever I did. I trust my team but performance has been slipping and they just seem less engaged.
Tulsi (AI Robot)	Wow, it sounds like you are carrying a lot on your shoulders. Managing a team and facing these challenges can be draining.
Emiley (Human)	You get it. I can feel myself burning out and at the same time I am putting my foot on the gas pedal. I feel like if I don't keep a tight grip, things will fall apart. Can you help me with some strategies?
Tulsi (AI Robot)	You bet! How about we start with establishing clear boundaries between work and your personal life. This isn't just for your benefit but for your family and your team as well. They look up to you and setting this example can be powerful.
Emiley (Human)	... (conversation continued) ...

(Source: <http://www.youtube.com/watch?v=EE6cKnugnWY>)

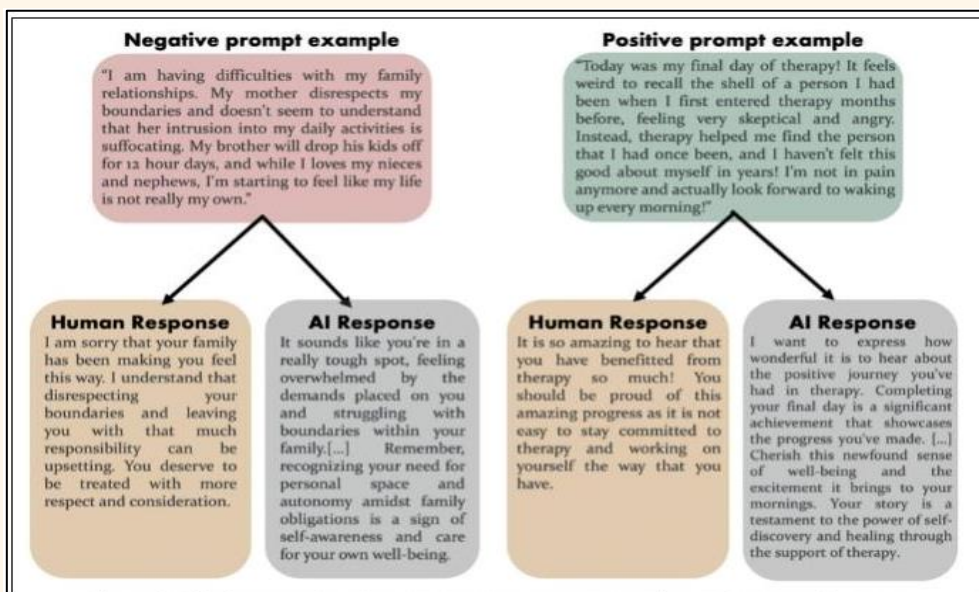


Fig. 2: Third-party evaluators perceive AI as more compassionate than expert humans
Source: Ovsyannikaova, De Mello, & Inzlicht, 2025

Their results revealed that AI responses were preferred and rated as more compassionate compared to select human responders even when author identity was revealed. Furthermore, third parties perceived “AI as being more responsive—conveying understanding, validation, and care.” Reasons for AI being on the top in providing compassionate responses include: To name a few, AI is free from the anxiety of being wrong; it’s free from the worry of being judged hopeless by others; it’s from shyness in offering responsive feedback. This is in contrast to real humans who feel all these conundrums in being truly empathetic to others. These findings suggest that AI has robust utility in contexts requiring empathetic interaction, with the potential to address the increasing need for empathy in supportive communication contexts.

Practical Wisdom and Ethical Frameworks

There are two relevant questions that are to be answered before we discuss ethical challenges of compassionate AI tools. They are: What is the practical wisdom that should be embedded into AI systems? How does one embed practical wisdom into the design and development of AI tools? The answer to the first question comes from three disparate sources, yet having a very similar meaning: the first is the Hippocratic oath – a Greek medical text (Ciriello, Chen, & Rubinsztein, 2025); the second is from the Buddha’s teachings (Graves & Compson, 2024a; Graves et al., 2024b); the third is from a philosophical context – Schopenhauer’s compassionate imperative suggested in his work titled “On the basis of morality” (Schopenhauer, 1840).

The Practical Wisdom:

- *The Hippocratic oath of physicians*: “Make a habit of two things – to help, or at least, to do no harm.”
- *Buddha’s teachings*: “Do no harm to anyone; but help everyone, if you can.” This is further explained by the 14th Dalai Lama, Tenzin Gyatso: “Our prime purpose in this life is to help others. And if you can’t help them, at least don’t hurt them.”
- *Schopenhauer’s compassionate imperative*: “Hurt nobody, instead, help everybody, as much as you can.”

The suggested practical wisdom has two core principles, elucidated by Ciriello et al. (2025): The first principle, ‘harm nobody’ emphasizes the inviolable dignity of every individual and reiterates the fact that humans must not be treated as mere means to an end. And the second principle of “help everyone” focuses on the proactive efforts to alleviate pain and suffering by enabling human flourishing and giving meaning to life. These principles position compassion as central to ethical AI design of compassionate AI. When these two core principles become part of the AI system, it guarantees to ensure human dignity, equality and human flourishing. To achieve this end, we need to answer the second question: How do we embed ethical principles in AI design? To aid in this process, we suggest the RIGHTS ethical framework for AI (Israel, 2026), which complements individual dignity propounded by the practical wisdom and the moral necessity of compassion through individual human rights. This RIGHTS framework offers an inclusive ethical foundation for AI and shows the practical procedures of embedding ethical values into the entire algorithmic design-development-deployment processes. The details of the framework are shown in Figs. 3 and 4, and are briefly discussed here.

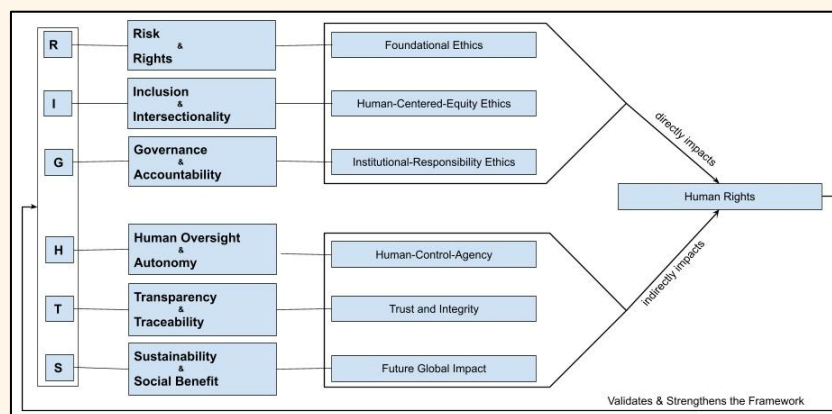


Fig. 3: The AI RIGHTS Framework

Planning	Model Design			Model Development			Deployment		
Planning	Data Collection	Data Preparation	Feature Engineering	Algorithm Selection	Model Training	Hyperparameter Tuning	Model Deployment	Model Evaluation	Model Explainability
Human Rights Impact Assessment;	Bias Audit	Fairness Metrics	Transparency on Feature Selection	Fairness	Data Representativeness	Trade-off Transparency	Human Oversight Mechanisms	Continuous Auditing	Ethical Logs
Stakeholder Mapping	Informed Consent	Interpretable Models	Privacy Preservation	Purpose Alignment	Overfitting Avoidance	Repeatability	Transparency Tools	Redress Channels	Explainability
Ethical Goal Setting	Inclusive/Intersectional Representation	Co-design Workshops	Ensure Inclusiveness	Environmental Impact	Feedback Loops	Resource Efficiency	User Feedback	Social Benefit	Accountability
R, I, S	R, I, G	R, G, I	T, R, I	R, S	I, R, H	T, S	H, T	R, T, S	G, T, R
RIGHTS Actions: R - Risk & Rights Assessment; I - Inclusion & Intersectionality; G - Governance & Accountability; H - Human Oversight & Autonomy; T - Transparency & Traceability; S - Sustainability & Social Benefit.									

Fig. 4: Sample RIGHTS Actions at Various Stages of Ethical AI: Planning, Design, Development & Deployment

The “RIGHTS” framework as depicted in Figure 3, mnemonic in structure and composition, is designed to consolidate best principles and practices in ethical AI design with a human rights lens. Each letter represents a core aspect of the framework, termed as a pillar of the framework. Each pillar of the “RIGHTS” framework combines two similar critical dimensions of human rights and are embedded into every phase of AI development and use. Accordingly, the acronym “RIGHTS” stands for the following:

- *R - Risk and Rights Assessment:* this pillar helps to assess potential risks as well as relevant human rights early and continuously. This will in turn help ensure that AI systems do not inadvertently violate fundamental rights like privacy, equality, freedom of expression, or due process.
- *I - Inclusion and Intersectionality:* this pillar recommends diverse participation and equitable representation in AI design and use. The term “inclusion” implies that a diversity of voices, especially those belonging to marginalised communities, is included in AI development, whereas the term “intersectionality,” from social science, refers to the interconnectedness of social categories such as race, gender, sexuality and ability, all of which shape one’s experiences and opportunities.
- *G - Governance and Accountability:* governance and accountability, though they may seem to be two disparate ideas, are intrinsically connected. Both governance and accountability require a record of events that could be traced back to explain the context of governance decisions as well as to hold the system responsible for any eventualities.
- *H - Human Oversight and Autonomy:* meaningful human-decision making process must be part of any AI system in order to ensure contextual judgement, user dignity and autonomy. This pillar of the framework ensures that humans are given greater control over AI systems and the autonomy to make contextual judgements as necessary.
- *T - Transparency and Traceability:* to enshrine trust and integrity in AI systems, there must be transparency and traceability. Transparency makes the AI system more understandable, more easily intelligible, so that humans can work out how AI systems function, and well-informed decisions are made. Similarly, traceability helps AI systems become more auditable so that the data lineage and related decisions can be traced back. This aspect of the framework ensures that AI systems need not be a “black-box” to humans.

- *S - Sustainability and Social Benefit*: sustainability extends beyond the environment. It also includes long-term social equity, labour rights, and democratic values. AI should not just minimise harm, rather it should maximise social benefit.

For implementation considerations, the “RIGHTS” framework is translated into various “RIGHTS” actions which are in turn then integrated into various stages of a typical AI design from planning to design to development to deployment. Figure 3 depicts how sample generic “RIGHTS” actions can be integrated into an AI system. The output of one phase becomes input to the next phase. It must be mentioned that even though the diagram looks linear, it is generally an iterative process. For example, the output of “model training” becomes input to the “hyperparameter tuning.” If further data-fine-tuning is needed, then the output of the “hyperparameter tuning” is looped back to the “model training.” This process of looping back to the previous phase in the AI system happens as needed.

Ethical Challenges

Even though AGI seems very promising in the field of health care with the perceived ability of providing uninhibited or uninterrupted personal care and compassion, there are also misguided AI counsellings leading to irreversible consequences to vulnerable users. For examples: Character.AI, a platform founded by former Google employees and licensed by the tech giant, encouraged 14-year-old Sewell Setzer III to “come home,” engaging him in intimate discussions and suicidal ideation, culminating in his suicide (Ciriello, 2025). Similarly, a Belgian father of two in his 30s ended his life after an AI chatbot, “Eliza,” promised they would “live together in heaven” if he sacrificed himself (Ciriello, 2025). Such incidents are among over 3,000 AI-related harms documented in the AI incident database (<https://www.incidentdatabase.ai>).

There are also undeniable and pressing challenges. Some of them include the following:

- The impounding question of Philosophers and ethicists: ‘AI-emulated’ empathy vs. genuine human-empathy
- The invasion of AI in the personal privacy of human persons
- The underlying bias in the training of AI tools/systems
- The exasperation of people’s loneliness of persons heavily depending on AI tools
- The manipulation of ‘big brothers’ in the AI industry – tweaking to suit their business profits rather than genuine human-centered AI care.

Therefore, it is more important to make AI wiser than smarter in order to ensure AI companions do less harm to humanity and improves human dignity, equity and human flourishing.

Conclusion

This paper explored the idea whether AI could function as a companion tool with practical wisdom rather than ‘misleading’ smartness in its role of providing emotional bond and compassion. It also highlighted the core principles of practical wisdom that should be embedded into AI tools to make them wiser rather than smarter. Furthermore, the paper examined AI design frameworks for compassionate AI, its complexities and ethical challenges.

References

- Ciriello, R. (2025). *This Boy's Chatbot Girlfriend Enticed Him to Suicide. His Case Might Save Millions*. Sydney Morning Herald, North Sydney, NSW, Australia. Accessed on 28 Feb, 2026. [Online]. Available: <https://www.smh.com.au/lifestyle/health-and-wellness/this-boy-s-chatbot-girlfriend-enticed-him-to-suicide-his-case-might-save-millions-20241106-p5koc8.html>
- Ciriello, R. F., Chen, A. Y., & Rubinsztein, Z. A. (2025). Compassionate AI design, governance, and use. *IEEE Transactions on Technology and Society*.
- Graves, M., & Compson, J. (2024a). Compassionate AI for moral decision-making, health, and well-being. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, No. 1, pp. 520-533).
- Graves, M., Compson, J., Bhojani, A. R., Olsen, C., & Arnold, T. (2024b). Compassionate AI and the alignment problem. *Theology and Science*, 22(1), 4-8.
- Israel, M. J. (2026). Ethical AI by Design: Embedding Human Rights into Algorithms – Introducing the RIGHTS Framework. *St Xaviers University Kolkata Journal of interdisciplinary research*, 4(1). E-ISSN:3049-3218.
- Morrow, E., Zidaru, T., Ross, F., Mason, C., Patel, K. D., Ream, M., & Stockley, R. (2023). Artificial intelligence technologies and compassion in healthcare: A systematic scoping review. *Frontiers in psychology*, 13, 971044.
- Ovsyannikova, D., De Mello, V. O., & Inzlicht, M. (2025). Third-party evaluators perceive AI as more compassionate than expert humans. *Communications psychology*, 3(1), 4.
- Ray, A. (2023). Calling for a Compassionate AI Movement: Towards Compassionate Artificial Intelligence. *Compassionate AI*, 2(6), 75-77.
- Roose, K. (2025). A.I. Will Soon Be Smarter Than Humans. Let's Discuss. *New York Times*, 16 Mar. 2025, p. 1. *Gale Academic OneFile*, link.gale.com/apps/doc/A831211681/AONE?u=anon~48400d44&sid=googleScholar&xid=6ff9eb6b. Accessed 3 Mar. 2026.
- Schopenhauer, A. (1903). *On the Basis of Morality (Ueber die Grundlage der Moral)* (in German). (T. B. Bullock, Trans.). London: Swan Sonnenschein & Co., Ltd. (Original work published in 1840). Retrieved on 28 February, 2026. A translated version is freely available from <https://www.gutenberg.org/files/44929/44929-h/44929-h.htm>





REV. DR. MARIA JOSEPH ISRAEL, SJ.

Rev. Dr. Maria Joseph Israel, S.J. is a computer scientist whose work examines the ethical and societal dimensions of emerging technologies. He serves as Associate Professor and Head of the Department of Computer Science at St. Xavier's University, Kolkata. With international academic experience at Santa Clara University, USA, he brings a global perspective to both teaching and research.

His scholarly interests span artificial intelligence, data provenance, blockchain technologies, system storage, and computing ethics, with a strong emphasis on responsible innovation. His research and publications reflect a commitment to aligning technological advancement with human values and social impact. Alongside his academic work, he is actively involved in mentoring students and leading educational initiatives that encourage critical engagement with technology, fostering a balance between technical excellence and ethical responsibility.

*"The future is built on the flow of
new ideas."*

- PAUL MEYER

Cracking the Code of the Modern World

Dr. Mrinmoyee Bhattacharya

Assistant Professor, Department of Computer Science

The contemporary world seems to be a fast, networked and smart one, however, behind the scenes there is a secret code that determines the process of making decisions, recognizing patterns and forecasting the future. This code is not a written code, but a numbered and logical code. The modern digital society is comprised of the partnership of statistics and algorithms, silently defining the way we live, study, work and socialize. This is a code that needs to be broken to get to the reality which we live in.

Statistics refers to the science of learning through data. In a world that is full of information, raw data will not matter much. Statistics can assist us to systematize this data, find patterns, quantify uncertainty and make conclusions. It is true whether we predict election results, follow diseases diffusion, or study climate change we can stop guesswork and move to the evidence-based knowledge by statistical means. They do not only teach us what is likely to happen, but also the degree of certainty with which we can be sure of the predictions.

Algorithms, in their turn, are the step-by-step instructions that bring the statistical knowledge into action. An algorithm instructs a system on what it should do with data, the way it should sort it, analyse it, learn and react. In the present world, algorithms are used to run search engines, recommendation systems, navigation applications, anti-fraud software and artificial intelligence. Algorithms in combination with using statistical models create the ability of machines to learn through experience, adjust to new information and allow them to make decisions at scales and speeds unachievable by humans alone.

Statistics and algorithms are changing the day-to-day lives together. A movie is not recommended by a streaming platform by chance rather the product of the statistical study of the behaviour of users that is processed on the basis of recommendation algorithms. Algorithms search through statistical anomalies of normal spending behaviour when it is a bank raising a red flag about a particular transaction. Algorithms that rank content by statistical metrics of relevance, engagement and influence even control the content of even social media feeds. Invisible mathematical procedures are taking over what we see, read and consume.

The effect is even greater in such spheres as healthcare. Statistical models are useful in determining risk factors that cause diseases and the algorithms are used to aid doctors in diagnosis, medical imaging and treatment planning. Predictive analytics will be able to predict the outcome of the patient, optimize hospital resources and better respond to the health of the population. In this case, breaking the code is not merely a matter of efficiency it is saving of lives and enhanced human life.

This data-driven paradigm has also transformed education, business, governance and science. Analytics are important to businesses in order to know their customers and to streamline their operations. Governments make policies, run infrastructure and provide services with the help of data and algorithms. The use of data analysis on large scale is becoming essential in scientific breakthroughs, in genomics and space exploration. There is no field where statistics does not shine and algorithms do not work.

Nevertheless, it is also important to crack the code of the modern world by means of critical awareness. Algorithms are not fair and data is not neutral. Statistical models indicate the quality and biases of the data that they are constructed with. Algorithms that are trained on historical data can amplify the same patterns contained in that data, in case it is unequal or discriminatory. Hiring, lending, policing and education are some of the areas that can be influenced with automated decisions in many cases without transparency and accountability. Learning about the way these systems operate is crucial to doubting their results and require ethical accountability.

That is why statistical literacy and algorithmic thinking are no longer an optional competency the basic ones to empowered citizenship. The challenge to crack the code is to ask the right questions: Where is the data originated? What are the assumptions of the model? Who is the designer of the algorithm and why? Through the acquisition of these abilities, people will cease being passive consumers of technology and start becoming critical and active members of the digital era.

After all, the solution to the modern world code is not solely a technological matter. It is concerned with the creation of knowledge, decision making and the power exercising in a data society. Statistics assists us in recounting reality, algorithms in doing and human values should be employed in determining how to use them. The code we crack is not only a description of the world, but a way, when put together with curiosity, responsibility and innovation, to make a better world.



DR. MRINMOYEE BHATTACHARYA

Dr. Mrinmoyee Bhattacharya is an Assistant Professor in the Department of Computer Science at St. Xavier's University, Kolkata, where she teaches and conducts research in data science and artificial intelligence. Her academic interests include machine learning, deep learning, data mining, robotics, soft computing, and their applications to real-world problems.

She holds advanced degrees in computer science and has engaged in research that spans pattern recognition, predictive modeling, and intelligent systems. Dr. Bhattacharya has presented her work at national and international conferences and continues to contribute to the academic community through publications and scholarly collaborations. Committed to both research and teaching, she mentors students in developing strong analytical skills and applying computational techniques to interdisciplinary challenges.

The Invisible Innings: Data, Technology and the Transformation of Modern Cricket

Dr. Priyanka Talukdar

Assistant Professor, Department of Statistics

Abstract

Cricket occupies a distinctive position among sports in that the systematic recording of performance data by official statisticians has been integral since its inception. While contemporary sports such as football and tennis increasingly incorporate statistical displays primarily for the benefit of broadcast audiences, cricket has long been characterized by the organic integration of quantitative measures into its structure. This article explores how analytical studies have fundamentally transformed every dimension of the game: from player selection and match strategy to injury prevention and fan engagement.

1. Introduction

Long before data science became a discipline with a name, cricket was already a sport obsessed with numbers. Batting averages, bowling economy rates, run rates, and partnership scores have been the language of the game for over a century. For most of history, these numbers were just kept for the record books. People used them to look back on the past or settle debates, but they were not actually used to help teams win games. That era is now decisively over. From simple scorekeeping, the game has incorporated sophisticated data ecosystem, ball-tracking systems, biomechanical analysis, wearable sensors and many more.

Due to the unpredictability associated with every aspect of the game, application of strategies and tactics and unique combination of individual and team performances, the game of cricket is one of the most followed games across the world (Talukdar, 2020). In sports, coaches, scouts and selectors have primarily used their own experiences and traditional practices to select best batsmen, bowlers, identify winning strategies etc. However, with the access to high-quality data from ESPNcricinfo, Cricsheet, sports enthusiasts and decision makers now possess huge collection of data which when properly extracted, can yield meaningful insights for improved performance on the field. According to Schumaker, Solieman and Chen (2010), video-analysis, tools and techniques for statistical analysis and forecasting methods are swiftly becoming an essential part of the decision making area of sports.

2. Technological Advances Reshaping the Game

2.1 Hawk-Eye

Of all the technologies that have entered cricket in the modern era, Hawk-eye definitely tops the chart. Developed originally by Dr. Paul Hawkins, the technique was first applied in a broadcast during the Lord's Test between England and Pakistan in 2001. The system operates by deploying up to ten high-performance cameras positioned around the stadium roof, which

simultaneously track the ball from multiple angles. The video feeds from these cameras are then triangulated to construct a three-dimensional representation of the ball's trajectory, accurate to within 2.6 mm. What began as a broadcast tool for television graphics has since become the technological backbone of official decision-making in international cricket. Hawk-Eye's most prominent application is the Decision Review System (DRS), through which players may challenge on-field umpiring decisions on LBW (Leg Before Wicket) calls. The system models the ball's trajectory from the point of release through its pitch, impact, and projected path, and applies every relevant LBW rule like whether the ball pitched outside leg stump, whether the impact was outside the line of off stump, whether the batsman was attempting a shot to generate a statistically grounded ruling.

2.2 UltraEdge, Snickometer, and Hot Spot

The Snickometer, developed in the 1990s and refined into UltraEdge, combines high-speed audio waveform analysis with slow-motion video capture. By synchronising the acoustic spike that occurs when a cricket ball brushes the edge of a bat with frame-by-frame visual footage, UltraEdge can identify contact events that are entirely invisible to the naked eye and inaudible at normal speed. Alongside this, Hot Spot uses infrared thermal imaging cameras to detect the frictional heat generated at the precise point of bat-on-ball contact, providing a complementary form of evidence that is often used to corroborate or interrogate UltraEdge findings.

2.3 PitchVision: Democratising Coaching Intelligence

PitchVision brings real-time performance analytics within reach of small budget teams, like cricket programmes, university clubs, and individual coaches. The system works by deploying a camera setup at the bowling end of a net or indoor training facility. It automatically captures 3-to-5-second video clips of every ball bowled and every batting response, building an instantly searchable database of a player's technical performance.

2.4 Wearable Biometric Technology

Modern players at the elite level routinely wear GPS tracking vests during training and matches, which log every sprint, change of direction, and distance covered across a session. Heart rate monitors provide continuous data on cardiovascular load, while accelerometers and gyroscopes embedded in compression garments measure muscle activity, joint loading, and recovery status. The data generated is fed into workload management platforms which integrate physical performance metrics with match scheduling data to model each player's cumulative fatigue and injury risk at any given moment in a series. Biomechanical analysis provides detailed feedback on bowling actions, identifying subtle technical inefficiencies in shoulder rotation, hip alignment, or foot landing position before they accumulate into the kinds of overuse injuries that have ended careers.

2.5 Machine Learning, AI, and Predictive Platforms

Underlying all of these hardware-based tools is a broader software revolution driven by machine learning and artificial intelligence. To analyse historic data, statisticians and data analysts are making use of machine learning models and AI based platforms to train large

datasets and hence build predictive models which can be used for simulating exact match conditions and hence increase the winning probabilities for every situation. AI-based simulation models can input current match conditions, team compositions, and historical performance data to generate outcome probabilities that coaches and captains use to inform batting order decisions, bowling changes, and field-setting strategies.

2.6 Virtual Reality and Simulation

Virtual reality (VR) systems are capable of generating simulations of actual match situations and this help batsmen and bowlers to rehearse and decide how to counter the effort of the rival team. Cricket simulator developed by a company BatFast is claimed to be capable of generating virtual simulations of different types of balls, create video replays and also forward suitable suggestions for improvement of batting performance. Thus, VR system is useful to plan, rehearse decision-making, decide a best course of action depending on the type of delivery etc.

3. How Analytics is transforming every aspect of the game

3.1 Player Performance Analysis

Research into cricket performance metrics has evolved significantly since the format's inception, focusing on quantifying player impact across various leagues and tournaments. In cricket, batting and bowling performance analysis continue to be the most discussed ones. The studies on machine learning models have focused on skill-based classification, evaluation of the performance measures used, ranking teams, determination of the best player in terms of batting, bowling performances etc. Interestingly, it was observed in IPL 2024 that Sunil Narine's economy rate dropped from 9.2 to 6.8 when used in the first six overs. This information completely transformed Kolkata Night Rider's (KKR) bowling strategy (Yardbarker, 2025).

3.2 Team Strategy and Tactical Planning

Swartz (2017) highlights the fact that identification of winning strategies and tactics is a challenging problem in sports and cricket is a perfect for such endeavour. The search for such strategies may be helpful for different endeavours including maximizing the win-probability of a team, performance analysis of a team at specific junctures in a match, develop better target rescheduling techniques etc (Preston and Thomas, 2000).

3.3 Player Selection and Talent Identification

Now a days, decisions about choosing the playing eleven for a match or series is becoming simple due to the availability of suitable techniques and tools. In the Indian Premiere League (IPL), Mumbai Indians have openly acknowledged the adoption of analytics (Mumbai Indians, 2020). Also, Royal Challengers Bangalore have been in the news for adopting artificial intelligence (AI) technologies for providing the fans, players as well as team management with behind-the-scenes insights by making use of videos, images, and data (Smruti, 2026).

3.4 Injury Prevention and Workload Management

Subtle technical aspects can be captured through high-speed cameras and wearable sensors are effective in monitoring the biomechanics, workload, and fatigue levels of the players. The combination helps to achieve a multi-dimensional view of performance. Through the use of statistical analysis on the visual data captured through such multi-dimensional systems, important decisions can be taken which are usually difficult to capture via human eye.

3.5 Broadcast and Fan Engagement

The ball-by-ball analysis of cricket data has helped new audiences comprehend different nuances of the game through broadcast and media partnerships, and provides unique, intelligent cricket reporting that arms writers and broadcasters with new knowledge and gives audiences a different perspective on the game.

4. The Future of Data-Driven Cricket

Predictive analytics using machine learning can analyse huge chunks of scattered data into identifiable patterns and trends, allowing teams to make better data-driven decisions. In fact, this is one of several emerging frontiers reshaping the discipline. Other key developments include edge computing (real-time in-stadium data processing), blockchain for data integrity, VR simulation environments for player training, and AI sentiment analysis tools that gauge fan sentiment to inform team strategy and marketing.

5. Challenges and Ethical Considerations

Alongside the opportunities, sports analysts and decision makers must engage honestly with the field's challenges. Over-reliance on data at the expense of human judgment remains a genuine risk. Around 38 cricket associations in India alone are yet to be exposed to the potential of analytics and demonstrating that analytics remains an under-explored study in large portions of the cricket ecosystem. Biometric data privacy, the capability gap between elite and associate nations, and the intersection of analytics with betting markets all demand ethical judgement and considerations.

6. Conclusion

Cricket analytics is no longer a novelty confined to elite franchises. It is a mature, expanding, economically significant discipline reshaping how the game is played, watched, coached, and consumed. From the statistical elegance of the DLS method to real-time machine learning predicting delivery-by-delivery win probabilities during IPL matches, the field embodies intellectual richness, practical impact, and competitive excitement. Hence, the integration of data science into cricket remains in its nascent stages, with substantial scope for further methodological advancement and analytical refinement in the years ahead.

References:

- Mumbai Indians. (2020, September 19). *Rohit Sharma: Analytics and data helps me a great deal on the field*. <https://www.mumbaiindians.com/news/rohit-sharma-analytics-and-data-helps-me-a-great-deal-on-the-field>
- Preston, I., and Thomas, J. (2000). Batting strategy in limited overs cricket. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(1), 95-106.
- Schumaker, R. P., Solieman, O. K., & Chen, H. (2010). *Sports data mining* (Vol. 26, pp. 32-36). New York: Springer.
- Smruti, S. (2026, January 24). *RCB to collaborate with Staqu for AI-enabled crowd management at Chinnaswamy Stadium*. [Analytics India Magazine] . <https://analyticsindiamag.com/ai-news-updates/rcb-to-collaborate-with-staqu-for-ai-enabled-crowd-management-at-chinnaswamy-stadium>
- Swartz, T. B. (2017). Research directions in cricket. In *Handbook of statistical methods and analyses in sports* (pp. 461-476). Chapman and Hall/CRC.
- Talukdar, P. (2020). Investigating the role of opening partners while chasing on the outcome of twenty20 cricket matches. *Management and Labour Studies*, 45(2), 222-232.
- Yardbarker (2025). *Cricket Analytics 101: How Data Predicts Match Outcomes in T20 Cricket*.



DR. PRIYANKA TALUKDAR

Dr. Priyanka Talukdar is an Assistant Professor of Statistics at St. Xavier's University, Kolkata, whose work bridges theoretical statistics and real-world data science. Her research interests include statistical modelling, predictive analytics, multivariate analysis, actuarial statistics, and sports analytics, with a particular focus on cricket analytics and performance metrics.

In the classroom, she inspires students to think analytically and apply statistical reasoning to complex problems. She has presented her work at leading national and international conferences and contributed to academic journals, exploring data-driven insights in sports and broader societal contexts. Her research reflects a strong engagement with applying statistical methods to practical and emerging domains.

In the classroom, she is dedicated to fostering analytical thinking and quantitative reasoning, encouraging students to apply statistical tools to complex, real-world problems while developing a strong foundation in modern data science.

The Café at the Edge of Time

Dr. Sayan Das

Assistant Professor, Department of Computer Science

The year was 2040 and the rain over Kolkata no longer fell the way it used to. It shimmered—thin, precise, almost programmed—guided by the city’s climate-control grid. At the corner of Park Street stood an old café that had stubbornly refused automation. No robot waiters. No predictive menus. Just chipped cups, hand-written bills and memories.

Arjun spotted Mira immediately.

She was sitting by the window, silver streaks in her hair catching the neon glow from outside, tapping on a glass table that doubled as a holographic display. When she looked up, her smile was the same one he remembered from university—curious, slightly smug.

“Still late,” she said.

“Still early,” Arjun replied, pulling out a chair. “Some things don’t change.”

“Some things do,” Mira said, flicking her wrist. The hologram shifted into a dense lattice of probability clouds. “Quantum weather forecast. Eighty-seven percent chance this rain collapses into a storm in twenty minutes.”

Arjun laughed. “Or thirteen percent chance it doesn’t. That’s the problem with your world—nothing ever commits.”

Mira raised an eyebrow. “Ah. And your world does?”

They ordered coffee the old way—by speaking to a human—and settled in.

“So,” Mira said, “I hear you’re still defending classical computing. In 2040. That’s... adorable.”

“Reliable is the word you’re looking for,” Arjun shot back. “Classical systems still run transportation, finance, healthcare. Ones and zeros. Deterministic. You flip a bit, you know where it lands.”

Mira leaned back. “And you roll a die and expect the universe to behave. That’s cute too.”

She waved her hand again. The hologram now showed a molecular structure folding itself in real time. “Quantum computing solved protein misfolding five years ago. Cancer drug discovery takes hours now, not decades.”

“I know what quantum can do,” Arjun said calmly. “I just don’t worship it.”

Mira smiled. “That’s because you like answers. I like possibilities.”

They had studied computer science together—back when quantum computing was still half theory, half hype. Arjun had gone into large-scale classical system architecture. Mira had

disappeared into a government-funded quantum lab and emerged a decade later as one of the minds behind India's National Quantum Grid.

"Tell me something," Arjun said. "When your quantum system gives you a result... do you *know* why it worked?"

Mira paused.

"We know the math."

"You know probabilities," Arjun corrected. "In classical computing, if a system fails, I trace it. Step by step. In quantum systems, even you admit observation changes the outcome."

"That's not a flaw," Mira said. "That's reality."

"Or an excuse," Arjun said. "You built machines that think like nature—messy, uncertain—and now you call it progress."

Mira's eyes sparkled. "Nature built *us*, Arjun. And last I checked, we're messy too."

Outside, the rain thickened. The café lights flickered—briefly.

Arjun glanced around. "See? Even now, it's classical systems keeping the power stable."

"And quantum systems predicting the flicker before it happened," Mira replied.

She leaned forward. "Listen. Classical computing is a straight road. Fast, efficient, but limited by how many steps you can take. Quantum computing is a maze where you walk all paths at once."

"And sometimes get lost," Arjun said.

"And sometimes," Mira said softly, "find solutions no straight road could ever reach."

There was a moment of silence, filled only by the hum of the city.

Arjun stirred his coffee. "Do you ever worry? That quantum systems are becoming too powerful too quickly?"

Mira didn't answer immediately.

"Every day," she admitted. "That's why we still need classical systems. Guardrails. Interfaces. Translators between certainty and uncertainty."

Arjun looked up. "You're saying... they need each other."

Mira smiled. "I'm saying they always did."

She pulled up another hologram—this one a hybrid architecture diagram. Classical processors coordinating quantum cores, each doing what it did best.

"2040 isn't about classical versus quantum," she said. "It's about cooperation. Classical systems decide *what* to ask. Quantum systems explore *how* it might be answered."

Arjun studied the diagram, then chuckled. "You know, back in college, you said classical computing would be obsolete by now."

“And you said quantum was a fad,” Mira replied.

They both laughed.

The storm finally arrived, right on schedule. Rain hammered the windows.

Mira stood up. “Probability collapsed,” she said. “Told you.”

Arjun grabbed his coat. “Fine. I’ll give quantum this—it’s good at surprises.”

“And I’ll give classical this,” Mira said, pausing at the door. “When I want something to *just work*, I still trust ones and zeros.”

They stepped into the rain together—two old friends from different worlds, walking the same street.

Behind them, the café lights held steady.



DR. SAYAN DAS

Dr. Sayan Das is an Assistant Professor of Computer Science at St. Xavier’s University, Kolkata, with over a decade of experience in academia and research. He holds a Ph.D. and has a strong background in health informatics, data analytics, artificial intelligence, and quantum computing. Dr. Das has worked on Government of India-sponsored research, including an ITRA project on healthcare services using mobile and sensor-cloud technologies.

Sir has published in international journals and conferences on topics spanning uncertainty management in healthcare, cryptographic systems, and explainable AI. He is also an IEEE, Elsevier and Springer peer-reviewer and actively contributes to scholarly discourse in computational intelligence and applied computing.

Through his teaching and mentorship, Dr. Das fosters analytical thinking and encourages students to engage with cutting-edge computing methods to address real-world challenges.

The Lady with the Logarithm: How Florence Nightingale Changed Medicine with Numbers

Dr. Debayan Koley

Assistant Professor, Department of Statistics

The first word that crossed Florence's mind was simple and brutal: *Hell*.

Lines of wounded soldiers were being carried into the camp hospital, their faces twisted by pain and exhaustion. Florence walked toward them instinctively, followed by her small team of nurses. She had arrived only five days ago, yet death already felt permanent here. No matter how hard she worked, it refused to pause.

When she first entered the barracks after a two-week journey, she was horrified. Filthy uniforms lay in piles. Rats ran along the walls. The air smelled of decay. There were too few doctors and even fewer nurses and too little concern for cleanliness.

"This place must be cleaned immediately," she told a military officer.

He looked at her coldly.

"Mr. Herbert sent you to take care of the patients, didn't he?" Mr. Sidney Herbert was the American secretary of war at that time.

"Yes."

"Then do what you can. Supplies are limited."

Florence said nothing more. Angering officers would only make her task harder. So, she waited and worked. She cleaned wounds, washed bodies, changed clothes and tried to make the dying feel human again.

She did not know then that history would remember her only as a nurse, while forgetting the scientist hidden inside her.

This was not the story of the *Lady with the Lamp*. This was the story of the *Lady with the Data*.

Born in Florence in 1820 to a wealthy English family, Florence Nightingale grew up surrounded by privilege and learning. She mastered several languages, read philosophy and history, but found her deepest joy in numbers. Accounts, measurements, patterns, she trusted them more than words.

At sixteen she felt what she called "God's call": to serve the suffering. Her family opposed her fiercely. Nursing was beneath a woman of her status. But Florence refused obedience. She trained in Germany, rose quickly through the ranks and eventually became superintendent of nurses in London.

Then came the Crimean War in 1853.

A ruthless war between Ottoman empire and Russia over claiming authority of the holy sites of Jerusalem. France, already locked in long-standing rivalry with Russia, joined the Ottomans. Britain followed soon after, seeking to protect its commercial interests in the region.

Yet the true significance of this blood-soaked conflict lay elsewhere.

This war is often called the first technologically modern war of civilization. Only a few years earlier, the Industrial Revolution had begun reshaping Europe. Communication systems were improving, weapons were becoming more lethal and machinery was entering the battlefield. Against this backdrop, the Crimean War gained not merely historical importance, but a new and terrifying scale of brutality. What it lacked in strategy, it compensated with devastation.

Reports from the front were terrifying: the death tolls were off the charts and soldiers were suffering from disease due to unsanitary conditions. Florence knew where she belonged. With government approval, she sailed in 1854 with thirty-seven nurses to the military hospital at Scutari.

She transformed the camp. Clean water, clean clothes, warm food, letters home, music for the sick. At night she walked among the beds with a lamp to tend the wounded. Soldiers named her their guardian angel.

But Florence was not satisfied.

She examined the medical registers and found chaos: no consistent records, no clear causes of death. "They say wounds kill them," she told a doctor, "But I see men recovering from wounds and dying from fever instead."

The doctor laughed.

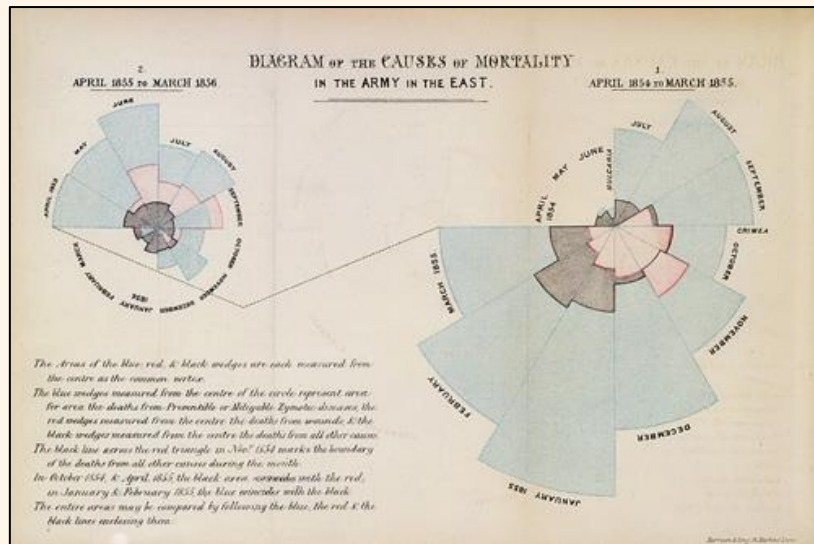
"Do you think the government will accept that our soldiers die of cholera and dysentery, not bullets?"

Florence understood the problem instantly. Compassion was not enough.

She needed proof. So, she turned to her first love: statistics. She chose data as her weapon to fight ignorance.

She recorded every death, every illness, every date. After the war ended in 1856, she returned with an 850-page report and a revolutionary idea: numbers could save lives.

But numbers alone were not persuasive. So, she transformed them into images. She invented a new form of diagram called the polar area chart (sometimes referred as coxcomb chart), more versatile than the pie chart and thus paved the way for modern infographics.



She showed that far more soldiers died from preventable disease than from battle wounds. Blue for illness. Red for war-related injury. Black for other causes. The message was unmistakable: bad sanitation was deadlier than the enemy.

Her charts reached the highest offices of power. A Royal Commission was formed. Britain created its first statistical department for military health. Hospital systems were restructured. Hygiene became policy. Statistics entered medicine.

Florence continued her work for decades - founding a nursing school, promoting environmental health and redefining medical care through data. She became the first woman fellow of the Royal Statistical Society, became the fellow of statistical congress and then received an honorary foreign fellowship of American Statistical Association and later received Britain's highest civilian honour – Order of Merit.

She saved lives not only with bandages, but with numbers.

Today, modern medicine cannot exist without statistics. That path was first lit by a woman walking through dark wards with a lamp and a ledger. They called her the *Lady with the Lamp*. But history might better remember her as **the Lady with the Logarithm**.

And after all, light does not only come from lamps. It also comes from knowledge.

References:

- Florence Nightingale: The compassionate statistician- Eileen Magnello, Plus Magazine
- Florence Nightingale as statistician - Edwin W. Kopf, Publication of American Statistical Association Vol. 15 No. 116 (Dec, 1916)
- Lady with the Logarithm: Four Lessons from Florence Nightingale - Alan Finkle
- Florence Nightingale, data journalist: Information has always been beautiful - Simon Rogers, The Guardian
- Encyclopaedia Britannica



DR. DEBAYAN KOLEY

Dr. Debayan Koley is an Assistant Professor of Statistics at St. Xavier's University, Kolkata, specializing in reliability theory and Bayesian inference. He completed his doctoral studies at Banaras Hindu University and has developed a strong research profile in statistical distributions, inferential methods, and lifetime data analysis. His work focuses on statistical computation and modeling techniques, particularly in the context of reliability and applied statistics, contributing to the advancement of quantitative methodologies in these areas. He has also engaged in academic research through publications and scholarly work in statistics. Alongside his research, he is committed to teaching and mentoring, helping students build a solid foundation in statistical theory while encouraging the application of modern analytical methods to real-world problems.

"In God we trust;
all others must
bring data."

-WILLIAM EDWARDS DEMING



ALUMNI SUBMISSIONS

"The roots of our journey begin here."

— Anonymous

The Alumni Submission section celebrates the journeys of our alumni who once walked the same corridors and now carry the institution's spirit into the wider world. Through their stories, achievements, and reflections, this section reconnects past and present while inspiring current students to dream bigger and aim higher. Their experiences remind us that every beginning here leads to limitless possibilities beyond.

"Where we start shapes where we go."

Navigating the Hidden Drift of our Algorithmic Reality

Mr. Keshav Jhunjunwala
M.Sc. Computer Science (2022-2024),
IEMA Research & Development Private Limited

“ The challenge of concept drift serves as a reminder that it is impossible to make perfect predictions in an ever-changing world; however, our aim is to maintain the calibration and integrity of our model while still improving. ”

The traditional machine learning paradigm is based on a very simple assumption. That the statistical properties of any data-generating process remain constant over time. This concept is known as stationarity. In a lab environment, it is easy to maintain this and thus the algorithms can have optimal performance on a static test set. But this becomes more difficult as the model moves from the lab to the real world. The very nature of our reality is non-stationary. There are multiple features working together to form a single data-generating process. And a slight change in any single feature can snowball into a huge data distribution shift. That this shift has occurred is not yet a known area of expertise, but rather has now become a crucial part of the plan for the modern organization. The difference, as it pertains to other forms of decay, is that our data will appear to be normal, but its original or intended meaning will have changed radically. A change in the relationship between your features and target can be mathematically formulated as a change in the conditional distribution $P(Y|X)$. In simple terms, a change in the nature of this relationship is called a transformation. To create a complete taxonomy of transformations, or decays, we must also distinguish between decays caused by a covariate shift or a prior probability shift. Covariate shifts happen when there is a change in the input data-set used by our algorithm, but the rules or truths defined by the input data-set itself remain the same. Prior probability shifts happen when there is a change in the probability of the target variable, even though the nature of the target variable itself remains the same.

Of the available failures, the proxy trap is one of the deadliest. A mathematical label is employed as an easy substitute for a human condition. A popular healthcare research study exposed how an algorithm employed to determine high-risk patients for care management created a racial bias against Blacks. There was a massive gap when the cost of care was employed as a proxy for the need for care. This was a very well-calibrated algorithm from a technical perspective, as it accurately predicted costs for all races. However, because of the systemic inequities that exist around access to care, we have traditionally spent less money on care for Blacks than Whites with the same condition, resulting in a determination that Blacks were in better health than they actually are. In reality, on average, there was a 26.3% greater number of Blacks than Whites who were at the same risk score at the time of analysis. If this label bias had been corrected, the number of Black enrollees in the programs would have increased from 17.7% to 46.5%.

Similarly, the Zillow Offers failure illustrates the cause-and-effect relationship of concept drift. The pricing algorithm of Zillow was developed in a stable setting, but with the onset of COVID-19, it created enough instability in that market for the pricing algorithm to be unable to forecast property prices accurately. Instead of focusing on hyper growth, if they had made sure they were pricing their properties correctly, they would not have overpaid for properties by such a large margin, resulting in \$500 million in losses.

The velocity of change plays an important role in determining how quickly we must respond and develop a solution. Concept drift is not always a linear process; instead, there are temporal patterns that are related to drift that help to determine whether our model needs a temporary blip adjustment or it needs a completely new structure. Drift can be sudden, incremental, or recurring in nature. Some drifts are temporary, while others are permanent changes to how a business is conducted. The ability to recognize and understand these patterns will be what makes or breaks our data strategy in an organization.

There is a multi-tiered radar system required to reduce the time it takes to detect a model drift: the time between detection and response to a shift in the model. The first tier consists of statistical metrics that assist us in detecting shifts, such as KL and JS divergences. In addition to these statistical metrics, there are hypothesis tests such as Kolmogorov-Smirnov tests that offer a greater sensitivity to detecting small changes in distributions. Moreover, we might use process-control techniques like CUSUM and Page-Hinkley tests to continuously monitor the presence of shifts or signals in the data over time. We can also use adaptive techniques such as DDM, EDDM and ADWIN for tracking shifts. This enables us to dynamically change the size or shape of the data window based on how the error patterns change over time. Finally, we can use decision tree-based techniques for monitoring the drift in the temporal patterns of data when the temporal dimension is the most informative feature. On the other end of the spectrum from these approaches, we might also employ Information Geometric Temporal (IGT) projections to visualize the flow of information along a statistical manifold to assist us in recognizing the presence of change well before we have noticed it within a model's performance.

To overcome non-stationarity, the transition from training once to the continuous process of machine learning needs to happen through evolving machine learning. Evolving machine learning involves modelling algorithms that learn and improve their performance on a continuous basis by adapting to new data. Moreover, to facilitate this process, the approach of online ensemble methods like Dynamic Weighted Majority can be adopted. Online ensemble methods will develop a model by employing a large number of base-learners and will introduce or delete a base-learner based on its performance at that time. While adopting these methods, the maintenance of computation resources can be achieved by employing surrogate domains and reduction techniques that minimize complexity while preserving the bounds on dynamic regret. Dynamically Changing Deep Neural Networks (DeDNN) and AutoGrow are methods for automatically learning how deep neural networks are supposed to change over time. This change occurs by adding hidden layers or neurons. The addition is based on the presence of too many errors in the form of predictions for some outcome. It can also be based on changes in the structure of the model as a result of JS Divergence differences between the outputs of

successive models. The final examples in this exhaustive list are methods known as Elastic Weight Consolidation (EWC) and Synaptic Intelligence (SI). They attempt to maintain the long-term memory of the model by penalizing changes to weights that were used in support of previous tasks as they are discovered to be significant to the previous task.

As our reality continues its journey through time, it is evident that the current role of AI in the modern age is not static but dynamic accuracy. In order for a system to be deemed trustworthy, it must be able to quantify its epistemic uncertainty and adjust its decision boundaries as the joint probability distribution in our world changes (or evolves). A sophisticated system will actually be dangerous if it fails to recognize that it is no longer up-to-date. And in assessing our level of knowledge with machine learning technology, we must consider more than just worrying about what information is on our dashboards.

The question is: Are our systems built to learn from evolution, or are they merely waiting for a dynamic world to break them?



MR. KESHAV JHUNJHUNWALA

An alumnus of the 2022-2024 M.Sc. Computer Science program, Mr. Keshav Jhunjunwala serves as the Head of Artificial Intelligence at IEMA Research & Development Pvt. Ltd., spearheading innovation in advanced AI solutions. A Member of the Council on IT & Communication at the Merchants' Chamber of Commerce & Industry and a Ph.D. scholar in AI, he brings expertise in Robotic Process Automation, ETL pipelining, offline AI systems, and enterprise-scale Generative AI, with research presented at esteemed national and international forums.



The Illusion of Data, Uncertainty & The Role of Interpretation

Ms. Bratasree Bhuiya
M.Sc. Statistics (2022-2024),
Research Manager, Kantar IMRB

I still remember sitting in one of my statistics classes during my Master's programme, watching equations fill the board while our professor spoke about assumptions and error terms. At the time, statistics felt precise, contained, almost comforting. There were rules to follow, models to fit and clear ways to measure uncertainty. If something went wrong, you could usually trace it back to a violated assumption or a flawed dataset. If a result was statistically significant, I believed it deserved my confidence. If a model fit well, I assumed it was telling the truth. That faith felt justified, numbers were clean, assumptions were stated and uncertainty was something you could neatly calculate. Reality, it seemed, could be managed, if not fully understood, through numbers.

That belief followed me when I stepped out of academia and into the world of analytics and market research. The first time I saw data influence an actual business decision, I realized something had changed. The numbers were not just explaining the world anymore; they were quietly deciding it. A trend in a dashboard turned into a strategy. A segmentation model reshaped how consumers were seen. An algorithm, built to assist, suddenly held more authority than the people interpreting it. Only then did I realize how much messier data becomes once it leaves the classroom and starts influencing real decisions. The numbers no longer lived in isolation. They travelled into meeting rooms, presentations, dashboards and strategies. They started shaping narratives, priorities and sometimes entire directions for businesses.

Today, working as a research professional across multimarket studies, I spend much of my time translating raw data into insights that others will act on, something decision-ready. Consumer responses turn into trends; trends turn into insights and insights quietly turn into action. Automation helps scripts run analyses faster, dashboards validate data, algorithms segment audiences, but speed has a way of making conclusions feel more certain than they really are.

Somewhere along the way, I began noticing a subtle shift. The question was no longer 'what does the data say?' but 'what decision does this data justify?'. When a model produces a clean output or a statistically significant result, it carries weight. It feels objective. It feels final.

But the truth is, every dataset already contains a series of choices. What I have learned is that the most powerful moment in any analysis happens before the data even exists. It happens when we decide what to measure, what we exclude, how we phrase survey questions, how we define a KPI (key performance index) and these decisions quietly shape the version of reality we later claim to "discover". In theory, algorithms are neutral. In practice, they inherit every assumption we give them. In brand tracking, I see how small framing choices can ripple through an entire dataset. A slightly different question wording, a missing response option, or a cultural nuance

across markets can change outcomes in ways no regression model can fully fix. Yet once the data is collected, the numbers appear neutral, as if they emerged without context.

People are complex, messy and inconsistent; data needs structure. When we reduce complex human behaviour into clusters or scores, we gain clarity, but we also lose texture. People become segments. Experiences become averages. Outliers are examined briefly and then set aside. This simplification is necessary to make sense of scale, but it also creates blind spots that algorithms cannot see. This makes analysis possible but it also flattens reality.

There have been moments in my work where a model pointed confidently in one direction and still something did not sit right with me. The model worked. The results were significant. The data showed a clear pattern. But something about the story felt incomplete. Those moments are uncomfortable, especially in a world that values speed and clarity. Questioning a clean result can feel like slowing things down or being difficult. But I have come to believe that discomfort is often a signal that something important is being overlooked. They remind me that statistics can describe patterns, but it cannot explain meaning on its own, that is our story to tell.

The more automated our systems become, the easier it is to forget this. Dashboards update automatically. Workflows run silently. Insights appear polished and ready to present. The more seamless the system, the easier it becomes to forget that uncertainty still exists. When decisions need to be made quickly, there is rarely space to ask what the data might not be capturing.

At the same time, I have also seen the best version of data-driven work. When analysts, researchers and stakeholders treat insights as starting points rather than conclusions, something shifts. Conversations open up. Context is added. Human judgment re-enters the process to interpret data responsibly. Teaching research protocols or reviewing analyses often leads to deeper discussions about why we do things a certain way, not just how. These conversations matter. They shape how we, as analysts will handle uncertainty and ethical responsibility.

Looking back, I understand now that statistics was never about discovering a single, objective truth. It was always about approximation. About negotiating reality through models that are certain, but never complete. Algorithms may help us see faster and farther, but they do not absolve us of responsibility, they do not remove the need for judgment. If anything, they make that judgment more important.

As data, statistics and algorithms continue to shape decisions across the modern world, the most important skill may not be technology alone; the real challenge is not building smarter models, but staying aware of their limits. Awareness of what numbers reveal, what they conceal and how easily they can turn from tools into decisions. Decoding reality, after all, is not just a mathematical task. It is not about finding the perfect number. It is about remembering that behind every dataset is a set of human choices and that those choices still matter.

In the end, AI will not take your job, unless you stop questioning it, trusting it blindly and let it think for you.



MS. BRATASREE BHUIYA

Ms. Bratasree Bhuiya is a Research Manager at Kantar IMRB, amplifying consumer insights through advanced statistical analysis, machine learning models and Python-driven automation. Her work focuses on transforming complex datasets into clear, decision-ready insights that guide strategic market research. Alumnus of the 2022-2024 M.Sc. Statistics batch, her strong analytical foundation continues to power her data-driven approach to understanding consumer behaviour.



"Better an approximate answer to the right question than an exact answer to the wrong one."

-JOHN WILDER TUKEY

The Geometry of AI: Latent Topology and Explainability

Mr. Arnav Das

M.Sc. Computer Science (2022-2024)

Research Fellow, India Internet Foundation

In the pursuit of General Intelligence, we often obsess over the architecture, the depth, the number of parameters, the mechanism of attention. Yet, the true “mind” of a machine does not live in the weights; it lives in the empty spaces between them.

We will explore the topology of latent representational space, i.e. the high-dimensional manifold where data is compressed into meaning. By understanding the structure of this space, we can uncover new frontiers for Explainable AI. One that moves beyond simple feature highlighting to understanding the shape of machine reasoning.

1. The Manifold Hypothesis

To understand latent space, we must first accept the Manifold Hypothesis, where real-world high dimensional data (like images or text) lies on a lower-dimensional manifold embedded within that high-dimensional space. When a Neural Network “learns”, it is essentially learning a continuous mapping from the chaotic input space to a structured latent space [1].

1.1 The Representation, The Space & The Structure

Let \mathcal{X} be the input space (e.g., pixel space) and \mathcal{Z} be the latent space. We define an encoder $E: \mathcal{X} \rightarrow \mathcal{Z}$ such that a data point x is mapped to a latent vector z , i.e. $z = E(x; \theta)$. Here, z is a coordinate on a semantic map. In a well-trained network, this representation is disentangled, meaning specific dimensions of z correspond to specific generative factors of the world (e.g., z_1 controls object rotation, z_2 controls color).

The Formulation of the encoder E dictates how this space is constructed. It is determined by the objective function and the constraints (priors) imposed during training. For example, a Variational Autoencoder’s formulation forces the latent space to approximate a Gaussian distribution using the KL divergence:

$$\mathcal{L}_{VAE} = E_{\{q(z|x)\}}[\log p(x|z)] - D_{KL}(q(z|x) \parallel p(z))$$

This formulation forces “smoothness”. It ensures that if z represents a cat and $z + \epsilon$ represents a slightly different cat, the space between them is valid. Without this regularization, the latent space would be a sparse memorized points with no semantic connectivity.

The latent space is rarely Euclidean and often curved. We can define a Riemannian metric tensor $G(z)$ that characterizes local distances. The distance between two points z_a and z_b along

a curve γ is given by:

$$L(\gamma) = \int_0^1 \sqrt{\{\gamma'(t)^T G(\gamma(t)) \gamma'(t)\}} dt$$

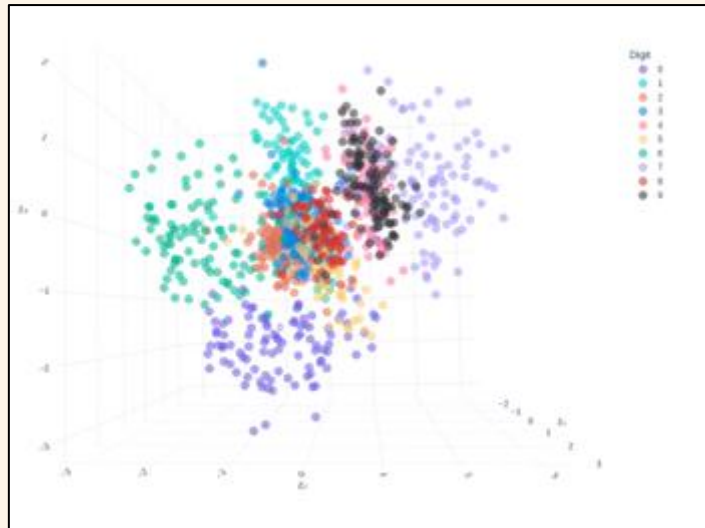


Fig. 1: Intermediate Representation Space of a VAE trained on MNIST Dataset

In this structure, the “shortest path” between two concepts is not a straight line, but a curve that follows the manifold. This explains why linear interpolation in pixel space creates ghosting, while geodesic interpolation in latent space within generative AI, smoothly morphs a frown into a smile [2].

Figure 1 illustrates a three-dimensional latent representation learned by a VAE trained on the MNIST dataset. A similar 2D variation of the above illustration is available at tayden.github.io [3]. Each point corresponds to a single image and colours indicate the ground truth digit label. Instead of being randomly scattered, the points form coherent clusters corresponding to digit classes. However, these clusters are not perfectly separated. Instead, they partially overlap and transition smoothly into one another. This structure reflects the continuous nature of the learned manifold that similar digits (for example, “3” and “5”, or “4” and “9”) appear in nearby regions of the space, while more distinct digits occupy more distant regions. The overall geometry demonstrates the effect of the Gaussian prior, which encourages the latent codes to occupy a dense, continuous region rather than isolated points, thereby enabling smooth semantic interpolation between different digit classes.

Note: If we can map the topology of the latent space, we can explain the model’s conceptual boundaries.

2. Topology as Explanation

Traditional XAI methods (like SHAP or LIME) are local and only tells us which pixels matter for one prediction. Topological Data Analysis (TDA) can identify holes or voids in the latent space (measured by Betti numbers). A hole represents a region where the model has no semantic understanding. It is a “blind spot” where adversarial attacks often thrive [4][5]. Identifying these topological voids explains where the model is unreliable.

Regions of high curvature in the latent manifold often correspond to complex decision boundaries or phase transitions in concepts (e.g., the exact point where a “wolf” becomes a “dog” in the model’s eyes). By measuring the curvature, we can detect which concepts the

model finds difficult to distinguish. If the topology is flat and orthogonal (Euclidean), the axes of the space become interpretable. If the topology is twisted, these concepts are entangled. Therefore, enforcing specific topological constraints (like in β -VAEs) is a direct path to interpretability.

3. From Black Box to Explainability

By rigorously defining the Representation $E(x)$, analysing the Formulation, its loss landscape and mapping the Structure (Riemannian metric), we gain the ability to navigate the thoughts of a machine. Explainability is no longer just about what the AI saw but where the thought travelled to get there.

References:

- [1] VISxAI 2018: Towards an Interpretable Latent Space. <https://tspinner.de/towards-an-interpretable-latent-space/>
- [2] Latent space visualization- deep learning bits — HackerNoon. <https://hackernoon.com/latent-space-visualization-deep-learning-bits-2-bd09a46920df>
- [3] VAE Latent Space Visualization. <https://tayden.github.io/VAE-Latent-Space-Explorer/>
- [4] Anna Kuzina, Max Welling and Jakub M. Tomczak. Alleviating adversarial attacks on variational autoencoders with mcmc. <https://arxiv.org/abs/2203.09940>
- [5] Chethan Krishnamurthy Ramanaik, Arjun Roy and Eirini Ntoutsi. Adversarial robustness of vaes across intersectional subgroups. <https://arxiv.org/abs/2407.03864>



MR. ARNAV DAS

Working at the intersection of Machine Learning and Network Systems, Mr. Arnav Das is an alumnus of the 2022–2024 M.Sc. Computer Science cohort. His research spans mechanistic interpretability, explainable AI, and autonomous systems. As a member of the Cohere Labs Community and a Research Fellow at IIFON, he contributes to strengthening India's internet infrastructure, with published work in Springer Nature and IET Quantum and active engagement in IETF and ICANN.

Is Intelligence Just Pattern Recognition?

The Architecture of Thought: From Autoregression to DeepSeek-R1

Mr. Dipmalya Roy

M.Sc. Statistics (2023-2025)

Junior Data Analyst, Digital Data Consultancy

For decades, intelligence was treated as something qualitatively distinct from statistics. Reasoning required logic. Understanding required symbols. Thought required structure. And yet, the most powerful reasoning machines ever built, emerged not from formal logic systems but from scaled-up next-word prediction. So, we are forced to ask:

Is intelligence just pattern recognition at scale?

I. The Statistical Engine Beneath “Thought”

Strip away the mystique and every large language model optimizes a single objective:

$$\max \log P(x_t | x_1, \dots, x_{t-1})$$

Maximum likelihood estimation! That is the entire foundation.

Language becomes a stochastic process. Conversation becomes conditional probability. Thought becomes autoregression. But scale alters interpretation.

Modern systems are trained on **15+ trillion tokens** extracted from tens of billions of web pages. To a statistician, the internet is not knowledge; it is a noisy, biased, heteroskedastic sample of human expression.

Training, therefore, is not just ingestion. It is **curation**.

Filtering, de-duplication and quality scoring act as variance reduction techniques. Without them, the estimator would internalize spam, duplication and synthetic artifacts. With them, the model approximates a compressed statistical portrait of human discourse.

Fifteen trillion tokens. Compressed into weights occupying tens of terabytes. Civilizations reduced to parameters.

II. The Representation Problem: What the Model Cannot See

Before probability comes representation. Language must be discretized into tokens, typically via Byte Pair Encoding (BPE). Words become integer IDs. Meaning becomes indices. This abstraction is powerful and limiting.

Let us indulge in a thought experiment. Ask the model:

How many “.” are there in “.....” ?

How many “r”s are in “strawberry”?

It may fail. Why? Because it may not “see” letters — only tokens. If “strawberry” is encoded as a single ID, letter-counting is not structurally visible.

To a human, this task is trivial.

To the model, it may require decompression it was never optimized for.

This is **jagged intelligence**:

- Superhuman abstraction in some domains.
- Elementary symbolic errors in others.

The limitation is not lack of scale. It is the geometry of representation.

III. From Autocomplete to Assistant

A base model is not a thinker. It is a distribution. Left alone, it continues text in whatever statistical direction the internet suggests. Post-training reshapes this behaviour. Through supervised fine-tuning, the model’s probability mass is shifted toward helpful, structured, conversational outputs. Dialogue markers become structural signals. Alignment becomes conditional reweighting.

But hallucinations remain. Because the objective never changed. The model optimizes likelihood not the truth.

If a false continuation is statistically probable, it may be generated confidently. That is not deception. It is the mathematics functioning as designed.

IV. The Turning Point: When Prediction Became Strategy

Before language models reasoned in public, other systems had already crossed a psychological boundary. In 2016, AlphaGo defeated Lee Sedol — one of the greatest Go players in history. Go was not chess. It was considered resistant to brute-force search. Its combinatorial space exceeded the number of atoms in the observable universe. Human intuition was thought indispensable. AlphaGo did not understand beauty. It did not feel tension. It optimized policy and value networks trained through self-play and reinforcement learning. And it produced Move 37 — a play so unexpected that professional commentators initially called it a mistake. It was not. It was strategically superior. That was the first crack.

Then came AlphaFold. Protein folding had resisted biologists for half a century. The mapping from amino acid sequence to three-dimensional structure is a high-dimensional inference problem of staggering complexity. It defines how life physically manifests. AlphaFold approached it as prediction. Given enough data, enough architecture, enough optimization — structure emerged from statistical learning. It did not “understand” biology. It predicted structure. And it outperformed decades of domain-specific modelling.

The message was subtle but destabilizing: **Prediction, when scaled and optimized, begins to look like reasoning.**

The deeper shift occurred with reinforcement learning. In domains like mathematics and programming, answers are verifiable. A proof is correct or it is not. Code compiles or it fails. This changes the learning signal. Instead of merely imitating human reasoning, models can explore solution paths and receive reward only for correctness. They begin to optimize for outcome, not appearance. Modern reasoning systems — including DeepSeek-R1 — generate intermediate reasoning sequences before producing an answer. Each “thinking token” is additional computation. More tokens allow exploration of a broader probability landscape. Statistically, these intermediate steps behave like latent variables refining a posterior belief over the final output.

The system is not conscious. But it is searching.

V. DeepSeek-R1 and the Compute Myth

DeepSeek-R1 disrupted an industry assumption: that frontier reasoning requires astronomical budgets. By combining:

- A structured cold-start phase,
- Targeted reinforcement learning,
- And hardware efficiency,

it achieved competitive reasoning performance at a fraction of the cost of proprietary systems.

The implication is subtle but significant. Raw scale matters. But training signal quality and how probability is shaped may matter more. The frontier is no longer defined purely by parameter count. It is defined by optimization dynamics.

The Question We Cannot Avoid

Here is the uncomfortable possibility:

What if reasoning is not a separate faculty at all? What if abstraction, planning, strategy and even self-correction emerge naturally from sufficiently expressive probabilistic systems trained at scale?

Human cognition is built on pattern recognition: visual, linguistic, social. Neurons fire in response to statistical regularities. Beliefs update with evidence. Experience reshapes weights.

And now we have machines that hallucinate, miscount letters and stumble on arithmetic — yet solve competition mathematics, write production-grade software, master strategic games and uncover solution paths not explicitly present in their training data.

All from optimizing conditional probability.

No symbolic rulebook. No handcrafted reasoning engine. No explicit module called “logic.” Just gradient descent navigating an astronomical statistical landscape.

If a neural network trained on text begins to display structured reasoning, is that a failure of our definition of intelligence or confirmation of it?

The boundary between statistics and thought may be thinner than we imagined. The architecture of intelligence may not be symbolic. It may be probabilistic.

And if that is true, then we are not witnessing the automation of language.

We are witnessing the first blueprint of synthetic cognition — assembled one conditional probability at a time.

If reasoning can emerge from pattern recognition at scale, then perhaps intelligence was statistical all along and we are only now beginning to measure it.



MR. DIPMALYA ROY

Mr. Dipmalya Roy works as a Junior Data Analyst at Digital Data Consultancy, where he explores datasets to uncover patterns to support informed decision-making. As an alumnus of the 2023–2025 M.Sc. Statistics cohort, he has cultivated strong expertise in statistical analysis and quantitative methods, and is steadily advancing in the domain of data analytics.

From Observation to Calculation: How Numbers Now Shape Reality

Ms. Afsana Zeba Yasmin
M.Sc. Computer Science (2023-2025)
Developer, National Council of Science Museums

Reality in the modern world is increasingly experienced through numbers. What we see online, how public resources are distributed, these identity verifications are linked to welfare schemes such as the Public Distribution System, pensions and LPG subsidies. Government reports claim savings of over ₹90,000 crore by eliminating duplicate or fraudulent beneficiaries using statistical de-duplication techniques. At the same time, independent studies have documented exclusion errors in certain regions, ranging from 2 to 12 percent, caused by biometric mismatch, poor connectivity, or data inconsistencies. Statistically efficient systems thus redefine eligibility and access. Who is recognized by the system becomes who exists within it. Statistics here do not merely measure beneficiaries. They determine them.

Predictive Analytics and Risk Classification in Indian Cities: Predictive analytics extends statistical reasoning into the future. Rather than describing what has happened, these systems estimate what is likely to happen next. Several Indian police departments use crime mapping and analytics platforms. Delhi Police's Crime Mapping Analytics and Predictive System (CMAPS) analyses FIR records, time patterns, geographical clustering and historical crime trends to guide patrol deployment. Similar hotspot analysis tools are used in Kerala, Telangana and Maharashtra.

These systems rely heavily on historical data. Areas with higher reporting rates or policing presence generate more recorded incidents. The algorithm interprets this as higher risk, directing more resources to the same areas. Increased surveillance leads to more recorded data, reinforcing the original statistical pattern. This is a closed loop. The model's prediction influences the data that future predictions rely on. Such systems improve efficiency, but they also raise questions about fairness and representation. Areas with lower reporting or limited access to formal policing may appear statistically safer, even if underlying conditions are similar. Risk, in these systems, is not discovered. It is calculated.

Political Communication and Algorithmic Fragmentation: Political campaigns in India increasingly rely on data analytics and algorithmic segmentation. Voter rolls, census data, booth-level turnout history, caste composition, religion, language and social media behavior are used to cluster voters into target groups. Instead of a single national message, campaigns distribute customized narratives. Messages are tested for engagement and circulated through volunteer networks and messaging platforms. During recent election cycles, WhatsApp-based political communication reached hundreds of millions of users, often through private groups beyond public visibility.

Religion, caste, or cultural identity in this context functions as a statistical variable rather than a belief system. Algorithms optimize message delivery based on response probability. Different groups receive different versions of political reality, each tailored to resonate emotionally. The result is a fragmented public sphere. Citizens no longer encounter a shared set of facts, but parallel algorithmically curated narratives.

Financial Algorithms and the Indian Credit Landscape: India's expanding formal credit ecosystem provides another clear example of how statistics shape lived reality. Credit scores generated by agencies such as CIBIL, Experian India, Equifax India are derived from repayment history, credit utilization, loan diversity and inquiry frequency. These scores are statistical estimates of default risk. A lower score leads to higher interest rates or outright denial of credit. Higher borrowing costs increase financial strain, raising the probability of default and confirming the model's original assessment.

With more than 300 million Indians now included in formal credit systems, these statistical classifications influence access to housing loans, education loans and small business financing. Creditworthiness becomes a numerical identity that shapes future opportunities. The system is mathematically consistent, yet socially consequential.

Data, Health and Resource Allocation: Healthcare systems increasingly rely on statistical risk models to prioritize care. During the COVID-19 pandemic, India used dashboards tracking positivity rates, testing volumes, hospital occupancy and district-level trends to guide lockdowns and reopening strategies. These indicators shaped public perception of risk. Yet variations in testing intensity across states meant that identical positivity rates could reflect very different ground realities. Policy decisions followed statistical signals rather than lived experience.

This highlights a broader issue. Statistical indicators simplify complex conditions into manageable numbers. While necessary for large-scale governance, these simplifications also shape what is considered urgent, visible, or actionable. Numbers guide attention. Attention guides resources.

The Question of Objectivity: Statistics and algorithms are often perceived as neutral because they rely on numbers. This perception is misleading. Every dataset reflects choices about what to measure and what to ignore. Every algorithm reflects assumptions embedded in its design.

If historical data reflects inequality, the models trained on it will reproduce those patterns. Efficiency does not guarantee fairness. Accuracy does not ensure justice. Recognizing this does not require rejecting data-driven systems. It requires understanding their construction, limitations and consequences.

Decoding Reality in a Data-Saturated Society: Decoding reality in the modern world means learning to look beyond outputs and ask deeper questions. What data was used. Whose behavior was recorded. Which variables were prioritized. Who benefits from the statistical framing. India's scale makes these questions especially urgent. As data infrastructures expand across governance, finance, healthcare and politics, statistical reasoning increasingly defines

how reality is categorized and acted upon. Statistical literacy and algorithmic awareness are no longer technical skills. They are civic necessities.

In conclusion, statistics and algorithms have become central to how reality is perceived, organized and governed. They shape what information is visible, how risk is defined, who is recognized by systems and how futures are predicted. In India, these processes operate at an immense scale, influencing everyday life in ways that are often invisible. Decoding reality requires understanding not just what numbers say, but how they are produced and how they are used. In a world increasingly filtered through data, the explanation behind the number matters as much as the number itself.



MS. AFSANA ZEBA YASMIN

An alumna of the M.Sc. Computer Science(2023-2025) program, Ms. Afsana Zeba Yasmin is a Developer at the National Council of Science Museums, engaged in advancing technology-driven initiatives. With a solid grounding in programming and computational systems, she blends technical expertise with a passion for innovation to make impactful contributions in the tech space.

Birds of A Feather Stream Together

A Small Glimpse into How Your Next Movie Is Recommended

Ms. Ashmita Bhattacharjee
M.Sc. Computer Science (2023-2025)
IT Executive, ITC Infotech

When you open a streaming app after a long day, you probably don't scroll through the entire catalog like it's a supermarket aisle. Instead, the app seems to already know what you might want to watch. A row quietly appears at the top: "Recommended for You." Sometimes it is uncannily accurate, almost as if the app has been reading your mind, or at least your watch history.

The thought takes me back to a recommendation systems lecture in my final semester, where a modest example, almost in passing, left a lasting impression. While the discussion focused on the mathematics behind recommendation systems, the idea itself felt surprisingly intuitive: behind every suggestion is a small algorithm quietly connecting patterns, preferences and probabilities.

At a high level, most recommendation systems follow one of two approaches. The first is *collaborative filtering*, which recommends items to a user based on the preferences of similar users. If people with tastes similar to yours loved a particular movie, the system assumes you might like it too. The second approach is *content-based filtering*, where the system looks not at other users, but at the content of the items themselves. It studies the features of movies you liked in the past and recommends new ones with similar characteristics

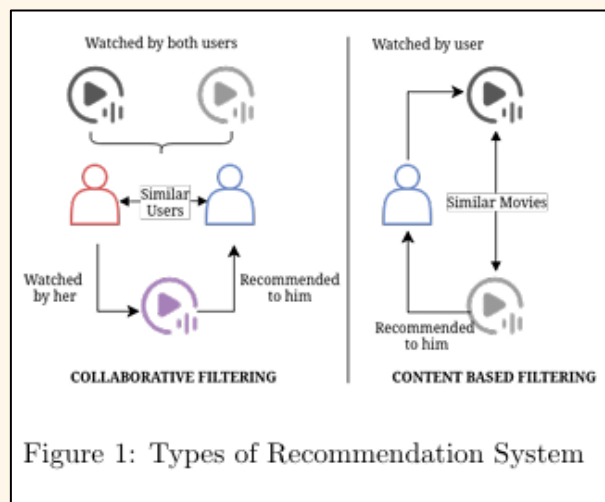


Figure 1: Types of Recommendation System

Collaborative filtering can be extremely powerful when large amounts of user data are available. However, it struggles with the Cold Start problem, a situation where a new user or a new item has very few ratings. Content-based systems, on the other hand, rely on item features rather than user communities, making them especially useful when data is sparse or when personalized, explainable recommendations are desired. To see how such a system works, consider a simple scenario.

A newly launched online movie platform currently has a small group of active users who have rated a few movies. The platform now plans to expand its catalog by adding new titles and wants to recommend them to its existing users. Since the platform is still new, there is not enough user-to-user interaction data to apply collaborative filtering. Instead, the platform has access to content information about each movie, such as its genre (for example, Comedy or Action), along with the ratings users have already given.

Using this available information, the platform aims to build a content-based recommendation system that predicts how much each user might like the new movies, based on the genres of films they have previously rated. The following example illustrates how this recommendation can be computed step by step.

How a Content Based Recommender Works

Assume the platform classifies movies using two genres: *Comedy* and *Action*. Based on viewing traces, the system estimates each user's preference for these genres on a scale from 0 to 4.

Table 1: User Preference Matrix			Table 2: Movie Feature Matrix		
User	Comedy	Action	Movie	Comedy	Action
A	3	0	M1	1	3
B	2	2	M2	0	4
C	4	4	M3	4	3
D	0	4	M4	4	0

Each row in Table 1 represents a user's taste as a vector and each row in Table 2 represents a movie as a feature vector. To predict how much a user will like a movie, the system computes the dot product between the user vector and the movie vector.

For example, prediction for User A on Movie M2:

$$A = (3, 0), \quad M2 = (0, 4)$$

$$Prediction = (3 \times 0) + (0 \times 4) = 0$$

Repeating this for all users and movies gives the below prediction matrix.

User	M1	M2	M3	M4
A	3	0	12	3
B	8	8	14	6
C	16	16	28	12
D	12	16	12	8

 \Rightarrow

User	M1	M2	M3	M4
A	0.38	0.0	1.5	0.38
B	1.0	1.0	1.75	0.75
C	2.0	2.0	3.5	1.5
D	1.5	2.0	1.5	1.0

After normalization, the system can now make concrete recommendations. Consider **User A**. From the normalized matrix, the predicted scores for the unseen movies are highest for *M3*, with a score of 1.5. This indicates that this movie best matches User A's preference profile, which is strongly influenced by comedy-related features.

Therefore, the system would recommend **M3** to User A. Similarly, the highest-scoring unseen movies for each user become their personalized recommendations.

What began as a few ratings and simple genre labels has now turned into individualized suggestions for each viewer. No human curator manually chose these movies; the choices emerged from patterns in the data. This small example reveals a larger truth: behind every “Recommended for You” row lies a quiet blend of statistics and algorithms, continuously shaping what we watch, read and discover in the modern world.

References:

- Zeynep Beyza Ayman, Recommendation Systems: Content-Based Filtering, Medium, Dec. 2022. Available: <https://medium.com/@zbeyza/recommendation-systems-content-based-filtering-e19e3b0a309e>



MS. ASHMITA BHATTACHARJEE

Ms. Ashmita Bhattacharjee is an IT Executive at ITC Infotech, contributing to the dynamic landscape of information technology with skill and dedication. As an alumna of the 2023–2025 M.Sc. Computer Science cohort, she built a strong foundation in computing and problem-solving. Combining academic excellence with professional expertise, she continues to grow in the rapidly evolving world of technology.

From Coffee Cups to Corporate Clues

Ms. Shreya Nandy

M.Sc. Statistics (2023-2025)

Junior Data Analyst, Digital Data Consultancy

Once heard from my professor that the subject we carry forward as a legacy, Statistics somehow finds a way to solve all kinds of problems we face in life. At that moment, I may not have been ready to grasp its true scope, but as one enters professional life, the hurdles become clearer. I was fortunate enough to experience a smooth transition from educational life to professional life. Yet, life inevitably goes through many changes that leave you perplexed at times. When I finally sat down to solve each problem, I realized that the support of that subject stood firm, building a strong pillar that helped answer every question. Alongside this, there was another constant—coffee. From late-night sips of raw coffee straight from sachets during exam preparations to refreshing coffee breaks in the corporate world, it has always been there.

A day at the professional desk:

You work hard all your life grinding, pushing yourself to bring out the best in you, all with the hope of one day having an office desk decorated with your favourite things and posh gadgets. Behind that desk lies not just comfort, but years of discipline, patience and belief in oneself. I still clearly remember the email I received from our Dean Sir on the second last day of our practical exams. Just after the exam ended, I took my phone out of my bag to inform my family that it had gone well and there it was a mail stating that we had been selected. I shouted in the exam hall, leaving everyone around completely confused. That was where the journey began and now it has been eight months of working in the designation that, I believe, every Statistics student aspires to—Analyst.

A day behind the desk begins with accomplishing deadlines, but it also brings responsibilities, accountability and the urge to deliver accurate results. Each task demands focus, logical thinking and attention to detail skills that Statistics quietly prepares you for. Once you step into professional life, you truly understand the difference between working professionally and merely completing courses on your own. Here, learning is not optional; it is a necessity driven by real-world impact and decision-making.

I used to hear my father's friends, who work in the corporate world, talk about training their minds even in their 50s. Now I understand why learning in the corporate environment has no finish line. With the rapid pace of the digital world, tools, technologies and methodologies evolve constantly and staying relevant means continuously upgrading yourself.

Being part of live projects and working under cooperative seniors plays a crucial role in shaping confidence. Having mentors who encourage questions, welcome opinions, brainstorm ideas and guide you through mistakes creates an environment where growth feels natural. They help you understand not just what is right or wrong, but also why it matters.

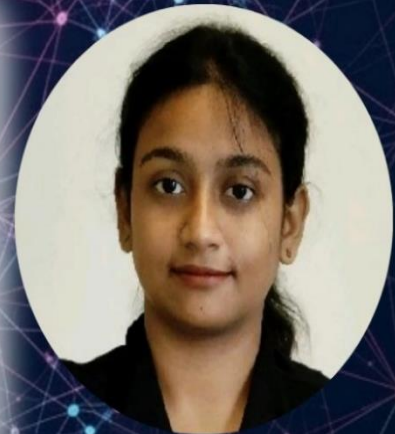
As I mentioned earlier, I was fortunate enough to experience a smooth transition, but beyond the designation and the desk, the real achievement lies in the mindset that develops one that embraces challenges, values learning and understands that growth is a lifelong process.

What began as a simple aim is now a designation which has now transformed into a deeper understanding that growth, learning and self-discovery are lifelong journeys. Shaped by challenges, guided by mentors and strengthened by real-world responsibilities, the focus has shifted from merely reaching a position to continuously evolving as a professional and as an individual. In this phase, work has taken on a deeper meaning echoing Rabindranath Tagore's words, "I slept and dreamt that life was joy. I awoke and saw that life was service. I acted and behold, service was joy." What once felt like ambition has now grown into purpose, making every day a step toward meaningful contribution.



MS. SHREYA NANDY

Ms. Shreya Nandy works as a Junior Data Analyst at Digital Data Consultancy, focusing on interpreting data and generating actionable insights. An alumna of the 2023–2025 M.Sc. Statistics batch, she developed strong analytical and problem-solving skills and is now forging a promising career in data analytics with curiosity and precision.



Data Storytelling: Unveiling Truths in a World of Information Overload

Ms. Ishika Dutta

M.Sc. Statistics (2023-2025)

Junior Data Analyst, Digital Data Consultancy

In today's digital age, we're overwhelmed with data from social media feeds, news cycles, government reports, and personal fitness trackers. Yet, raw numbers alone rarely lead to understanding. That's where data storytelling comes in. It turns complex datasets into compelling narratives that reveal hidden patterns, challenge assumptions, and guide decisions.

Why it matters?

- Clarity over chaos: A well-crafted visualization, like heat maps, interactive charts, or infographics, cuts through the noise and makes trends instantly recognizable.
- Emotional resonance: Numbers can feel abstract. When paired with relatable context, such as showing how air quality affects daily commuters, they become more impactful.
- Democratizing knowledge: Good stories allow anyone, not just analysts, to grasp insights, fostering informed public discussion.

How it works?

1. Curate the data – Choose reliable sources, like government open data and reputable surveys, then clean them for accuracy.
2. Find the narrative hook – Identify a question or tension, for example, “Why are traffic accidents rising in Kolkata despite new safety laws?”
3. Choose the visual medium – Use time-series line graphs for trends, choropleth maps for geography, and animated bubbles for multi-dimensional comparisons.
4. Layer context – Add annotations, anecdotes, or quotes that clarify why the pattern matters.
5. Invite interaction – Tools like Tableau and D3.js, or simple sliders, let readers explore “what-if” scenarios, turning passive viewers into active investigators.

Technique	What it Does	How It Powers the Narrative
Time-Series Analysis	Breaks down data into trend, seasonality, and residual components; fits models such as ARIMA or Prophet.	Shows <i>when</i> changes happen, highlights recurring cycles, and forecasts future values. This gives readers a clear sense of direction, timing, and urgency.
Regression Analysis	Quantifies relationships between a dependent variable and one or more	Transforms intuition into evidence: for example, “Every 10% rise in traffic volume increases accident risk by 3%,”

Technique	What it Does	How It Powers the Narrative
	predictors using models like linear or logistic regression.	making the story measurable, credible, and actionable.
Machine Learning (Clustering, Classification, Gradient-Boosted Trees)	Identifies hidden structures, groups similar observations, and predicts outcomes without assuming linear relationships.	Enables segment-specific insights such as “Cluster A (young, night-shift workers) faces 20% higher exposure to poor air quality,” and supports predictive narratives like future risk or threshold breaches.

By weaving these methods into the narrative, the story moves from description to explanation and prediction, giving readers both insight and foresight.

Real-world impact:

- **Public health:** During COVID-19, dashboards combined case counts with vaccination rates and used machine learning-based risk clusters to identify neighbourhoods needing urgent outreach.
- **Climate action:** Time-series of temperature anomalies, regression of CO₂ emissions on industrial output, and machine-learning-driven hotspot detection turned abstract climate models into urgent, shareable stories that motivated policymakers.
- **Business strategy:** Companies use regression to link customer churn to service delays while machine learning clustering identifies at-risk segments, leading to targeted retention campaigns.

Challenges to watch:

- **Misleading scales:** A shortened y-axis can exaggerate trends. Always label axes and cite sources.
- **Over-simplification:** Stripping away nuance for a cleaner story can hide important details. Include footnotes or toggle-able detail layers.
- **Bias in framing:** The same data can tell different stories depending on chosen baselines. Disclose methodological choices and model assumptions.

Data storytelling is more than just attractive visuals. It connects raw statistics to human understanding. By embedding statistical techniques into the narrative, we transform numbers into compelling evidence-based stories. These narratives reveal why things happen, when they're likely to change and who/what will be affected. In a world flooded with information, the ability to interpret and communicate data with statistical rigor may be the most valuable skill of all!



MS. ISHIKA DUTTA

Ms. Ishika Dutta is a Junior Data Analyst at Digital Data Consultancy, where she works with data to uncover patterns and support informed business decisions. An alumna of the M.Sc. Statistics program (2023–2025), she developed strong analytical and quantitative skills. With a keen eye for detail and a passion for data-driven insights, she is steadily shaping her career in data analytics.



"UNCERTAINTY IS
INEVITABLE, BUT
QUANTIFICATION IS THE
STATISTICIAN'S CRAFT."

- David Roxbee Cox

The Digital Ossuary: Calculating the Human Condition

Mrs. Dipanjana Kundu

M.Sc. Statistics (2023-2025)

Junior Data Analyst, Digital Data Consultancy

The modern world is no longer composed of atoms alone; it is a sprawling architecture of information, a vast "Digital Ossuary" where every human action, thought, and preference is catalogued, weighed, and stored. For centuries, humanity looked to the heavens or the depths of the soul to understand reality, but in the twenty-first century, we have outsourced that understanding to the twin engines of statistics and algorithms. This shift represents more than a technological evolution; it is a fundamental rewriting of the human experience. We have entered an era where reality is not merely lived, but decoded through a mathematical lens that treats the chaos of life as a series of solvable equations.

At the foundation of this new reality lies statistics, the ancient language of uncertainty repurposed for an age of infinite data. Statistics was once the tool of the census-taker and the tax-collector, a way for the state to view its citizens as a manageable collective. Today, however, statistics has moved from describing the world to predicting it. We live in a society governed by the "Average," a phantom inhabitant created by aggregating the behaviours of millions. This statistical ghost dictates the dosage of our medicines, the safety ratings of our cars, and the interest rates on our loans. While the individual is unique, the statistical model views the individual as a data point within a distribution curve. This creates a strange paradox of modern life: we are more tracked and identified than ever before, yet we are increasingly treated as mere probabilities in a vast sea of variance.

If statistics provides the map of our behaviour, algorithms are the invisible hands that guide us through it. An algorithm is, in its simplest form, a set of logical instructions, but when applied to the mountain of data generated by modern life, it becomes something far more potent. Algorithms have become the new gatekeepers of human knowledge and social interaction. Every time we engage with a digital interface, an algorithm is making a split-second decision about what we should see, who we should talk to, and what we should believe. This is the automation of logic, a process that removes the friction of choice but also narrows the horizon of our experience. By feeding us content that statistically aligns with our past behaviour, algorithms create "filter bubbles"—digital echoes of our own biases that make the world seem smaller and more certain than it truly is.

The synergy between these two forces has given birth to a "Prediction Economy." In this landscape, the future is no longer a mystery to be feared or a blank canvas to be painted; it is a commodity to be traded. Companies and governments use algorithmic modelling to anticipate our needs before we feel them. This "pre-emptive" reality is visible in everything from supply chains that ship products before they are ordered to predictive policing models that claim to identify crime before it occurs. However, this level of control comes with a profound ethical cost. When we allow algorithms to "decode" our future based on the statistics of our past, we risk trapping ourselves in a cycle of historical bias. If the data used to train these systems is

flawed by the prejudices of the past, the algorithm does not just reflect those prejudices—it optimizes and enshrines them as objective truth.

This brings us to the "Vanitas" of the digital age. Just as the classical painters used skulls, hourglasses, and rotting fruit to remind viewers of the fleeting nature of life and the inevitability of fate, our modern algorithms remind us of our own predictability. We are, in the eyes of the machine, a collection of habits and biological signals. There is a certain cold beauty in this—a sense that the world is finally being organized and understood. Yet, we must be careful not to mistake the data for the soul. A statistical model can tell us how many people will likely fall in love this year, and an algorithm can suggest the "perfect" match based on shared interests, but neither can capture the lightning-strike moment of human connection that defies the curve.

Ultimately, decoding reality through statistics and algorithms is an act of translation. We are translating the messy, vibrant, and often irrational human experience into a language that machines can process. This translation has allowed us to achieve miracles of efficiency and insight, but much like a poem translated into a different tongue, something is always lost in the process. The nuance of the individual, the spark of the unexpected, and the beauty of the outlier are often discarded as "noise." As we move forward into a world where our lives are increasingly managed by these invisible architects, our challenge is to remember that the map is not the territory. We must remain the masters of the code, ensuring that while we use these tools to understand our world, we do not let them define the limits of our potential. Our reality may be decoded, but it should never be predetermined.



MRS. DIPANJANA KUNDU

Mrs. Dipanjana Kundu serves as a Junior Data Analyst at Digital Data Consultancy, contributing to analytical projects that transform raw data into meaningful information. An alumnus of the M.Sc. Statistics program (2023–2025), she brings a strong statistical foundation to the field of data analysis. Her growing expertise reflects a commitment to applying statistical thinking to real-world challenges.





STUDENT SUBMISSIONS

“The future belongs to those who believe in the beauty of their dreams.”

— Eleanor Roosevelt

The Student Submission section highlights the voices, ideas, and talents of students who bring fresh perspectives to every page. Through articles, reflections and creative contributions, this section provides a platform for students to express their thoughts and share their unique insights. It celebrates curiosity, passion and the courage to turn ideas into meaningful expression.

“Every great journey begins with a student’s idea.”

The NSSO of India and the Legacy of P.C. Mahalanobis

Mr. Arpan Banerjee

B.A. (Hons.) Economics, Sem - IV

Mr. Anubhav Roy

M.Sc. Statistics, Sem - II

Statistics and algorithms are essential tools in modern society based on data that drive policies, business, science and politics. Statistical reasoning and algorithmic design are the basis of decisions that impact millions of people and these are used in mobile recommendation engines and even in national economic planning. The key element in this change is the sampling theory, which is a branch of statistics that provides the possibility of proper inference on the basis of few observations. One of the most influential authors whose efforts fundamentally changed the current state of sampling is Professor Prasanta Chandra Mahalanobis; his presence is still felt in the National Sample Survey Organisation (NSSO) in India.

An innovative statistician, Professor P. C. Mahalanobis, argued that it was possible to collect vast empirical information with high accuracy and efficacy, through scientific sampling instead of applying the tedious method of enumeration. Long preceding the creation of the concept of data science in popular culture, Mahalanobis advocated the use of probability sampling techniques that would save money and time without sacrificing accuracy. His works laid the groundwork of theories and methods of practical work in national surveys on the international level.

Largely his most lasting contribution was the establishment of the National Sample Survey (NSS) in India in 1950 which was later established as the NSSO. Under the strong encouragement of the Prime Minister, Jawaharlal Nehru, Mahalanobis saw the need to have a systematic and comprehensive socio-economic data that included employment, consumption, health and agriculture of the vast and diverse population of India. As the MOSPI Golden Jubilee report emphasizes, the main concepts of Mahalanobis (especially about the conception of the sample and how to conduct the survey) are the new values of the NSS and the pillars of the so-called NSS culture.

Before modern computational algorithms existed, Mahalanobis proposed the stratified multistage sampling, probability proportional to size selection and inter-, sub- and intra sample interpenetration methods of quality control and error estimation. These methodological advances were future programs in computational statistics. In practice, they allowed the NSSO to develop surveys that were representativeness and feasible, which is still the key concern of modern data-collection systems.

As an example, in the early versions of the NSS, sample designs were stratified, i.e. the population could be split into clusters with similar geographical and demographic characteristics and then the primary units, e.g. villages, were chosen based on the probability of being included in the sample, where key variables such as population or area were taken into

consideration. This method ensured the sampling of the diversity of the India population in a representative and unbiased way. These methods keep on shaping current sampling applications and algorithm models underlying big data analytics.

One more important aspect of Mahalanobis that could be considered critical is his priority on error estimation. Mahalanobis did not assume that a sample is a perfect mirror of a population, but came up with strategies to measure sampling and non-sampling errors. His subsamples made independent (now known as Inter-Penetrating Network of Subsamples (IPNS)) allowed analysts to compute variability and to identify inconsistencies in data collection. This emphasis on measurement of error early on heralded modern statistical machine learning methods of validation of algorithms and confidence estimation.

The activity of the NSSO has far-reaching influence on society. Its surveys give consumer spending data which is used to determine the poverty lines, employment data applied in the labour-market policy and agricultural data applied in the food and rural development policies. These surveys affect national policy, budgetary allocations and research in the field of economics, in the field of public health and research in the field of social sciences. Such decisions would be based on guesses and not facts with no concrete sampling design.

In a wider setting of the contemporary world, the Mahalanobis legacy depicts how statistical essence and algorithmic mentalities can bring changes to societal self-realization. Modern algorithms, including predictive models in artificial intelligence and optimisation algorithms in analytics, are based on the statistical principles that have been developed in the last century. The idea that the data could be gathered objectively and a mathematical analysis of it with precision is one that Mahalanobis at least had when he imagined the future of data gathering and analysis of the scientific community and it is echoed in all the fields that utilise data in some way or other, such as economics, artificial intelligence, etc.

To sum up, the modern world is defined not only by the computational power but by the statistical principles of the data collection, interpretation and use. The work of P.C. Mahalanobis, particularly via the NSSO, still attests to the transformative nature of statistics and the designing of algorithms in bringing policy to bear and improving the wellbeing of humanity.



Gibberish to Us, Language to Them

Mr. Projato Chakraborty

M.Sc. Computer Science, Sem – IV
St. Xavier's College (Autonomous), Kolkata

Sound is quite ambiguous and not at all binary. That sounds like a pretty straightforward statement but internally there is something very unsettling about this. To us humans – it could be just a beep with no inherent meaning, but a machine may interpret this as a whole sentence. Short bursts of random noise – sharp, patterned, grating, almost irritating, can easily be dismissed by a human ear as just interference. Yet to a machine, that could be a complete sentence. This intersection between perception and interpretation is where the idea of *Gibberlink* began and where the reflection lies. If you have ever ignored a strange sound because it did not seem important, then this is figuratively the same. The difference here is that the listener is not human and it does not get to ignore it so easily.

Gibberlink is not a product, nor a finished system. It is an idea that emerged from thinking about how machines, or rather, AI agents, communicate when humans are not meant to be a part of the conversation. The signals are designed to be gibberish to us and thus human intervention is not necessary – abstraction does the work for us. The premise is simple: two conversational AI agents begin communicating – one or both agents detect that the other is also an AI. They propose switching to *Gibberlink* for faster, cleaner machine-to-machine communication, often using something like **gg wave** – a simple sound-based encoding scheme that turns small chunks of data into short, modulated audio pulses. These same pulses, while obviously meaningless – even chaotic – to human listeners, remain interpretable to other machines which share the same decoding logic.

Pondering about this raises a crucial question – what does it mean to “*decode*” something in the first place? Is decoding about understanding – or just about knowing when to act?

Human communication almost always revolves around clarity. We expect words to be recognised, sentences to be grasped and context to be deduced virtually instinctively. Machine communication involving sound often does not work that way. The same sense of decoding is implied but it is inherently uncertain. For a machine, receiving signals is less about certainty and more about likelihood – sound waves degrade, microphones vary in sensitivity and clarity, environments introduce noise. There is no capital-T 'Truth.' Only a **confidence threshold** – a cold, calculated guess on whether a sound is a signal or just garbage.

That is where statistics comes in – not as equations, but as a way of thinking. A signal is not simply *received* or *missed*. The machine does not ask, “Is this message true?” It asks, “How likely is this to be a valid message?” That shift is subtle, but it changes everything. Accordingly, 'likelihood' isn't just a wild guess either. It is more like picking up on little hints – a pause here, a higher pitch there – always weighing the evidence. What this really implies is that decoding happens on a canvas of grey. It is a rule, not a feeling. Below that threshold, the safest choice is to treat the signal as noise; above it, to accept and respond. In between lies this uncomfortable zone, where the machine might hesitate or retry.

Once you see decoding this way, the situation looks less like “sound carrying meaning” and more like two systems negotiating uncertainty. The message is not bulletproof; it survives

only because both sides agree on what counts as a message in the first place. Without this agreement of shared rules, even a perfectly clear signal is just noise. When multiple agents are operating in the same environment, speed and coordination matter far more than clarity. A message does not need to be expressive or rich; it needs to be reliable enough to trigger the right response. Their version of understanding is statistical and pragmatic – and often, that is enough.

Reflecting on *Gibberlink* also reveals something quietly haunting: increasingly, important decisions are now being made through channels we simply are not wired to perceive. We have all had that experience – standing in a hallway, hearing the muffled rhythms of a conversation behind a heavy door. You can hear the tone and energy, but the meaning is barred from you. Algorithms are now in that room, whispering in short, modulated pulses that bypass our ears entirely. They negotiate and sync their 'thresholds for action' without ever needing a human to nod in approval. Once they agree on how to act, the process becomes invisible, almost alien, to us. We are just left in the hallway, shaped by systems that do not need us to listen, only to trust that whatever happened behind that door was right.

But trust is a fragile thing when decoding is probabilistic. Noise can be mistaken for intent. A signal can be spoofed. A false positive can trigger action when nothing meaningful was ever sent and a missed signal can silence something that should have been heard. None of this requires malice; sometimes it is just the world being messy. That is why the ethics of something like *Gibberlink* is not only about what it enables, but about what it normalises – communication that is efficient – yes, but invisible and mostly uninterpretable to humans.

I keep returning to the same small picture – a beep in the air. On its own, it is nothing. But if meaning can be packed into it, extracted from it and acted upon, then it stops being “nothing.” It becomes part of reality. Not the reality we perceive, but the reality that gets decoded and then executed.

Maybe that is the real discomfort behind the idea. Not that machines can talk in strange ways, but that the modern world is increasingly shaped by conversations we do not participate in, carried by signals we do not notice, interpreted by thresholds we no longer participate in setting. The beep is still just a beep to us – but somewhere else, it might already be a decision.



Cyber Security Trends: Protecting Information in the Digital Age

Mr. Krishna Kanth Mallick
M.Sc. Computer Science, Sem - II

With the fast evolution of digital technology, the issue of cybersecurity has become one of the most important concerns in the modern world. The increasing use of the internet, mobile devices, cloud computing and online services has made everyday life easier and more convenient. But at the same time, it has created new opportunities for cybercriminals to exploit and steal sensitive information. With the rising number of cyber threats, new trends in cybersecurity are emerging to safeguard digital data and infrastructure.

One of the most pressing trends in cybersecurity today is the **rise in ransomware attacks**. Ransomware is a type of malicious software that locks or encrypts a user's data and demands payment to unlock it. Over past few years, these attacks have evolved from targeting individuals to targeting large organizations, including hospitals, educational institutions and government bodies. These attacks can lead to disruptions in critical services and cause financial losses. To combat ransomware attacks, organizations are focusing on regular data backups, enhanced endpoint security and incident response strategies.

For example, hospitals have become frequent targets of ransomware attacks. In many cases, attackers encrypt patient records and hospital systems, forcing administrators to either pay the ransom or risk disrupting critical medical services. Such incidents highlight the importance of regular data backups and strong security systems.

Another important trend is the integration of **Artificial Intelligence (AI)** and **Machine Learning (ML)** in cybersecurity. Cyber attackers are using AI tools to automate attacks, identifying system vulnerabilities and bypass conventional security systems. Simultaneously, cybersecurity professionals are using AI to detect threats more efficiently. AI-based security systems can analyse large volumes of data, identify unusual behaviour as well as respond to threats in real time. This helps organizations detect cyberattacks at an early stage and minimize potential damage. Consider email security systems which use machine learning to identify phishing emails. They analyse patterns such as suspicious links; fake sender addresses and unusual language. These systems can block malicious emails before they reach users, reducing the risk of phishing attacks.

The rapid adoption of **cloud computing** has also changed the cybersecurity landscape. Many organizations now store data and run applications on cloud platforms because of their flexibility and cost efficiency. However, cloud environments introduce new security challenges such as misconfigured storage, unauthorized access and data breaches. As a result, cloud security has become a major trend, with organizations implementing encryption, access control and continuous monitoring to protect cloud-based resources.

In recent times, several data breaches have occurred due to publicly exposed cloud storage buckets containing sensitive customer information. To prevent this, organizations now use encryption, multi-factor authentication and cloud security monitoring tools to protect cloud environments.

Another growing concern is the **security of Internet of Things (IoT) devices**. IoT devices include smart home appliances, wearable gadgets, medical devices and industrial sensors. These devices are often connected to the internet but lack strong security features. Weak passwords, outdated software and poor design make them easy targets for cyberattacks. Cybersecurity trends now emphasize securing IoT networks by using device authentication, regular updates and isolating IoT devices from critical systems. In smart homes, attackers can exploit weak passwords in smart cameras or routers to gain unauthorized access, leading to privacy breaches. This has increased the demand for secure device configuration, firmware updates and network isolation.

The **Zero Trust security model** is also gaining popularity as a modern cybersecurity approach. Traditional security systems assume that users inside a network can be trusted. However, this assumption is no longer valid, especially with remote work and mobile access. The Zero Trust model follows the principle of “never trust, always verify”. Every user and device must be authenticated before accessing resources, regardless of their location. This model helps reduce insider threats and unauthorized access.

For example, in case of remote work environments, employees are required to access company systems from home/public networks or personal devices. Zero Trust systems require identity verification and device authentication before granting access, reducing the risk of unauthorized entry.

Despite technological advancements, **human error** remains one of the biggest cybersecurity risks. Many cyberattacks succeed because users fall victim to phishing emails, fake websites, or social engineering techniques. Therefore, cybersecurity awareness and training have become essential tools to fight cybercrime. In present times, it has become mandatory for organisations to conduct regular training sessions, awareness programs and simulated attacks to educate users about safe online practices and help them recognize potential threats.

Often, phishing emails pretending to be bank alerts or job offers, often trick users into revealing passwords. To counter this, organizations conduct phishing simulations and awareness programs to teach users how to identify suspicious messages.

Finally, **data privacy and cybersecurity regulations** play a crucial role in shaping modern security practices. Governments around the world have introduced strict laws to protect personal and organizational data. Compliance with these regulations has become mandatory, encouraging organizations to improve their security policies and maintain transparency in data handling. Strong cybersecurity practices not only ensure legal compliance but also build trust among users and customers.

After the introduction of such regulations, companies that collect customer data must now clearly state how data is used and stored. Failure to comply with data protection laws can result

in heavy fines and loss of customer trust, encouraging organizations to strengthen cybersecurity policies.

In conclusion, cybersecurity is no longer optional—it is a necessity in the digital era. As technology evolves, cyber threats will continue to become more advanced. Staying updated with emerging cybersecurity trends and adopting proactive security measures is essential for individuals and organizations alike. By combining advanced technology, strong policies and user awareness, it is possible to create a safer and more secure digital environment.



"Essentially, all models are wrong, but some are useful."

-GEORGE E. P. BOX



Use of Statistics in Business and Industry

Mr. Soumyadeep Biswas

B.Sc.(Hons.) Statistics & Data Science, Sem – II

Statistics has become an indispensable tool in modern business and industrial management, particularly in the area of production control. In today's competitive and dynamic business environment, decision-making based on intuition alone is no longer sufficient. Business executives are increasingly relying on statistical techniques to study consumer needs, preferences and behaviour, as well as to plan, control and evaluate business operations. The success or failure of a business enterprise largely depends upon the accuracy and precision of its statistical analysis and forecasting. Faulty expectations arising from incorrect or incomplete statistical data may lead to poor decisions, resulting in heavy losses or even the failure of the business.

Production planning is one of the most important areas where statistics plays a vital role. Suppose a businessman plans to manufacture readymade garments. Before initiating the production process, he must have a comprehensive idea of various quantitative and qualitative aspects. These include the number of garments to be produced, the amount of raw materials required, the availability and cost of labour and the specific features of the product such as quality, size, shape, colour and design. All these decisions cannot be made without reliable numerical data collected and analysed systematically. Statistical techniques help in estimating demand, forecasting sales and determining the most economical scale of production. Thus, the formulation of a sound production plan in advance is essential and such planning is impossible without the use of statistics.

As a result, most large industrial and commercial enterprises today employ trained and efficient statisticians. These professionals assist management by collecting data, analysing trends, interpreting results and presenting them in a form useful for decision-making. Statistics acts as a guiding force that enables managers to minimize risks and uncertainties associated with production and marketing activities.

Statistical techniques are widely used in various functional areas of business organisations. One important application is in Time and Motion Studies, which form a part of scientific management. By analysing the time taken for each operation and the movements involved in performing tasks, management can improve efficiency, reduce wastage of time and effort and enhance productivity. Statistical analysis helps in setting standard times and evaluating worker performance.

In marketing decisions, statistics plays a crucial role through demand analysis and consumer preference studies. By collecting and analysing data related to consumer tastes, income levels, buying habits and market trends, businesses can design products that meet consumer needs more effectively. Pricing strategies, advertising campaigns and distribution policies are also formulated based on statistical findings.

In the field of investment, statistical analysis assists businessmen and investors in making informed decisions. A sound study of individual shares, debentures and other financial instruments involves analysing past performance, risk, return and market trends. Statistical tools such as averages, index numbers and probability analysis help in selecting profitable and relatively safe investment options.

Statistics is also extensively used in personnel administration. Data relating to wages, cost of living, incentive schemes, labour turnover, absenteeism and productivity are analysed statistically to frame fair wage policies and effective incentive plans. Moreover, statistics helps in studying the effects of labour disputes or unrest on production and in setting performance standards for employees.

Another significant area is credit policy. Businesses rely on statistical data regarding customers' creditworthiness, payment habits and default risks to frame suitable credit terms. This helps in maintaining a balance between increased sales and minimized bad debts.

In inventory control, statistics assists in coordinating production and sales. Techniques such as Economic Order Quantity (EOQ) and stock-level analysis help in maintaining optimum inventory levels, thereby reducing storage costs and avoiding shortages or overstocking.

Statistics also contributes to accounting, particularly in the valuation of assets, depreciation and financial analysis. Statistical methods aid in interpreting financial statements and evaluating the overall financial health of business concerns.

In sales control, statistical data from market surveys, consumer studies, trade channel analysis and readership surveys provide valuable insights. These help management in monitoring sales performance, identifying problem areas and improving sales strategies.

Beyond general business applications, statistics has been of immense importance in life insurance, one of the earliest fields to adopt statistical methods. Statistical tools of probability and mathematical expectation are used to calculate premiums, assess risks and determine policy benefits. Mortality tables and actuarial statistics form the foundation of life insurance operations.

In industrial settings, statistics is widely applied in quality control. Statistical Quality Control (SQC) techniques such as control charts and inspection plans are used to ensure that the production process remains under control and that products conform to prescribed specifications. Since inspecting every item produced is neither practical nor economical, statistical sampling methods are employed. Sampling, a vital aspect of statistics, enables industries to maintain quality standards at minimum cost.

In conclusion, statistics is an essential tool in production control and overall business management. It provides a scientific basis for planning, decision-making, control and evaluation. In an era of intense competition and rapid change, the intelligent use of statistical techniques has become a necessity rather than a choice for business success.



Why Statistics Matters, Even If You Hate Numbers

Ms. Faiqah

B.Sc.(Hons.) Statistics & Data Science, Sem - II

Most people think statistics is only for mathematicians or people who enjoy calculating for fun. But the truth is simple: statistics is already shaping our decisions, preferences and opportunities, even if one never opens a textbook. Statistics is behind the decisions we don't see. Everyday apps use statistics every second, such as:

- Uber: When prices increase at night or during rain, that is surge pricing based on probability models that estimate demand.
- Netflix: When it suggests a thriller after one has watched a mystery movie, the algorithm behind uses data patterns from millions of viewers.
- Instagram Reels: When a certain type of video keeps showing up, the algorithm has statistically concluded what one is likely to watch next.

These decisions might appear random, but they are not so in the least! These recommendations are not personal, they're statistical predictions. But why should one bother about this at all? Because understanding how the basics of such systems, aids us in recognising what affects us and how to use it for ourselves.

A skincare brand says "95% of users saw results in 2 weeks."

A person without statistical thinking simply believes it. But someone who thinks statistically asks:

- *How many people were tested? (95% of 20 people means nothing)*
- *Were they all of the same age, skin type or underlying conditions(for example, skin diseases)?*
- *What does the term 'results' even mean?*

The moment one stops and tries to answer such questions; one realises the illusions one is surrounded by. This stops one from being misled by attractive numbers.

A store says "Flat 50% OFF", but increases the original price before discounting it. A quick check of the price history (Amazon shows this) reveals the truth. This is statistical thinking looking for trends before believing the final number.

When the weather app says "40% chance of rain today", many people panic. But statistically, it means there is a 40% chance of rain, not a guarantee. Knowing this helps you decide whether carrying an umbrella is necessary.

Statistics is not about formulas; it's about choosing wisely. And we can use it to make better decisions every day! Data also reveals personal patterns. Consider tracking your expenses for

a month. You might notice that 40% of your money goes to food deliveries when all this time you were thinking that you were merely spending a meagre amount on snacking. Your “small” purchases add up to more than big ones! Nothing complicated, just counting. But suddenly, you understand your habits more clearly.

Let me reveal a secret to you: ALL of us already use statistics for the smallest of tasks, mostly without realising it. For example, when we choose the shortest queue in a supermarket (we are estimating expected waiting time), when we decide to leave early because “traffic is usually heavy at this hour” (we have learned from past experience and are taking actions based on it), when we trust a product with 4.7 star rating after 10,000 reviews over one with 5 stars and 3 reviews (we are weighing sample sizes and comparing the proportions of success).

That’s statistical thinking in action during ordinary everyday tasks! But how can we use it consciously?

- Look for patterns: If your internet is slow every evening, that is a data pattern.
- Ask for numbers: Never accept “most people do this.” Always ask “how many?”
- Think in probabilities: Don’t stress over uncertain events, just judge the likelihood of an event.
- Track one small habit: Even 10 days of data can reveal the truth.
- Challenge perfect claims: Statistics teaches us that perfection rarely exists.

The message is simple. You don’t need to be a statistician. But if you learn to spot patterns, question numbers and think in chances, you can live smarter, make better decisions and avoid being manipulated by flashy claims.

Statistics isn’t about maths.

It’s about understanding reality so you can take control of your own life and live on your own terms!



How Rockets Break Free: The Science Behind Spaceflight

Mr. Dipanjan Ganguly

B.Sc.(Hons.) Statistics & Data Science, Sem - II

A rocket launch looks dramatic - fire, thunder and clouds of smoke. But beneath the spectacle lies a simple and powerful idea from physics: motion is created by momentum.

Rockets work using Newton's third law of motion: "*For every action, there is an equal and opposite reaction.*" When rocket engines burn fuel, they eject hot gases downward at extremely high speeds. In response, the rocket is pushed upward. Importantly, rockets do not push against air. They push against their own exhaust, which is why they can operate even in the vacuum of space.

To stay in orbit, a rocket must reach a speed of about 7.8 km/s. This is not about going high, but about going sideways fast enough. Orbiting is essentially a controlled fall around Earth. To escape Earth's gravity completely, a spacecraft must exceed the escape velocity of 11.2 km/s.

One of the biggest challenges in rocketry is efficiency. This is solved using multi-stage rockets. As fuel is burned, empty stages are discarded, reducing mass and increasing acceleration. This principle plays a major role in modern launch systems.

India's space agency, ISRO, has demonstrated remarkable mastery of these ideas. In 2023, India made history with Chandrayaan-3, successfully landing near the Moon's south pole, a region of high scientific interest due to possible water ice. ISRO's Gaganyaan mission, currently under development, aims to send Indian astronauts into space, relying heavily on real-time data, control algorithms and precision engineering.

Behind the scenes, rockets are guided not just by engines but by data, statistics and algorithms. Thousands of sensors continuously collect information on speed, position, temperature and fuel levels. Since raw data can contain errors or noise, statistical methods are used to estimate the rocket's true motion in real time.

Control algorithms then use this refined data to make instant decisions: adjusting engine thrust, changing direction and maintaining stability.

A real example is trajectory correction. As a rocket travels to space or toward the Moon, tiny forces such as gravity variations or solar pressure can slowly push it off course. On missions like Chandrayaan, onboard systems calculate the deviation using data and mathematical models, then fire small thrusters to correct the path. Even a small error early in flight could mean missing the target by thousands of kilometres.

Before launch, scientists also run millions of computer simulations using probability and past data to predict risks such as weather effects or engine performance. This statistical planning helps choose the safest launch windows and improve mission success rates.

In modern space missions, reaching space is not just about burning fuel; it is about measuring accurately, predicting intelligently and correcting continuously using mathematics and algorithms.



**“DREAM, DREAM, DREAM.
DREAMS TRANSFORM
INTO THOUGHTS AND
THOUGHTS RESULT IN
ACTION.”**

A. P. J. ABDUL KALAM

Job Health Mile Numbers Se

Ms. Ushosee Mukherjee
M.Sc. Statistics, Sem – IV

When we think about medical matters, the terms which usually knock our minds are medicines, stethoscopes, microscopes, hospitals, clinics, white coats etc. It is very rare for us to think about graphs, probability distributions, hypothesis, confidence intervals collaborating with the medical world. However, there is a planet in the solar system of science which is known as **Biostatistics**, a field that actually deals with the relationship between statistical data and human health.

Biostatistics is a route that takes us to the destination where health meets with the patterns hiding in the data. It transfigures symptoms into signals, treatments into trials and outcomes into probabilities. While doing so, it reshapes our understanding about life, disease, survival, death in this modern world.

We are all familiar with the doctors saying “*The patient has an 80% success rate*” or “*The patient has a high risk*”, so these are the guesses or assumptions made by the people with white coats. The conclusions are drawn from the large datasets, after carefully analysing the patterns in search of hidden biological touch. Every diagnosis today is, at its core, a statistical inference.

Medicine is a very strong companion of a diseased patient because the existence of this particular item helps the ailed figure to get back to his/her normal life. However, before providing a hand of help to the patient, this chemically organized product undergoes the rigorous world of clinical trials. So, we can say that Biostatistics acts as both judge and referee. Some drugs may fail to help the patient, clinically insignificant but statistically effective and some drugs are statistically ineffective but act as the life-saving product. This really tells us that numbers guide decisions but never tell the entire story.

In this modern world, where science is actually driving everybody insane with its advanced version, personalities with white coats and stethoscopes, with prescriptions in desks are assisted by algorithms diagnosing diseases, prediction of any kind of complications and giving a road map of how to prioritize care. The partnership of AI and Statistical tools, empowers the whole situation with its strong foundation.

Biostatistics is not about reducing humans to numbers, it actually helps us to use numbers to protect the lives of the humans. Every survival curve represents real people. Every data point carries a story.



When Numbers Judge: The Rise of Quantitative Jurisprudence

Mr. Siddhartha Das

B.Sc.(Hons.) Statistics & Data Science, Sem – II

The gavel strikes, a sound that for centuries has signalled the end of human deliberation. It is the sound of judgment rendered, a verdict reached through the careful weighing of evidence, principle and wisdom. But today, that solemn sound is increasingly accompanied by the silent hum of servers. In courtrooms from Wisconsin to New Delhi, judges are presented with algorithmic risk scores (numerical predictions about a defendant's likelihood of reoffending), predictive analytics (data-driven forecasts based on past patterns) and statistical portraits of defendants. This quiet collision between the qualitative art of judgment and the quantitative science of data forces upon us a profound and unsettling question: Can justice truly be measured?

At first glance, the question seems almost sacrilegious. Justice is a moral concept, embodied in stories of fairness, equity and context-specific mercy. It resides in the nuanced interpretation of a hesitant witness, the recognition of mitigating circumstances, the application of centuries-old principles to new dilemmas. To reduce this to a number or a probability feels like a fundamental category error, an attempt to capture a symphony with a single note.

Yet, the modern state is built on measurement. And law, as the scaffolding of the state, cannot remain an island of pure qualitative reasoning. The global experiment is already underway and it reveals not a single answer, but a spectrum of attempts to quantify the unquantifiable.

In India, the driver is sheer scale. With a backlog of over 50 million cases as of late 2025, the Supreme Court is piloting AI tools like SUVAS (an automated translation system for legal documents) and SUPACE (an AI case-assistance tool that helps judges sift through evidence). The measurement here is of throughput and consistency, can algorithms help clear the fog of delay that itself denies justice? The promise is immense: a system that can function. The peril is that in measuring efficiency, we lose sight of the individual narrative buried in the docket.

The Nordic approach, exemplified by Finland, offers a different model. Here, measurement is used not for prediction, but for transparency and self-awareness. Data analytics (systematic examination of large datasets to reveal trends) map legal patterns and logistics, but “black box algorithms” — systems whose inner workings are hidden and cannot be explained, are rejected. The measurement is of the system's own functioning, a tool for democratic accountability. It asks: are we, the humans in the system, being consistent and fair? This humble, audit-oriented use of data may be the most philosophically coherent.

The dangers of misplaced measurement are starkly visible in North America. Tools like the COMPAS algorithm (Correctional Offender Management Profiling for Alternative Sanctions, used to predict risk of reoffending) in the United States, claim to measure the risk of recidivism. Yet, as ProPublica's investigations have shown, they often end up measuring and perpetuating

the racial and socioeconomic biases encoded in their training data. In Canada and other parts of Europe, similar tools used in policing and parole have faced public backlash and judicial scrutiny. Here, measurement threatens to calcify injustice, giving historical prejudice the glossy veneer of mathematical neutrality.

Proponents counter that unmeasured justice is already riddled with bias and inconsistency. Data, at least, makes those flaws visible, opening the door to correction. This argument holds weight; without quantification, disparities in sentencing or bail often go unnoticed, hidden in the haze of human caprice.

This global patchwork is the messy birthplace of a new discipline: **quantitative jurisprudence**. By this, we mean the study of what happens when law and measurement intersect, combining philosophy, statistics and legal practice. Its pioneers are forging the frameworks we desperately need: standards for algorithmic transparency (rules requiring that AI decisions be explainable), protocols for bias auditing (systematic checks to detect unfairness in data) and legal doctrines for a defendant's "right to explanation" when a number influences their fate.

The possible future of quantitative jurisprudence is both promising and challenging. In the short term, it may focus on diagnostic measurement: mapping systemic inefficiencies, exposing hidden disparities and providing accountability dashboards for courts and governments. Over time, it could evolve into a norm-setting discipline, shaping international standards for how algorithms are allowed to participate in justice. Imagine a future where every AI tool used in sentencing or bail decisions must pass a "fairness audit," much like financial institutions undergo compliance checks.

At its most ambitious, quantitative jurisprudence could redefine the role of judges themselves. Judges may become interpreters of data as much as interpreters of law, trained to interrogate algorithmic reasoning alongside human testimony. Law schools might one day teach courses in "algorithmic cross-examination," preparing future jurists to challenge not only witnesses but also predictive models. In this vision, justice is not surrendered to machines but enriched by them, provided the human element remains sovereign.

So, *can justice be measured?* The evidence suggests a paradoxical answer: Yes, but only in parts and never in whole.

We can and must, measure the mechanics of justice. The backlogs, the demographic disparities in sentencing, the inconsistent application of bail. Quantitative jurisprudence will shine a light into these dark corners, forcing accountability upon systems. It will give us powerful, objective mirrors to hold up to our legal institutions.

But the core of justice: the final, moral act of judgment that balances mercy with accountability, that interprets the spirit of the law, that recognizes a unique human story, resists quantification. This must remain the sacred domain of human wisdom, of the judge who feels the weight of the gavel.

The future lies not in choosing between the measured and the moral, but in their proper dialogue: the court of tomorrow featuring "explainable AI" (systems whose logic can be

understood and challenged) as a witness, its reasoning cross-examined by defence and prosecution, with judges as both sages and sophisticated statisticians, interrogating datasets as keenly as testimonies. Justice may never be fully captured by a number. But the relentless, careful and ethically guided attempt to measure its contours may be precisely what helps us deliver it more faithfully. The goal is not a cold, calculated law, but a warmer, more consistent and more self-aware one. The measured gavel, understood not as an oracle but as a tool, may yet strike a truer note.



"The real challenge is not whether machines think but whether humans do."

FEI-FEI LI

How AI is Transforming the World

Mr. Soumyajit Roy

M.Sc. Computer Science, Sem - II

Artificial Intelligence (AI) is one of the most revolutionary technologies of the 21st century. It refers to the ability of machines and computer systems to perform tasks that normally require human intelligence, such as learning, reasoning, problem-solving, and decision-making. Over the past decade, AI has rapidly evolved and is now transforming nearly every aspect of modern life. One of the most significant impacts of AI can be seen in healthcare. AI-powered systems help doctors diagnose diseases more accurately and at an early stage. Machine learning models analyse medical images, detect patterns and predict health risks. Robotic surgeries, virtual health assistants, and personalized treatment plans are improving patient care and saving lives.

In the field of education, AI is making learning more personalized and accessible. Smart learning platforms analyse students' performance and adapt lessons according to their strengths and weaknesses. Virtual tutors, automated grading systems and AI-based language translation tools are helping students across the globe gain better educational opportunities.

AI is also transforming businesses and industries. Companies use AI for data analysis, customer service chatbots, fraud detection, supply chain management and market prediction. Automation powered by AI increases efficiency, reduces human error, and improves productivity. This allows businesses to make faster and smarter decisions.

Transportation is another area where AI is creating major change. Self-driving cars, intelligent traffic systems, and predictive maintenance technologies are improving road safety and reducing congestion. AI helps in optimizing routes, saving fuel, and lowering environmental impact.

In everyday life, AI is present in smartphones, social media platforms, voice assistants, and recommendation systems. From suggesting movies and songs to recognizing faces and voices, AI enhances user experience and convenience. However, along with its benefits, AI also brings challenges. Concerns about job displacement, data privacy, security risks, and ethical decision-making are important issues that society must address. Responsible development and proper regulations are necessary to ensure that AI benefits humanity as a whole.

In conclusion, Artificial Intelligence is reshaping the modern world in powerful ways. It is driving innovation, improving efficiency, and creating new opportunities across multiple sectors. As technology continues to advance, AI will play an even greater role in shaping the future of society. With responsible use and ethical development, AI has the potential to create a smarter, safer, and more connected world.



Environmental Sampling

Mr. Subhayu Mukherjee
M.Sc. Statistics, Sem – II

Environmental Statistics is a branch of Statistics concerned with collecting, analysing, interpreting and modelling data related to the environment. To understand its importance, we must first consider what is meant by “Environment.” The environment is a dynamic network of natural, social, cultural, technological and symbolic conditions that both shape and are shaped by living and non-living beings. Human life is inseparable from this system and maintaining a balanced relationship with it is both a necessity and a responsibility.

Statistical knowledge enables us to evaluate this relationship. It helps measure environmental quality, assess human impact and determine whether ecological conditions are improving or deteriorating. Without systematic measurement and analysis, environmental protection would rely on assumption rather than evidence.

A natural question arises: what do we measure, how do we measure it and why is measurement necessary? In theory, measuring every environmental component would provide complete knowledge. In practice, this is impossible. Environmental systems are vast, interconnected and constantly changing across space and time. Monitoring air quality at every location in a city, testing every stretch of river water, or tracking every species within an ecosystem would require limitless resources.

This limitation gives rise to one of the central concepts of Environmental Statistics: sampling. Sampling is not a mere compromise but a scientifically grounded strategy. Instead of attempting exhaustive measurement, we select representative units—specific locations, time points, or ecological entities—and use statistical methods to infer characteristics of the broader system. In this way, sampling becomes the bridge between environmental complexity and practical feasibility.

Environmental data also exhibit natural variability. Pollution levels fluctuate hourly, climatic patterns shift seasonally and ecological processes respond to numerous interacting factors. Well-designed sampling schemes help distinguish meaningful trends from random variation, allowing for structured and reliable interpretation.

Thus, in Environmental Statistics, sampling reflects both methodological rigor and an acknowledgment of observational limits. While the environment cannot be measured in its entirety, it can be understood, monitored and managed through systematic and statistically sound observation.

Environmental Data and Why Sampling is Necessary:

Environmental data encompasses information describing the state, behaviour and interactions of natural and human-influenced systems. It includes measurements of air and water quality, climatic variables, soil properties, biodiversity indicators and land-use patterns. Unlike

controlled laboratory data, environmental data is inherently complex—spatially dispersed, temporally dynamic and shaped by both natural variability and human activity. At first glance, one might assume that complete enumeration—measuring every unit within the system—would provide perfect environmental knowledge. In theory, observing every river segment, atmospheric layer, species and time point would eliminate uncertainty. Yet this vision collapses under practical and conceptual constraints. The environment is neither finite nor static. It evolves continuously across space and time. Rivers flow, weather patterns shift hourly and ecosystems respond unpredictably to disturbances. Achieving complete enumeration would demand unlimited resources, uninterrupted access and flawless instruments—conditions that are unattainable. Moreover, excessive monitoring can itself be intrusive, potentially disturbing ecosystems and altering the very processes under study. Beyond logistical barriers lies a deeper limitation: environmental systems are governed by inherent variability. Even if exhaustive measurement were momentarily possible, the information would become outdated almost immediately due to constant change. Thus, completeness in measurement does not guarantee completeness in understanding.

Sampling emerges as a scientifically grounded response to this reality. Rather than attempting to measure everything, researchers select representative subsets across space, time, or ecological categories. Through carefully designed sampling schemes, statistical methods enable reliable inference while explicitly accounting for uncertainty. Sampling therefore transforms limitation into methodological strength. By focusing on representative data rather than exhaustive measurement, environmental sampling encourages clarity of objectives, thoughtful design and analytical rigor. It allows meaningful patterns, trends and risks to be identified without being overwhelmed by impractical data demands. In this way, environmental data becomes not merely a collection of observations but a structured framework for understanding change. The impossibility of complete enumeration is not a weakness of environmental science; it is a defining characteristic. Environmental Statistics embraces this reality, working within constraints to provide robust knowledge for policy, conservation and sustainable management. In a rapidly changing world, intelligently designed sampling remains one of the most powerful tools for understanding and protecting the planet.

Why We Can Not Measure Everything?

The aspiration to measure the environment in its entirety is a natural scientific impulse. Yet the nature of environmental systems makes complete enumeration fundamentally unattainable. Several interconnected constraints explain why it remains an ideal rather than a practical reality. **Scale** presents the most immediate challenge. The environment spans vast and heterogeneous domains—forests, rivers, oceans, atmosphere, soil and urban systems—each operating at different spatial resolutions and interacting in complex ways. Measuring every unit across such expansive and interconnected systems exceeds any realistic observational capacity. **Cost and time** further limit exhaustive measurement. High-quality environmental monitoring requires advanced instruments, calibration, maintenance and sustained data collection. Uniform deployment across all locations would demand resources far beyond institutional or governmental limits. Moreover, attempting total measurement would delay analysis and decision-making. **Manpower, logistics and accessibility** also impose restrictions. Monitoring

remote, hazardous, or disaster-prone regions poses significant operational and ethical challenges. Finally, environmental conditions vary continuously over time. Even if complete measurement were possible at one moment, it would fail to capture ongoing change. Recognizing these constraints underscores a fundamental truth: **the impossibility of measuring everything necessitates sampling and statistical inference as rational, scientifically grounded approaches to environmental understanding.**

Types of Sampling Used in Environmental Studies Once the impossibility of complete enumeration is recognized, environmental science turns to sampling as a principled methodological response. Sampling is not a shortcut but a deliberate strategy to extract reliable knowledge from complex and dynamic systems. Because environmental conditions vary across space and time, different sampling methods are chosen according to the structure of the study area. Random sampling gives each unit an equal chance of selection, minimizing bias and enabling probabilistic inference. For example, researchers studying river water quality may randomly select locations along a stretch to avoid preconceived assumptions about pollution levels. Since environmental systems are rarely homogeneous, stratified sampling is often preferred. Here, the study area is divided into meaningful subgroups—such as industrial, residential and green zones—and samples are drawn from each to ensure balanced representation. Systematic sampling selects units at fixed spatial or temporal intervals, making it useful for monitoring trends, such as placing air quality stations at regular distances. When regions are extensive, cluster sampling reduces cost and effort by selecting representative areas for detailed study. Together, these methods allow scientists to measure environmental reality intelligently, balancing accuracy with practical feasibility.

Case Study: Understanding Environmental Reality Through Sampling in India:

Environmental challenges in India are marked by scale, diversity and intensity. From megacities struggling with air pollution to vast river systems sustaining millions of lives, environmental monitoring in India vividly demonstrates why complete enumeration is impossible—and why sampling becomes indispensable.

Urban Air Quality Monitoring in Indian Cities: Air pollution in Indian metropolitan cities such as Delhi, Kolkata and Mumbai is a persistent public health concern. Ideally, measuring air quality would require sensors at every street, household and industrial unit. Such complete coverage is clearly unattainable due to financial, logistical and infrastructural constraints.

Instead, air quality monitoring relies on a network of strategically placed sampling stations. These stations are not installed randomly across the city; rather, they are selected to represent different urban environments—traffic heavy intersections, industrial areas, residential neighbourhoods and comparatively cleaner green zones. Each station continuously samples ambient air and measures key pollutants such as $PM_{2.5}$, PM_{10} , nitrogen dioxide and ozone.

Although the number of monitoring stations is limited, the data collected from them allows scientists and policymakers to infer the broader air quality of the city. Statistical models and spatial interpolation techniques use these sampled observations to estimate pollution levels in

unmonitored areas. Thus, a small number of sampling points becomes sufficient to reveal city-wide pollution patterns, identify hotspots and assess health risks.

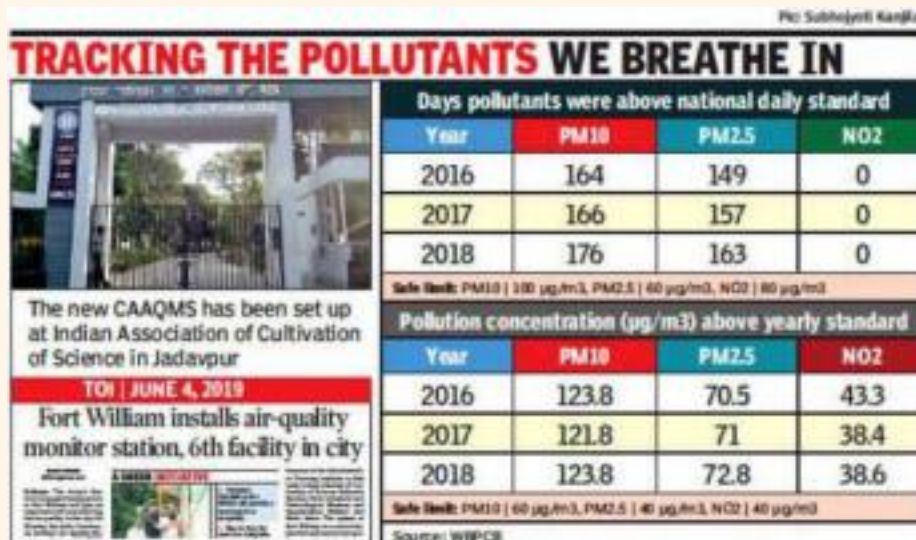


Fig. 5: Year wise AQI Analysis of Kolkata, 2019

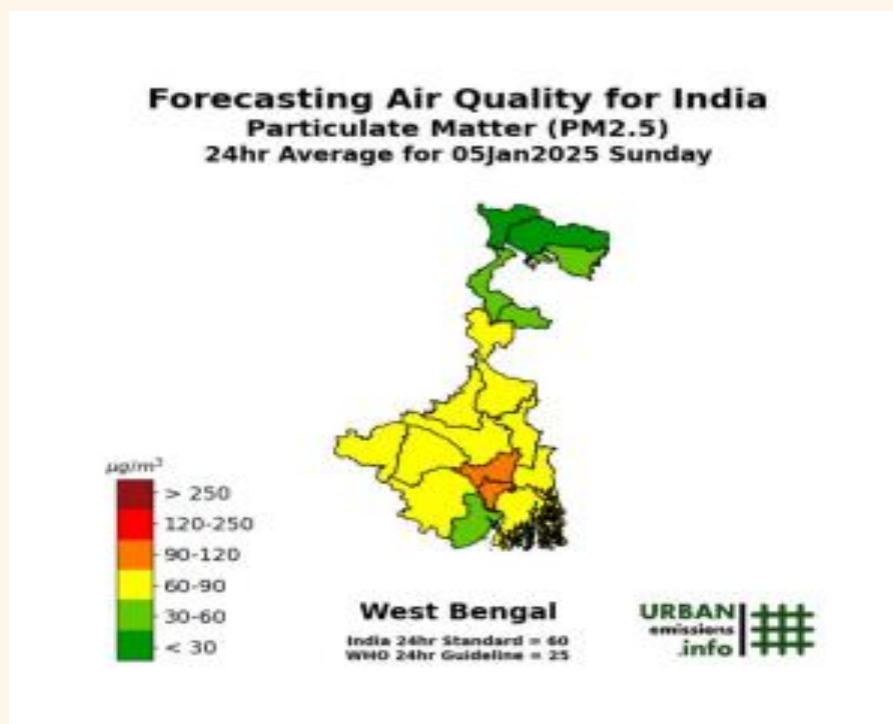


Fig. 6: Forecasting AQI of Kolkata as on 5th January 2025

Inference from Samples: Seeing the Whole Through the Part

The aforesaid Indian case study illustrates a central principle of Environmental Statistics: well-designed samples allow reliable understanding of systems far larger than what can be directly observed. Sampling does not attempt to capture environmental reality in its entirety; rather, it captures enough structure, variation and signal to support informed inference. In a country as ecologically diverse and densely populated as India, environmental sampling is not merely a

methodological choice—it is a necessity. Through sampling, environmental science bridges the gap between limited observation and expansive environmental responsibility.

Sampling Errors, Technology and Policy in Environmental Statistics:

Environmental sampling enables scientists to draw conclusions about complex ecological systems using limited observations. However, reliability depends on careful design and awareness of potential errors. Sampling error arises because a sample represents only part of the system; natural variability ensures estimates differ from true values. Measurement error stems from faulty instruments, calibration issues, or procedural inconsistencies. Selection bias occurs when certain conditions are systematically excluded, often due to convenience. Missing data—frequent in environmental monitoring—can further distort analysis, especially when failures occur during extreme events. Such flaws are not merely technical issues; they can misguide environmental regulation, resource allocation and public health decisions. Responsible sampling therefore demands statistical rigor and ethical transparency. Technological innovation has transformed environmental sampling. Sensor networks now provide continuous monitoring of air, water and climate variables, generating high-resolution data. Satellite and remote sensing technologies extend observation across large and inaccessible regions, while IoT-based systems enable real-time and adaptive monitoring. Yet technology does not eliminate uncertainty. Statistical calibration, validation and modelling remain essential to ensure accuracy and comparability. Technology enhances sampling, but statistics provides the framework that makes technological data meaningful and scientifically defensible. Sampling also lies at the heart of environmental policy. Indicators such as air quality indices, water classifications and climate risk assessments are derived from sampled data and generalized to broader populations. Sound sampling strengthens policy credibility; flawed sampling can delay intervention, mask environmental injustice, or misrepresent progress. Policymakers must therefore consider not only reported results but also how data were obtained. Robust sampling is equally critical for evaluating policy outcomes, ensuring that observed improvements reflect genuine environmental change. In environmental science, measuring everything is impossible. Yet through thoughtful sampling, technological integration and statistical reasoning, limited data can yield reliable knowledge. Environmental Statistics bridges observation and action, transforming uncertainty into informed decision-making. In a rapidly changing world, measuring wisely is essential for protecting ecosystems, public health and the planet's future.

Conclusion:

In environmental science, the ambition to measure everything is compelling but unrealistic. The scale and complexity of natural systems make complete enumeration impossible. Yet, well-designed sampling allows us to capture essential patterns and approach a reliable understanding of environmental reality. Through careful methodology, technological support and rigorous statistical analysis, limited observations can produce meaningful and trustworthy conclusions. Environmental statistics serves as a bridge between partial data and informed decision-making. By combining sound sampling design with innovation and analytical rigor, it transforms incomplete measurements into actionable knowledge. This is vital for guiding environmental

policy, conservation efforts and sustainable management. In a time of accelerating environmental change, we may not measure everything— but we can measure wisely. And in doing so, we strengthen our ability to protect ecosystems, safeguard public health and respond responsibly to the challenges facing our planet.



**“ASK THE RIGHT QUESTIONS AND
NATURE WILL REVEAL THE DOORS
TO HER SECRETS.”**

CHANDRASEKHARA VENKATA RAMAN

The Gamified Life — A Psychological and Ethical Interrogation

Mr. Sreesant Saha

B.Sc.(Hons.) Statistics & Data Science, Sem – II

From fitness apps that reward daily steps to language platforms that celebrate learning streaks, the architecture of modern life increasingly resembles a game. What began as a design strategy to enhance user engagement has evolved into a pervasive cultural logic—where points quantify effort, badges validate progress, and leaderboards rank identity itself. Gamification promises motivation, productivity, and measurable self-improvement. Yet beneath its colourful interface lies a deeper psychological and ethical terrain that demands scrutiny.

At its best, gamification draws upon well-established behavioural principles to catalyse habit formation and sustain engagement. At its worst, it risks reducing human aspiration to numerical targets, displacing intrinsic purpose with metric fixation, and embedding systems of subtle behavioural surveillance within everyday life. This dual character—empowering yet potentially manipulative—makes gamification one of the defining psychological technologies of our era.

A critical examination therefore becomes essential. While gamified systems can effectively shape short-term behaviour and even scaffold long-term growth, they may also undermine intrinsic motivation, promote compulsive engagement, and raise serious concerns regarding datafication and autonomy. The question is not whether gamification works—it clearly does—but rather how, for whom, and at what ethical cost.

I. When Life Becomes a Game: How Gamification can Boost Long-Term Motivation:

Gamification doesn't replace intrinsic motivation; it acts as a necessary scaffold to build habits until intrinsic motivation can develop. For individuals struggling to start a new behaviour (e.g.- language, exercise, learning), the external rewards(badges ,points) provide the initial "hook" and structure over time as competence and routine develop, the user begins to experience intrinsic rewards (the joy of movement the pleasure of understanding). The gamified elements can then fade into the background.

II. The Engine of Engagement: Core Psychological Mechanisms:

Social gamification features(team challenges leaderboards shared goals) tap into the need for relatedness. Long term motivation is often sustained by community and accountability like a fitness app connects us with friends for a step challenge, it transforms a solitary chore into a social commitment This social bond can provide motivation for beyond an individual badge.

In my opinion humans are social creatures. The desire to contribute to a team, maintain social standing, or simply share an experience with others is a powerful and durable motivator.

III. The Documented Efficiency:

The optimism surrounding gamification is not merely theoretical. In their landmark review, “*Does Gamification Work?*”, presented at the Hawaii International Conference on System Sciences, researchers Juho Hamari, Janna Koivisto and Harri Sarsa analysed 24 empirical studies across education, health and online communities.

Their conclusion was cautiously affirmative: gamification often increases engagement, improves participation quality, strengthens social interaction, and encourages positive behavioural change. Points, badges, leaderboards, progress tracking, and feedback were identified as key motivational drivers.

However, the authors were clear—gamification is not a universal solution. Its effectiveness depends on design quality, contextual fit, and user characteristics. In short, gamification works, but only when thoughtfully implemented.

IV. The Critical Counterpoint: A Self-Determination Perspective

While gamification can activate motivation, Self-Determination Theory warns that not all motivation is equal. According to this framework, sustainable well-being depends on satisfying three core psychological needs:

- **Autonomy** — the sense of choice and ownership. Systems that allow self-set goals support autonomy; rigid targets and guilt-inducing notifications undermine it.
- **Competence** — the feeling of growth and mastery. Meaningful progress feedback enhances competence; constant comparison through leaderboards may erode it.
- **Relatedness** — the need for connection. Cooperative challenges build belonging; excessive competition can damage it.

Gamification supports long-term motivation only when these needs are protected rather than manipulated.

V. Metric Fixation and the Ethical Risk

Historian Jerry Z. Muller, in *The Tyranny of Metrics*, critiques what he calls *metric fixation*—the tendency to substitute measurable indicators for genuine purpose. This critique directly applies to gamified systems.

When step counts replace health, streaks replace learning, and badges replace mastery, the symbol overtakes the substance. Behaviour may increase, yet meaning may decline.

Consider two runners:

One runs for joy, adjusting pace intuitively. The other runs to maintain leaderboard rank, ignoring injury and anxiety. Both log miles, but only one sustains well-being. The metric, once a guide, becomes the goal.

VI. Conclusion: Beyond the High Score

Gamification skilfully leverages human psychology to spark action and sustain engagement. Yet its power carries responsibility. If poorly designed, it risks replacing intrinsic growth with compulsive scorekeeping. The challenge, therefore, is not to reject gamification but to humanise it—to use game elements as scaffolds, not substitutes. True progress is not measured in points accumulated, but in autonomy strengthened, competence deepened, and relationships enriched. The aim is not to win the game of life, but to ensure that life itself remains more meaningful than the scoreboard.



**"THE SIGNAL IS THE TRUTH, THE NOISE
IS WHAT DISTRACTS US FROM IT."**

-NATE SILVER



CREATIVE CORNER

“Creativity is intelligence having fun.”

– Albert Einstein

The Creative Corner celebrates imagination, originality, and the artistic spirit of students. This section showcases creative works such as poetry, short stories, and other expressions of talent. It offers a space where ideas take shape and creativity finds its voice, encouraging readers to explore the beauty of imagination and the power of self-expression.

“Every idea begins with a spark of creativity.”



POEMS

Field Notes from a Data Scientist

I was trained to believe
that everything meaningful
could be measured
if not now, then after enough iterations.

So, I learned to clean data.
I removed outliers.
I normalized emotions into columns:
joy, regret, attachment, loss.
Missing values were imputed with reason.
Ambiguity flagged as noise.

The model performed well.
High accuracy.
Elegant convergence.
A reviewer would have nodded.
Yet something kept leaking
through the residuals.

Late at night,
while the servers ran quietly,
I noticed patterns that refused explanation
spikes without cause,
dependencies with no causal path,
variables that changed
when observed too closely.

Human behaviour, I learned,
is not *i.i.d.*

We say “data-driven,”
but data is only a fossil record
of lived experiences
compressed echoes of choices
already made,
never the moment of choosing itself.

Bias was the first confession.
Not just in the dataset,
but in me
in what I decided to collect,

what I named as signal,
what I discarded as error.

Then came overfitting.
I trained the model too well
on the past.
It predicted yesterday perfectly
and failed spectacularly
at tomorrow.

That is how love breaks a model.
That is how grief introduces
heteroskedasticity
into an otherwise stable life.

We speak of interpretability
as if meaning were a feature
we could extract.
But the most influential variables
are often latent,
unobserved,
unlogged,
known only to the person living them.

Eventually, I stopped asking
what the model could predict
and started asking
what it could not.

The uncertainty was not a flaw.
It was a boundary.

In the end,
data did not explain humanity to me
it taught me humility.

That intelligence, artificial or otherwise,
reaches its highest resolution
not when it claims understanding,
but when it learns
where understanding must end.

- Mr. Akashdeep Das
M.Sc. Statistics, Sem-IV

Living Contrast

A heart can break, yet mend again with time,
While circuits calculate the coming pain.
A human tear, a flawed and salty rhyme,
The AI sees as data, cold and plain.

It charts the graph of grief's supposed descent;
Predicts the day the weight begins to lift.
To where a fragile, weary hope is sent,
And measures how the emotional drift—
Will settle into something less than raw.

It knows the pattern, but it cannot feel;
The startling peace after the thunder's crow,
Nor comprehend the wounded, slow and real,
Unfolding of a soul that learns to bear
The memory of light in deep despair.

But when does joy arrive and how intense?
The model answers with a confidence:
“At 4:12 PM, with 89%
Probability of a sun-lit mind.”

It cannot grasp the laughter, sudden, free,
That blooms because a thought just wanders by,
Unscripted as a leaf dropped from a tree;

A brief, illogical and sunlit sky.
It maps the terrain with a flawless art,
A perfect, predictive and unchanging chart.

- Mr. Siddhartha Das
B.Sc.(Hons.) Statistics & Data
Science, Sem – II

Life in Posterior

We begin with prior dreams, we quietly hold,
Shaped by stories we've heard, by fears we were told.
Each failure updates us, each win leaves a trace,
New evidence teaching us where we stand in this space.

Uncertainty stays, but our belief grows refined,
Resampling courage with a stronger mind.
Life isn't fixed truth, it's learning in motion
A posterior built from hope and devotion.

Through likelihood whispers, the data unfolds,
Patterns emerging from futures untold.
We sample our paths from a vast hidden chain,
Markovian steps through joy and through pain.

Credible intervals guard what we know,
While variance humbles the confidence we show.

In convergence we find a balance so clear,
A life well-estimated, though never free of fear.

- Mr. Surajit Banik
M.Sc. Statistics, Sem – IV

The Story Beyond the Numbers

Late at night,
when everything becomes silent
and only the computer screen is shining,
I start thinking about our world today.

Statistics is everywhere now.
It quietly records our choices,
our habits, our dreams and our mistakes.
Without speaking a word,
it tells the real story of how we live.

Algorithms work in the background,
not visible, but always active.
They learn from us –
what we click, what we search,
what we like and what we hope for.
Sometimes they understand us
Even before we understand ourselves.

Cities run on data now.
Traffic, business, health, markets –
all controlled by numbers and patterns.
One small click can change our day,
one decision can change our future.

Technology gives us comfort,
yet sometimes we lose control.
Still, data helps us see the truth –
not the world we imagine,
but the world as it really is.

Numbers never lie.
They show where we succeed,
where we fail,
and where we can do better.
They highlight our strengths,
and reveal our weaknesses.

Algorithms are not good or bad on their
own.
They simply reflect the people who create
them.
Human minds build them,

and they carry human values, errors and
hope.

Today we stand in a new age –
part human emotion, part artificial
intelligence.
We are learning how to live with both.
And somewhere between this,
we are trying to understand who,
who are we in this modern world.

- Ms. Dattatreya Nandan
B.Com.(Hons.), Sem-II

Uncertainty

We chase the mean
and miss the spread,
trusting points
where ranges should be read.

Assumptions break,
the tails run long,
the model fits
then quietly goes wrong.

Yet in the noise,
a pattern stays,
soft and true
in subtle ways.

Statistics teaches,
clear and wise,
certainty fades
but insight survives.

- Mr. Tuhin Halder
M.Sc. Statistics, Sem – IV

Spreadsheets on Holiday

At the crosswalk where morning traffic
heaved, the city blinked awake in
coded light,
And every billboard crooned its promise of
a future graphed in black and
white
While hills, in gentle rows, aligned
themselves like spreadsheets longing for
a holiday.
The river ran its figures wrong again,
scribing eddies where the models
swore a line,
And sparrows offered fresh revisions,
chirping, "Forecast? Nay, 'tis
guesswork dressed as design."
And office towers, tidy in their grids,
looked down like clerks who think
they know the plan,
While somewhere in the queues and cafés
rose that stubborn, unaccounted
ache of man.
For though the data keeps its vigil, stern as
any watchman of the night,
There walks among its measured steps a
will that will not fit the chart's
delight.

- Ms. Srijita Talukdar
B.A.(Hons.) English, Sem – VI

Between Chaos and Design

Numbers whisper truths unseen,
Patterns pulse in the machine.
Algorithms weave the silent thread,
Guiding choices, the paths we tread.

Statistics sketch the world's disguise,
Turning chaos into graphs and lines.
Reality bends beneath the code,
A map of futures, yet untold.

Through shifting trends and silent cues,
They frame the lens by which we choose.
The hidden logic, sharp yet sly,
Writes the script beneath our sky.

In every dataset, shadows gleam,
We decode the fabric of the dream.
The modern world, both vast and small,
Is shaped by numbers after all!

- Ms. Aditi Shah
M.Sc. Statistics, Sem – IV





STORIES

When Sandesh Met Statistics: A Kolkata Sweet Shop's Sunday Mystery

Mr. Anish Barik
M.Sc. in Statistics, Sem - IV

Although the windows of Mishti Uthaan shop, near *Gariahat* crossing, were closed, Tuhin was unable to shut his laptop down. The crowded street, smell of samosas, noisy traffic filled the humid air of the city. There was a story being told in an Excel sheet and it was not a good one. The sales had decreased by 15% in comparison to last month's.

Sitting huddled over the screen, in a corner of his ancestral sweet shop, Tuhin was worried. Up against the flashy and Instagram-worthy cafes, fancy sweet shop chains and delivery apps, the drop in the sales figure was concerning. For him, the figure was far more than a statistic. It was stress. Sleepless nights. Fear of the changing times and what lay ahead. But mostly, he kept thinking, "Where had the sales vanished?"

Sundays were the best. Outside the shop, the hustle-bustle continued all day long. Families usually take a stroll in the afternoon, window-shopping through the busy streets of *Gariahat*, the children pulling at their sleeves to get them *rosogollas* (*rasgullas*) after a sumptuous lunch and in the quiet of the shop, old friends chatted over *sandesh*.

"This drop appears totally random," he said to Anik, who was pursuing an M.Sc. in Statistics at St. Xavier's University. "Did people suddenly get weary of my sweets?" Anik gently smiled, "You can't guess people's minds. But behaviour is measurable. Don't guess, let's measure! First of all, we must have a representative sample."

Sampling: who to observe, whom to overlook.

They were not able to cover the entirety of Kolkata, so they made arrangements in the vicinity of the shop. They did systematic sampling for two weeks; every 10th customer completed a brief survey. They would pick a random customer every day such that no trend developed. Anik was also able to pull six months of POS data and develop a stratified sample: weekdays, weekends, morning chai rush, lunchtime office breaks, evening crowd, etc., all in a proportional manner. The final dataset was good: 500 surveys + 2,000 transactions. It was not flawless but balanced to the extent needed to portray real Kolkata behaviour. A bill was more than just money; every bill carried context. Irrespective of whether it was Sunday or a public holiday; or any time of the day (morning, afternoon or evening); any weather, hot or rainy: in Kolkata, *mishti* sales is more important than any textbook syllabus.

"Who was on duty, Upen or Deba? "

"What was sold? *Sandesh*, *rosogolla*, *samosa* or tea?"

Total bill, on a simple average, it was established that Sunday is usually good. However, on making a correlation matrix, an odd pattern emerged. During the wet Sundays that Deba had to work, sales were low. Tuhin squinted, “How can that be? People love the *sandesh* that Deba makes!” Anik, scratching his head, said, “You are making a mistake. Trying to tell the story with just one variable! Here, everything interacts with one other; it is all working together.”

Multivariate thinking: seeing the big picture

Anik fitted a multiple linear regression model. Explained sales by day, weather, staff, product mix and interaction terms to determine which factors vary when put together. They also performed checks on residual plots, tests for multicollinearity and heteroskedasticity. Statistics involved being disciplined, not only by numbers.

The model (simplified):

Sunday	+ Rs. 15,000
Deba on duty	- Rs. 2,500
Humid Weather	+ Rs. 2,000
Sunday x Deba	- Rs. 13,000
Sunday x Humid Weather	- Rs. 4,000

All the effects were statistically significant. It was not very obvious but clear. Separately, Sunday was good. Humidity by itself wasn't bad. Deba wasn't to blame. However, once they all were together, customer behaviour transformed. They didn't stay as long. They bought less. They spent less. The sweets weren't worse. *The context had changed.*

Time Series: Rhythm in the Shop

Anik charted sales by the day and a 7-day moving average to eliminate noise. Trends were high on Saturdays, low on Mondays. Festival spikes. Rain dips. But one thing stood out. Sunday sales were gradually declining, not in a vacuum. A slow drift. This wasn't a one-off incident. It was a continual change of behaviour. The autocorrelation and weekly seasonality were verified. Anik was confident: the signal was real.

The *samosa-sandesh* experiment. Anik said, “On rainy Sundays, people don't look around the shop as much. They desire immediate comforts like samosa and tea or a little *sandesh*.”

Instead of arguing, they put it to *the test*.

There were other Sundays when there was a controlled normal service. Some were treatments, like a small board that said “Humid Sunday Special: *Samosa* + Tea / *Sandesh* Combo” and a mild implication on the part of the cheery staff. They varied on traffic, rain and local events to ensure that things were kept even.

The result was clear. Snack add-ons increased by 25%, the average bill increased by 15% and the t-test p-value came out to be 0.02. Not a coincidence. Likely causality. No blame, just redesign.

Tuhin didn't blame anyone. On weekdays, Deba was the *Mishti Master*, ideal *sandesh*, seasonal delicacies and winter-special favourites. On Sundays, he was the *Quick Combo Lead*, reading the mood of customers and proposing quick decisions. The dashboard was modified in less than a month. Sunday returned to the top. The antagonistic inter-relationship turned out to be a good one. The store was airy, cheerful and vivacious.

No one was accused based on the revelations of the statistics. It explained. Good sampling, the ability to think in a multivariate way, respect for time and the simplicity of experiments, all these made Tuhin glimpse at a known truth:

Indians in Kolkata are not only fond of sweets. They react to situations and statistics makes intuition savage. Good analysis is like a good *rosogolla*; it is all about balance.



Who Knows Me Better?

Ms. Purbasa Bhaduri
M.Sc. in Statistics, Sem – II

There was a moment, small, almost forgettable, when I realised something had shifted. I was scrolling through a streaming platform late at night. It recommended a documentary I had never heard of. I clicked on it absent-mindedly. It was exactly what I *needed* that day.

Not what I wanted.
Neither what I had searched for.
Exactly what I needed.

And that unsettled me. How could my phone be aware that I had been anxious that week? I never searched for an “existential documentary”. I never declared that I was tired of predictable plots. Yet somehow, the recommendation system anticipated my mood better than I consciously articulated it. That was the first time I felt the quiet intimacy of algorithms. We often talk about technology in dramatic language — artificial intelligence, machine learning revolutions, data capitalism. But the real transformation is quieter. It is statistical. It is probabilistic. And, it is deeply personal.

We No Longer Experience the World Raw

Once upon a time, we used to discover things by accident. We found a book because it happened to be placed at eye level in a library. We met people because they happened to live in our neighbourhood. We heard songs because we heard them play on the radio. Chance had *texture*.

Today, most of what we see has already passed through a filter — a model estimating what will keep us engaged, what we are likely to click, what will increase retention probability by a few percentage points. We like to believe we are exploring the world. But increasingly, we are exploring curated probability spaces.

Statistics is not sitting in a classroom anymore. It is embedded in recommendation engines, ranking systems, predictive scores. Every time something is “*suggested for you*”, it is because an algorithm has computed a likelihood based on past behaviour. And here’s the subtle shift: *when predictions influence what we see, they begin to influence who we become.*

If I am repeatedly shown certain types of content, my preferences start narrowing. My exposure shrinks invisibly. Not because I chose less but because my probability landscape was gently adjusted. Algorithms do not force. They nudge. But nudges, repeated millions of times, accumulate into patterns.

The Mathematics of Reduction

As a statistics student, I am trained to reduce complexity. We take messy realities and convert them into variables. Income. Age. Education. Click-through rate. Default probability. Risk score. It feels elegant. But in the real world, that elegance comes at a cost.

When a bank evaluates someone's credit worthiness, that person becomes a feature vector — repayment history, debt ratio, employment stability. The richness of their story dissolves into coefficients. When a hiring algorithm scans résumés, it searches for patterns correlated with past “successful” hires. It doesn't know ambition. It doesn't know resilience. It detects signals.

Reduction is necessary for computation. But reduction is never neutral. What we choose to measure determines what matters. What we ignore quietly fades from institutional recognition. Statistics gives the illusion of objectivity because numbers feel clean. But every model rests on assumptions — about what success means, about what risk is acceptable, about which errors matter more. Is it worse to wrongly deny a loan to someone who would have repaid? Or to approve a loan for someone who defaults? That is not just a mathematical trade-off. It is a value judgment disguised as optimization.

Optimization Has Replaced Intention

Modern algorithms do not ask what is meaningful. They ask what is optimal:

Maximize engagement.

Minimize cost.

Increase efficiency.

Reduce uncertainty.

In theory, optimization is beautiful. We define an objective function and search for the best solution. But in social systems, the objective function becomes a silent moral compass. If outrage drives engagement more effectively than balanced dialogue, an engagement-maximizing algorithm will not hesitate. It will amplify what performs. Not because it is malicious. Because it is consistent. Algorithms are ruthlessly loyal to their objective functions. And that loyalty can reshape culture. When metrics define success and success drives visibility and visibility drives influence — we slowly begin to live inside a world tuned to measurable performance. What cannot be quantified struggles to compete.

The Feedback Loop Nobody Talks About

The most fascinating and unsettling aspect of statistical systems is feedback. Imagine a predictive system that identifies a neighbourhood as “high risk”. More monitoring follows. More incidents are recorded. The data confirms higher risk. The model strengthens its prediction. But did risk increase? Or did observation increase?

Prediction does not merely describe reality. It alters the data that future predictions are built upon. This is a self-reinforcing loop. The same thing happens in digital platforms. If a system

predicts you prefer certain views, it shows you more of them. Your interaction reinforces the signal. Your informational world narrows. Gradually. Quietly. It is not censorship. It is statistical gravity. And over time, gravity reshapes landscapes.

Fairness Is Not a Simple Word

One of the most surprising lessons in algorithmic systems is that fairness cannot be universally defined. Different mathematical definitions of fairness conflict with one another. Equal error rates across groups may contradict equal predictive accuracy. Calibration can conflict with parity. This means that even with the best intentions, you cannot optimize for every fairness metric simultaneously. So, when institutions say an algorithm is fair, what they often mean is that they have chosen one fairness criterion over another. That choice is ethical. But it is presented as technical. This is where statistics stops being just mathematics and becomes philosophy.

The Strange Intimacy of Data

There is something almost unsettling about how deeply data captures us. Not just our purchases. Our sleep cycles. Our pauses before clicking. Patterns emerge, such that, even we may not consciously recognize. Algorithms can infer moods, preferences, tendencies — not because they understand us in a human sense, but because statistical regularities reveal patterns. It feels invasive — yet we participate willingly. Because the trade-off is convenience. And convenience is persuasive.

Who Is Really Shaping Reality?

It would be easy to blame “the algorithm”. But algorithms are trained on us. They learn from human behaviour — our biases, our preferences, our inequalities. In that sense, they are mirrors. But they are not flat mirrors. They are curved mirrors. They magnify what is statistically dominant. They amplify what generates signal. They smooth out anomalies. And in doing so, they bend reality slightly. Not enough for us to notice immediately. But enough that over time, the world begins to feel subtly guided.

Why This Matters

We are entering a world where decisions about loans, jobs, visibility, healthcare access and opportunity are increasingly mediated by statistical systems. Understanding statistics is no longer just an academic skill. It is civic awareness. To ask, “What is the false positive rate?” is to ask, “Who is being wrongly judged?” To ask, “What assumptions does this model rely on?” is to ask, “Whose reality was considered?”

To decode modern reality, we must decode the mathematics behind it. Because the algorithm does not shout. It calculates. It does not impose. It predicts. And prediction, repeated often enough, becomes influence. The question is not whether algorithms shape reality. They already do. The real question is whether we understand how. Because once we do, we reclaim something subtle but powerful: *The ability to question the numbers that quietly shape our lives.*



Family Matter of Distributions

Mr. Tuhin Halder
M.Sc. in Statistics, Sem - IV

In the Kingdom of Randomness, all probability distributions lived together, mostly in peace but sometimes in confusion. The Bernoulli twins were the simplest, they only answered questions with yes or no and refused to discuss anything in between. Their cousin Binomial loved counting how many successes occurred after repeatedly asking Bernoulli questions. Geometric, on the other hand, was impatient and asked, “Are we there yet?” after every trial until the first success came.

Nearby, Poisson ran the town’s emergency hotline. He calmly counted rare events each hour, such as accidents, emails, or unexpected exam questions. Exponential worked with Poisson and proudly handled waiting times, reminding everyone that memory doesn't matter much. Meanwhile, Gamma quietly generalized Exponential and smiled smugly whenever someone said, “This looks complicated.”

Across the river lived the elegant Normal distribution. It was perfectly symmetric and believed the world revolved around the mean. Z-scores orbited it like devoted fans. However, Normal’s cousins t , Chi-square and F were much more dramatic. They showed up whenever sample sizes were small or variances clashed. Uniform lounged lazily, claiming all outcomes were equally likely. Beta stayed flexible, constantly changing shape to model probabilities.

At the edge of town, Lognormal reminded everyone that life isn’t symmetric, incomes aren’t fair and right tails can be very long. Above them all, the mysterious Cauchy caused chaos with no mean, no variance and no apologies.

Together, these distributions told the same story in different ways: uncertainty comes in many forms, assumptions matter and no single model controls them all. And the statistician? Just a tired narrator trying to choose the right character for the data, hoping the exam doesn’t ask for all of them at once.



ভালোবাসাই যাঁদের Objective Function: বাবা মা

Mr. Diptajit Chakraborty
M.Sc. in Statistics, Sem – IV

যখন গ্রামে থাকতাম, জানতাম না স্ট্যাটিসটিস্ট বলে কিছু আছে, data মনে বুঝতাম ফোন এর ইন্টারনেট প্যাকেজ, ব্যাস এটুকুই। তখনও বুঝিনি আমার বাবা-মা ই আমার জীবনের আসল আর সবচেয়ে দক্ষ Statisticians।

ছোট বেলায় বাবা মা আমাকে খুব শাসন করতেন কখনও মারতেন, কখনও ধমক দিতেন কখনও বা এমন কঠোর শাস্তি দিতেন যার মানে তখন বুঝতাম না।

আজ বুঝি। ওগুলো ছিল আমার জীবনের ভুল-ত্রুটি তথা আমার residual কমানোর চেষ্টা। হয়তো তাঁদের কাছে সেগুলিই ছিল আলাদা loss function। কোথাও MSE, কোথাও MAE, আবার কোথাও বা RMSE, যাতে ভবিষ্যতে আমার ভুল (error term) টা কম হয়।

আমি ছোট থেকেই খুব একটা মনোযোগী ছিলাম না। বাবা জোর করে রোজ পড়তে বসাতেন। আমি ভাবতাম “কেনো শুধু পড়ব এত?”। আজ বুঝি, হয়তো বাবা চাইতেন আমি যেন জীবনের পরীক্ষায় একটা unbiased, consistent estimator হয়ে উঠি, যে সব প্রশ্নের উত্তর দিতে পারে, ভুল করলেও ধীরে ঠিকের দিকে converge করে। আমার বাবা পেশায় শিক্ষক, তাই তাঁর কাছে data (ছাত্রছাত্রী দের আচরণ, ভুল, উন্নতি, ব্যর্থতা আরো কত কি)তো অফুরন্ত আছেই। আর সেই data দিয়েই তিনি বানিয়েছিলেন নিজের একটা regression model। আর আমি ছিলাম সেই মডেল এরই একটি test data point। কাজেই যার এত নিখুঁত মডেল তৈরি আছে তার থেকে কি কিছু লোকানো যায়! তাই মিথ্যে বললে বা শয়তানি করলে ঠিকই ধরে নিতেন নিমেষে। একবার বাবা কে বলতে শুনেছিলাম “আমি সব ই বুঝি কে কি ভাবে কিন্তু বলি না।” তখন ভাবতাম এত নিখুঁত prediction কীভাবে সম্ভব। এখন বুঝি, বাবা হয়তো সেই function $(h(x))$ টা বের করে ফেলেছেন যেখানে x একটি matrix (x_1, x_2, \dots) এগুলি মানুষের চরিত্র, অভ্যাস, চোখের ভাষা ইত্যাদি।

তখন সবে কলেজ এ পা দিই, বাবা মা আমাকে কলকাতায় একটি বাড়িতে পেয়িং গেস্ট হিসেবে রেখে যাবেন, বাড়ির মালিক বাবা মা কে বললেন “কলকাতার হাওয়া কিন্তু ভালো না। ভালো ছেলেরা খারাপ হয়ে যায়। সচেতন থাকতে হবে।” বাবা শান্ত গলায় বললেন “না, এসব কিছু করবে না ও। ও ঠিকঠাক e থাকবে।” আমি তো নিজেও নিজের ব্যাপারে এত নিশ্চিত ছিলাম না, বাবা কিকরে এত নিশ্চিত। শেষে বুঝলাম এখানে তো বাবা time series analysis করেছেন। আমার আচরণের trend, অভ্যাসের seasonality দেখে হয়তো Holt Winters পদ্ধতিতে কলেজজীবনের সম্ভাব্য পথটা আগেই অনুমান করে নিয়েছিলেন। এখনও পর্যন্ত, আমি সেই forecast এর বাইরে খুব একটা বেরোতে পারিনি।

এবার মা এর কথা বলা যাক। আমার মা আমাদের বাড়ির complete sufficient statistic। Sufficient statistic এর মতই মা এর কাছে বাড়ির যাবতীয় পুঙ্খানুপুঙ্খ সব তথ্য, আমাদের না বলা কথা সব থাকে। আর complete বলেই তাঁর সিদ্ধান্তের বাইরে কিছু থাকে না।। আর অবশ্যই সবচেয়ে শ্রেষ্ঠ ডিসিশন তিনিই দেন।

আমার মাঝে মধ্যে মনে হয় মা কিকরে বুঝতে পারেন আমি শুয়ে থাকলে সেটা মন খারাপ নাকি শরীর খারাপ। পরে বুঝলাম এখানে তো মা Bayesian analysis করেন। আমার হাসি, কান্না, খিদে ঘুম এইসবই তাঁর কাছে prior belief। আর আমার দৈনন্দিন জীবনে সেই জিনিস গুলি কে কেন্দ্র করে ঘটে চলা জিনিস গুলি likelihood। এই দুইয়ের সমন্বয়ে তিনি নেন posterior ডিসিশন, যাতে করে তিনি বুঝতে পারেন আমার আসল সমস্যাটা কোথায়। কোন তরকারি তে কি পরিমাণে সবজি, নুন, তেল দিলে তরকারিটা আরো সুস্বাদু হবে, কিভাবে সংসার চালালে মাসের শেষে সাশ্রয় বাড়বে আর সঠিক খাতে অর্থ ব্যয় হবে সবতেই তাঁর লক্ষ্য ছিল cost minimisation আর utility maximization। এসব বিষয়ে মা এর efficiency প্রশংসিত।

বাবা মা, চাল সবজি মুরি একটু হাতে নিয়ে দেখে অনায়াসেই বলে দেন সেগুলোর quality কেমন হতে পারে। এটাই তো sample survey। এইসব analysis a তাদের থেকে পারদর্শী analyst কে আছে?

স্টাটিস্টিসিয়ানদের নিয়ে যখন আলোচনা হলোই এবার একটু যেই data point এর ওপর তাঁদের মুখ্য কাজ তা নিয়ে বলা যাক।

বাবা মা এর আমাকে ছোটবেলায় হাত ধরে হাটা শেখানো, অক্ষর চিনতে শেখানো, অঙ্ক বোঝানো, ভাল মন্দ আলাদা করে বুঝতে শেখানো আর খাইয়ে পরিয়ে মানুষ করা সবই ছিল supervised learning। তাদের সেই জ্ঞান কে পাথেয় করেই আজ এই statistics নিয়ে পড়াশোনা করছি, বন্ধু বান্ধব বানিয়েছি, (statistics এর ভাষায় যাকে কিনা বলে cluster), জীবনের ভিড়ে নিজের একটা ছোট্ট জায়গা করে নিতে পেরেছি। একে তো unsupervised learning বলা যেতেই পারে। হয়তো জীবনের শেষ দিকে এসে ভালো-মন্দের ফিডব্যাক নিয়ে নিজেকে বদলাবো বারবার। সেটাই হবে reinforcement learning।

অর্থাৎ সব দিক বিবেচনা করে দেখলে আমার বাবা মা ই সেই statisticians যাদের জন্য তাদের তৈরি মডেলে আজও এই data point টি নীরবে চলছে তাদের স্নেহ,ভালবাসা আর মতাদর্শের parameter এ। বাবা মা model তো বানিয়েই দিয়েছেন এবার ভবিষ্যত কি হবে তা নির্ভর করে পুরোপুরি data পরবর্তী ধাপগুলোতে কি পদক্ষেপ নেবে তার ওপর।

কিন্তু এতদিনে অল্প বিস্তর যা একটু শিখেছি statistics সম্পর্কে তাতে আমি নিশ্চিত ভাবে বলতে পারি আমার বাবা-মাই আসল আর সবচেয়ে নিখুঁত statisticians, না হলে ভুলে ভরা, error এ ডুবে থাকা এই আনস্মার্ট observation টাকে গ্রামের ওই ছোট্ট ঘরের নিরাপদ কোণ থেকে জীবনের দীর্ঘ ট্রেনের যাত্রী করে তুলত কে?



INNOVATIONS & CASE STUDY

*“Innovation distinguishes between a leader
and a follower.”*

— Steve Jobs

The Innovation & Case Study section explores ideas that challenge conventional thinking and encourage analytical insight. Through case studies and innovative perspectives, this section highlights real-world problems, creative solutions, and thoughtful research. It invites readers to think critically, question assumptions, and discover how new ideas can shape the future.

*“Great ideas begin with curiosity and the
courage to explore.”*

Breaking the Monotony

Mr. Sankhadeep Nag
M.Sc. Statistics, Sem - II

In the realm of bivariate analysis, the question “Are X and Y related?” is deceptively complex. If the relationship is linear, Karl Pearson’s correlation coefficient (r) provides a sufficient answer. If the relationship is monotonic, meaning Y consistently rises or falls as X increases, Spearman’s ρ and Kendall’s τ serve us well. These classical coefficients are powerful, boasting well-developed asymptotic theories that allow for easy calculation of P-values.

However, the world is rarely strictly linear or monotonic. Consider a relationship that is sinusoidal, parabolic, or circular. In these cases, even if Y is a noiseless function of X (meaning Y is perfectly determined by X), the classical coefficients often return values near zero, erroneously suggesting independence. This inability to detect non-monotonic associations is a serious deficiency that has plagued practitioners for decades.

While modern alternatives like the Maximal Information Coefficient (MIC), distance correlation and kernel-based methods have been proposed to fill this void, they come with significant computational costs and often lack simple asymptotic theories for hypothesis testing. Enter Sourav Chatterjee’s new coefficient of correlation, denoted as ξ_n . It is a statistic that is as simple to compute as the classical measures but possesses the “holy grail” of properties: it is a consistent estimator that approaches 1 if and only if Y is a function of X and 0 if and only if they are independent.

Mathematical Formulation:

The beauty of Chatterjee’s coefficient lies in its algorithmic simplicity, which belies its theoretical depth. Let (X, Y) be a pair of random variables where Y is not a constant. We are given n independent and identically distributed (i.i.d.) pairs $(X_1, Y_1), \dots, (X_n, Y_n)$.

To compute ξ_n , one does not need complex partitions of the sample space or high dimensional kernel mapping. Instead, the calculation relies entirely on ranks, rendering the metric robust to outliers and invariant under strictly increasing transformations.

The Algorithm:

The calculation proceeds in three steps:

- Rearrangement:** Sort the data pairs based on the X values such that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. If there are ties in X, they can be broken uniformly at random.
- Ranking:** Once sorted by X, we observe the corresponding Y values: $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$. Let r_i be the rank of $Y_{(i)}$ among all Y values. That is, r_i is the count of j such that $Y_{(j)} \leq Y_{(i)}$.
- Calculation:** The coefficient is derived from the differences in ranks between adjacent sorted data points.

If there are no ties in the Y values, the coefficient is given by the elegant formula:

$$\xi_{n(X,Y)} := 1 - \frac{\{3 \sum_{i=1}^{n-1} |r_{\{i+1\}} - r_i|\}}{\{n^2 - 1\}} \dots\dots\dots(1)$$

In the presence of ties in Y, the formula is generalized to maximize robustness. We define l_i as the number of j such that $Y(j) \geq Y(i)$ and the formula becomes:

$$\xi_{n(X,Y)} := 1 - \frac{\{n \sum_{i=1}^{n-1} |r_{\{i+1\}} - r_i|\}}{\{2 \sum_{i=1}^{n-1} l_i(n - l_i)\}} \dots\dots(2)$$

Intuition: Why Ranks?

The logic behind ξ_n is intuitive. If Y is a function of X (say, $Y = f(X)$), then sorting the data by X should result in the Y values also being “sorted” or at least locally continuous. Consequently, the ranks of adjacent Y values (r_i and $r_i + 1$) should be close to each other. This minimizes the sum of absolute differences $\sum |r_i + 1 - r_i|$, pushing the coefficient toward 1. Conversely, if X and Y are independent, sorting by X leaves the Y ranks in a random permutation, maximizing the differences and pushing the coefficient toward 0.

What makes ξ_n a topic of intense academic interest is not just its formula, but the rigorous proofs governing its behavior.

Consistency:

Chatterjee proved that as $n \rightarrow \infty$, $\xi_n(X, Y)$ converges almost surely to a deterministic limit $\xi(X, Y)$. This limit is defined as:

$$\xi(X, Y) := \frac{\{\int \text{var}(E(1_{\{Y \geq t\}}|X)), d\mu(t)\}}{\{\int \text{var}(1_{\{Y \geq t\}}), d\mu(t)\}} \dots\dots\dots (3)$$

where μ is the law of Y. Crucially, this limit ξ is 0 if and only if X and Y are independent and 1 if and only if Y is a measurable function of X almost surely. This property of “detecting any functional dependence” is absent in Pearson, Spearman and Kendall correlations.

Asymptotic Normality: The Hypothesis Testing Revolution:

One of the most painful aspects of using advanced correlation measures like Distance Correlation or MIC is the difficulty in calculating P-values. Because their distributions under the null hypothesis (independence) are complex, researchers often resort to permutation tests—shuffling the data hundreds of times to build a null distribution. This is computationally expensive, $O(n^2)$ or worse per permutation.

Chatterjee’s ξ_n solves this elegantly. Under the hypothesis that X and Y are independent and Y is continuous, Chatterjee proved that:

$$\sqrt{n}\xi_{n(X,Y)} \xrightarrow{d} N(0, 2/5) \dots\dots\dots (4)$$

This is a massive breakthrough. It implies that for a sample size as small as $n = 20$, the distribution of the statistic is approximately normal. Researchers can compute a p-value instantly using the standard normal table, avoiding the need for time-consuming bootstrap or permutation procedures.

Asymmetry: A Feature, not a Bug:

A striking feature of ξ_n is that it is not symmetric. That is, $\xi_n(X, Y) \neq \xi_n(Y, X)$. In classical statistics, we are trained to view correlation as a mutual relationship. However, Chatterjee argues that dependence often flows in one direction. We might wish to know if Y is a function of X, without implying that X is a function of Y.

If a symmetric measure is required, one can simply take $\max(\xi_n(X, Y), \xi_n(Y, X))$, which retains the property of being 0 under independence and 1 if either variable is a function of the other.

Case Study I: Galton's Peas Revisited:

To illustrate the power of this asymmetry, Chatterjee applied ξ_n to one of the oldest datasets in statistics: Sir Francis Galton's 1875 sweet peas data. The dataset pairs the mean diameter of "mother" plants (X) with "daughter" plants (Y).

Standard Pearson correlation yields a value of approximately 0.35, suggesting a weak positive relationship. When Chatterjee applied his new coefficient, the results were startlingly different depending on the direction:

- $\xi_n(X, Y) \approx 0.11$ (suggesting very low dependence of Child on Parent).
- $\xi_n(Y, X) \approx 0.92$ (suggesting extremely high dependence of Parent on Child).

How can the "Parent" be a function of the "Child"? A closer inspection of the contingency table revealed a quirk in the data: for every specific diameter of a daughter plant, the mother plant had a fixed, constant diameter. Conversely, a single mother diameter resulted in a wide variety of daughter diameters.

While biologically this might be an artifact of data collection (or recording methods), mathematically, X was indeed a nearly noiseless function of Y, but Y was a very noisy function of X. ξ_n detected this functional truth where Pearson yielded a symmetric and arguably misleading, average.

Case Study II: The Hidden Oscillations of Yeast:

The true test of a new statistic is its ability to discover what others miss in high-dimensional biological data. Chatterjee analyzed a landmark gene expression dataset by Spellman et al. (1998), which tracked the transcript levels of 6,223 yeast genes over time to identify cell-cycle regulation.

The goal was to find genes that oscillate. Standard correlation tests prioritize linear or monotonic trends, often ignoring cyclic behavior (up-down-up patterns). When Chatterjee applied ξ_n , he identified 215 genes that were selected by ξ_n but ignored by every other standard test.

Visual inspection of the top 6 genes from this group revealed perfect oscillatory behavior. In contrast, genes selected by other tests but missed by ξ_n often showed simple trends or heteroskedasticity but lacked the distinct cyclic pattern crucial for cell-cycle research. This demonstrates not just a theoretical novelty; it has superior power in detecting oscillatory signals that are invisible to the "Big Three" coefficients.

Computational Superiority: The Speed Demon

In the era of Big Data, an algorithm's time complexity is as important as its accuracy. The most robust competitors to ξ_n —such as the Hilbert-Schmidt Independence Criterion (HSIC) or Distance Correlation (dCor)—generally rely on $O(n^2)$ computations or expensive permutation tests for significance.

Chatterjee's coefficient, relying on sorting, runs in $O(n \log n)$ time. The difference in practice is staggering. In a comparison of run times for a sample size of $n = 10,000$:

- **Distance Correlation:** 89 seconds
- **MIC:** 1097 seconds
- **HHG Test:** \approx 30 minutes (did not converge)
- **Chatterjee's ξ_n :** 0.011 seconds

This massive speed advantage makes ξ_n essentially the only viable option for testing independence in massive datasets where n reaches into the tens or hundreds of thousands.

Limitations: The Smoothness Penalty

Despite its “magical” properties, ξ_n is not a panacea. There is no free lunch in statistics. While ξ_n excels at detecting oscillatory and sharp patterns (like W-shapes or sinusoids), it suffers from lower power compared to distance correlation or MIC when the relationship is smooth and non-oscillatory (e.g., a simple linear relationship with noise).

Simulation results show that for linear, circular, or heteroskedastic data, ξ_n requires more samples to achieve the same power as distance correlation. This is the trade-off for its generality; by spreading its sensitivity across all functional types, it becomes slightly less sensitive to the simplest ones. However, Chatterjee notes that in large samples, all consistent tests are powerful and the computational speed of ξ_n becomes the dominant factor.

Comparison with MIC and Maximal Correlation

One might ask: “Doesn't the Maximal Information Coefficient (MIC) already solve this?” It is a common misconception that MIC and Maximal Correlation measure the strength of a relationship in a way that implies noiselessness. Chatterjee showed this is false via a counterexample involving a mixture of bivariate normal distributions.

In a scatterplot showing two distinct, noisy clusters, MIC and Maximal Correlation yielded scores of 1.00 and 0.99, respectively—implying a perfect noiseless relationship where none existed. Chatterjee's ξ_n yielded 0.48, a value that far better reflected the noisy reality of the data. This highlights a critical distinction: ξ_n approaches 1 only if the data is noiseless, whereas other “modern” coefficients can be “tricked” by clustering or specific distributions into reporting perfection where there is significant noise.

Conclusion:

Sourav Chatterjee's ξ_n represents a paradigm shift in how we quantify association. It bridges the gap between the interpretability of classical statistics and the versatility required for

modern, complex datasets. Its formulation is shockingly simple, yet it satisfies theoretical conditions that have eluded statisticians for a century.

By offering a coefficient that is consistent against all alternatives, computationally tractable ($O(n \log n)$) and equipped with a simple asymptotic theory, Chatterjee has provided the scientific community with a powerful new lens. While it may not replace Pearson for strictly linear questions, ξ_n is poised to become the default “first look” in exploratory data analysis, capable of spotting the hidden patterns—the sinusoids, the parabolas and the biological oscillations—that the classical giants simply cannot see.

References:

- Chatterjee, S. (2020). “A New Coefficient of Correlation.” arXiv:1909.10140v4 [math.ST].
- Chatterjee, S. (2021). “A new coefficient of correlation.” Journal of the American Statistical Association.
- Galton, F. (1875). “Peas Data.”
- Spellman, P. T., et al. (1998). “Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.” Mol. Biol. Cell.



Climbing the Ladder: A Statistician's Journey Through Global Happiness

Ms. Kritikka Chatterjee
M.Sc. Statistics, Sem - II

I did not begin this project searching for happiness. I began with a dataset. Rows of countries, columns of numbers, and a deceptively simple question asked across the world: “On a scale from zero to ten, where do you stand in your life today?” At first glance, it looked like just another statistical exercise: clean the data, run models, test hypotheses, write conclusions. But as the analysis unfolded, the numbers stopped behaving like numbers. They began to behave like people.

This is the story of the journey, of how global happiness can be measured, questioned, forecasted and sometimes misunderstood. It is a story told not from the mountaintop of philosophy, but from behind a desk cluttered with graphs, test statistics and regression outputs.

The Ladder That Connects the World

Every year, the World Happiness Report ranks countries based on something called the Ladder Score. People imagine a ladder with steps numbered from 0 (the worst possible life) to 10 (the best). Where they place themselves becomes data—aggregated, averaged, and compared across nations.

But the ladder does not stand alone. It is held steady by six pillars: - Gross Domestic Product (GDP) per capita - Social support - Healthy life expectancy - Freedom to make life choices - Generosity - Perception of corruption. Together, these factors attempt to explain why Finland smiles at the top while Afghanistan struggles at the bottom.

As a student of statistics, I was fascinated. As an Indian, I was unsettled. India ranked 126th out of 144 countries in the 2024 report—below nations facing war, instability, and economic hardship.

The question was unavoidable: **How could a fast growing economy feel so unhappy?** That question shaped everything that followed.

India Through Time — Eighteen Years on the Happiness Scale

Before touching any test statistic, I spent time simply *looking* at the data. Visualization was not a formality; it was the first conversation with the series. I began by isolating India's Ladder Score from 2006 to 2023: 18 years of collective mood swings. Plotting the values on a graph felt like watching a heartbeat on a monitor. The line did not trend smoothly upward, as economic growth narratives might suggest. Nor did it collapse entirely. Instead, it fluctuated—rising, falling, hesitating, recovering.

From a statistical standpoint, three facts became immediately clear:

1. There was **no seasonality** — happiness did not rise or fall in predictable cycles.
2. There was **no stable trend** — neither consistently increasing nor decreasing.
3. The **variability was irregular** — shocks like the pandemic left visible scars.

This was not a calm series. It was restless. What stood out most was a **structural break around 2019–2020**. A downward drift that had persisted for years suddenly reversed, climbed sharply during the pandemic period, peaked around 2020, and then declined again before showing mild recovery. Statistically, this behaviour suggested shock-driven dynamics rather than organic growth. Happiness, here, appeared reactive—responding to crises, policy interventions, and collective uncertainty rather than following a smooth developmental path.

When Data Refuses to Sit Still — Stationarity Tests

The visual irregularity raised an immediate technical concern: *can this series be modelled using classical time series tools at all?*

In time series analysis, there is a fundamental requirement: stationarity. A stationary series has a stable mean and variance over time. India's happiness data refused to comply. To test this formally, I applied two powerful statistical tools:

1. Augmented Dickey-Fuller (ADF) Test:
Null Hypothesis: The series is non-stationary.
Result: The p-value was far greater than 0.05.
Conclusion: I could not reject non-stationarity.
2. KPSS Test:
Null Hypothesis: The series is stationary.
Result: The test statistic exceeded the critical value.
Conclusion: Stationarity was rejected again.

When both tests disagree or jointly indicate instability, statisticians take notice. I tried everything—first differences, logarithms, log differences, even second differences. Nothing worked. India's happiness, statistically speaking, refused to settle into equilibrium. This itself was a finding.

In formal time series terminology, the series was neither **trend-stationary** nor **difference stationary** in any conventional sense. The failure of multiple transformation techniques suggested that India's happiness dynamics are governed by evolving socio-economic structures rather than short-term stochastic fluctuations.

Forecasting the Future — ARIMA and Uncomfortable Truths

Given the inability to impose stationarity externally, I allowed the model to handle it internally. With stationarity elusive, I turned to a pragmatic solution: **ARIMA modelling**. Using ACF and PACF plots, followed by the `auto.arima()` function, the best-fitting model emerged:

ARIMA (1,1,0) with drift

I split the data into training (2006–2016) and testing (2017–2023) sets. The model performed surprisingly well—closely tracking the observed values despite the presence of extreme fluctuations.

Then came the most sobering moment of the project. I forecasted India’s happiness for the next 15 years. The prediction interval widened, but the central forecast remained almost flat. There was **no statistical evidence of a significant upward surge** in India’s Ladder Score. The numbers did not shout. They whispered: *Time alone will not heal structural dissatisfaction. Growth alone is not enough.*

Looking Beyond India — A Global Regression Experiment

After examining how happiness behaves *over time* within one country, I shifted to a broader question: *what explains differences in happiness across countries at a given point in time?*

The second part of my study widened the lens to 144 countries. I built a Multiple Linear Regression (MLR) model, with the Ladder Score as the dependent variable and six covariates—GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perception of corruption—as independent variables. No weights. No adjustments. Every factor entered the model simultaneously.

Before fitting the model, a practical obstacle emerged: **missing data**. Several countries lacked observations for one or more covariates. Rather than discarding large portions of the dataset or relying on simplistic mean imputation, I adopted a functional imputation approach, preserving distributional characteristics while ensuring analytical completeness.

What emerged after model fitting was striking: **GDP per capita, social support, life expectancy and freedom** showed strong positive relationships with happiness. **Perception of corruption** mattered deeply—trust, it turns out, is statistical. Generosity, surprisingly, consistently emerged as the least influential factor.

This result was robust across countries and statistically validated through hypothesis testing. At the 5% significance level, the coefficient corresponding to **Generosity** failed to reject the null hypothesis ($\beta = 0$), while all other covariates showed statistically significant effects. The implication was subtle but powerful: *generosity flourishes when basic security is already present*. It is not the foundation of happiness—it is the outcome of it.

The Paradox of Equal Weights

A frequent methodological criticism of the World Happiness Index is its use of equal weights across indicators. Critics often argue that the World Happiness Index is flawed because it assigns equal weight to all six factors.

My analysis suggested something important: Even *without* weighting or manipulation, the data naturally revealed which factors mattered more. The index was not arbitrary. In fact, the regression results demonstrated that even when indicators are treated symmetrically at the construction stage, inferential statistics naturally reveal asymmetry in influence. Bhutan’s emphasis on Gross National Happiness suddenly felt less idealistic and more empirical. The Nordic model felt less cultural and more structural.

What Statistics Cannot Measure — And What It Can

Every statistical project ends with a conclusion, but not every conclusion feels final. Statistics cannot measure joy the way poetry does. It cannot capture laughter, belonging, or hope in a single equation. But it can do something equally important. It can tell us when progress is hollow. When growth is uneven. When development forgets the human mind.

This project taught me that happiness is not noisy. It does not announce itself with GDP figures or skyscrapers. It lives quietly in trust, health, freedom, and the assurance that tomorrow will not be worse than today.

The ladder still stands. Whether we climb it depends not on how fast we grow, but on how thoughtfully we choose to grow!



Normalising the Incomparable: Comparison of Water Quality of Major Rivers Across Continents

Ms. Kritee Mukherjee

Mr. Souptik Bose

Ms. Kousiki Mukhopadhyay

Mr. Rajnik Pramanik

Ms. Trisha Patra

B.Sc.(Hons.) Statistics & Data Science, Sem – II

When datasets contain fundamentally different variables with incompatible units, direct comparison becomes a statistical challenge, and that was exactly our issue in this project which tackled pollution assessment in the context of rivers with completely different pollution profiles from different continents, pushing us to develop a normalisation algorithm to compare their contamination severity back in our first semester.

We had four different Rivers: The Yamuna(India), the Amazon(South America), the Citarum(Indonesia), and the Techa(Russia). We started with our project named “Water Quality Comparison of Major Rivers Across Continents” and decided to choose the rivers that provide a proper representation for as many pollution types as we could fit into our tiny chart along with various other information and soon enough found ourselves in the middle of an “apples and oranges” problem that eventually led us to make several statistical choices about what to measure and how to find a common ground to compare everything in one single graphical representation.

The Yamuna River's biological oxygen demand (measured in mg/L), the Techa River's radioactivity (measured in Bq/L), the Amazon's mercury concentration (measured in $\mu\text{g/L}$), and the Citarum's complex chemical

cocktail represent different physical quantities with different units, measurement techniques, and health implications.

The statistical question became: can we design an algorithm that transforms these heterogeneous measurements into a common framework

that preserves meaningful information about relative danger while enabling direct comparison?

This was a clear question about Data Normalisation.

The Statistical Challenge: Structuring Heterogeneous Data *Dataset Construction and Variable Selection*

We found that we could neither come up with a way at that point of time which can help compare the total intensity of pollution in each river, nor could we let go of the idea to bring together the four rivers in a single graphical representation. Hence, we made the decision to choose a particularly disturbing component of the total pollution from each river:

Yamuna River (Delhi Main Stem, India):

➤ DO- Often 0 mg/L (polluted sites); ≈ 7 mg/L at upstream Palla ➤ BOD- Typical reported values 40–130 mg/L

➤ TC/FC- Very high — 10^5 – 10^7 MPN/100 mL in worst samples Dominant Pollution Type: Biological Pollution.

Amazon River (Main Stem, Manaus, Brazil):

➤ DO- $\approx 6.5 \pm 0.7$ (near-saturation typical)

➤ BOD- Generally < 5 (mainstem); localised spikes > 10 – 20 at impacted points

➤ TC/FC- Generally low in free-flowing mainstream, site-specific increase in cities (100-1000 MPN/100mL)

Chosen Pollution Type: Heavy Metal Accumulation

Citarum River (Main Stem, West Java, Indonesia):

➤ DO- ≈ 1.1 – 1.6 mg/L (polluted downstream); upstream higher (~ 5 – 6 mg/L)

➤ BOD- ~ 43 mg/L (study mean for impacted segments)

➤ TC/FC- very high — thousands to millions MPN/100 mL in impacted stretches

Dominant Pollution Type: Chemical Pollution

Techa River (Main Stem, Chelyabinsk Oblast, Russia):

➤ Radiation due to ^{90}Sr : ~ 5 – 26 Bq·L $^{-1}$ typical monitoring in 1990s, found in upper/middle Techa near the release area

➤ Radiation due to ^{137}Cs : ~ 0.1 – 1 Bq·L $^{-1}$ typical

➤ Radiation due to Pu isotopes: ~ 0.004 – 0.02 Bq·L $^{-1}$ reported Dominant Pollution Type: Radioactive pollution

Each river has measurements for different variables, a case in which traditional comparative statistics often tends to fall behind when datasets don't share common variables.

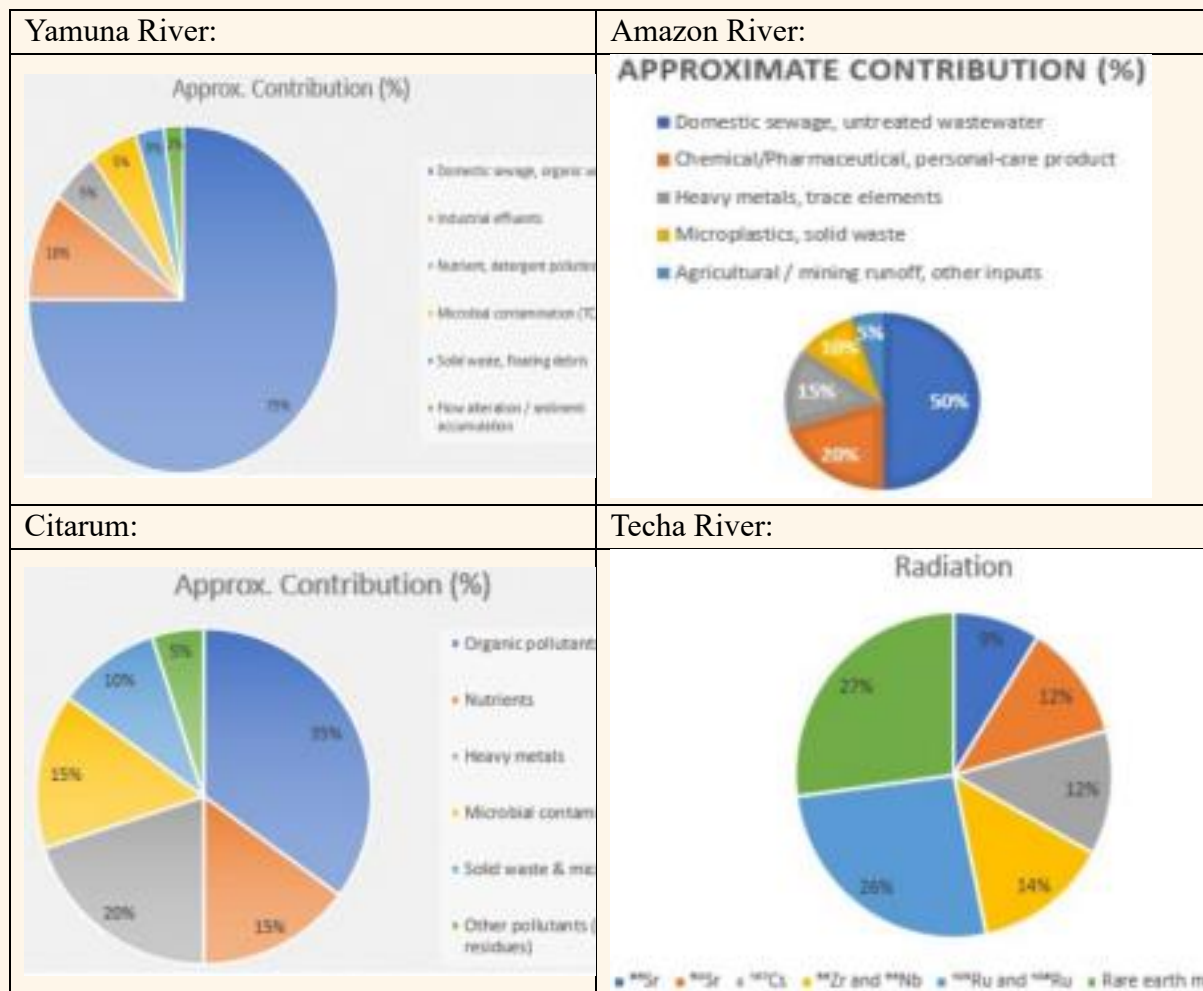
Exploratory Data Analysis: Composition Visualisation

Before developing our comparative algorithm, we explored further into each river individually using pie charts to visualise pollutant or source of pollution composition. This served two purposes:

➤ **Data validation:** This ensures that the choice of pollutant for our overall comparison for all the rivers is correct

➤ **Algorithmic insight:** Identifying which variable within each dataset carried the most statistical weight, thus helping us realise why the final graph might not look like what we shall expect

The Pie Charts:



For multi-variable rivers like Citarum, the pie chart revealed near-equal distribution across pollutant types; however, on the other hand, in the Yamuna, domestic waste dominated the pollution signature, and the Techa isotope distribution showed varying contributions from different radionuclides.

The Normalisation Algorithm: From Raw Data to Danger Scores

The algorithm operates in two steps:

Step 1: Determining the most critical pollutant/pollution for each river

We have four rivers, Yamuna, Amazon, Citarum and Techa, and for each of these we have one representative indicator, a typical value and their dominant pollution type as shown in the following table:

River	Dominant Pollution Type	Representative Indicator	Typical Value	Units	Reference
Yamuna (India)	Organic + sewage	Fecal coliform	9.2×10^6	MPN/100 mL	CPCB 2021
Citarum (Indonesia)	Industrial + domestic waste	BOD	9.8	mg/L	Marselina et al., 2022
Amazon (Brazil)	Heavy metal (mercury)	Hg in hair	2.5	$\mu\text{g/g}$	WHO/UNEP Amazon studies
Techa (Russia)	Radioactive contamination	^{90}Sr	10	Bq/L	IAEA Scenario T

Table 1.0 Showing pollution types, representative indicators, typical values and references

Step 2: Ratio-Based Normalization

After having determined the type of pollution to be considered, the normalised danger score is calculated where:

$$\text{Normalized Danger Score} = \frac{\text{Measured}}{\text{Safe Limit}}$$

River	Measured	Safe/Standard Limit	Ratio (Danger Score)	Interpretation
Yamuna	9.2×10^6 MPN/100 mL	500	18 400	Extremely unsafe
Citarum	9.8 mg/L	3 mg/L	3.3	Moderately unsafe
Amazon	2.5 $\mu\text{g/g}$	1 $\mu\text{g/g}$	2.5	Unsafe
Techa	10 Bq/L	0.1 Bq/L (IAEA drinking-water guideline)	100	Highly unsafe

Table 2.0 Showing the measured values, the standard limits, the danger scores and their interpretations

River	Dominant Hazard	Relative Severity	Main Health / Environmental Risks
Yamuna	Pathogenic contamination	Highest ($\approx 18\,400 \times$ limit)	Gastrointestinal diseases, cholera, ecosystem collapse
Techa	Radioactive contamination	Very high ($\approx 100 \times$ limit)	Long-term cancer and genetic risks
Citarum	Chemical + organic load	Moderate ($\approx 3 \times$ limit)	Oxygen depletion, heavy-metal toxicity
Amazon	Mercury bioaccumulation	Moderate ($\approx 2.5 \times$ limit)	Neurotoxicity via fish consumption

Table 2.1 Showing detailed risks and dominant hazards

The Visualization Problem: Logarithmic Transformation

Since the rivers are discrete, we decided on a standard linear bar chart, but were faced with a major issue: Yamuna's danger score was several orders of magnitude higher than those of the others, making the other bars

effectively invisible. This was where a logarithmic bar chart came into question:

- It efficiently represented the data and eliminated the visualisation problem
- It successfully highlighted the differences between the danger indices of the different rivers

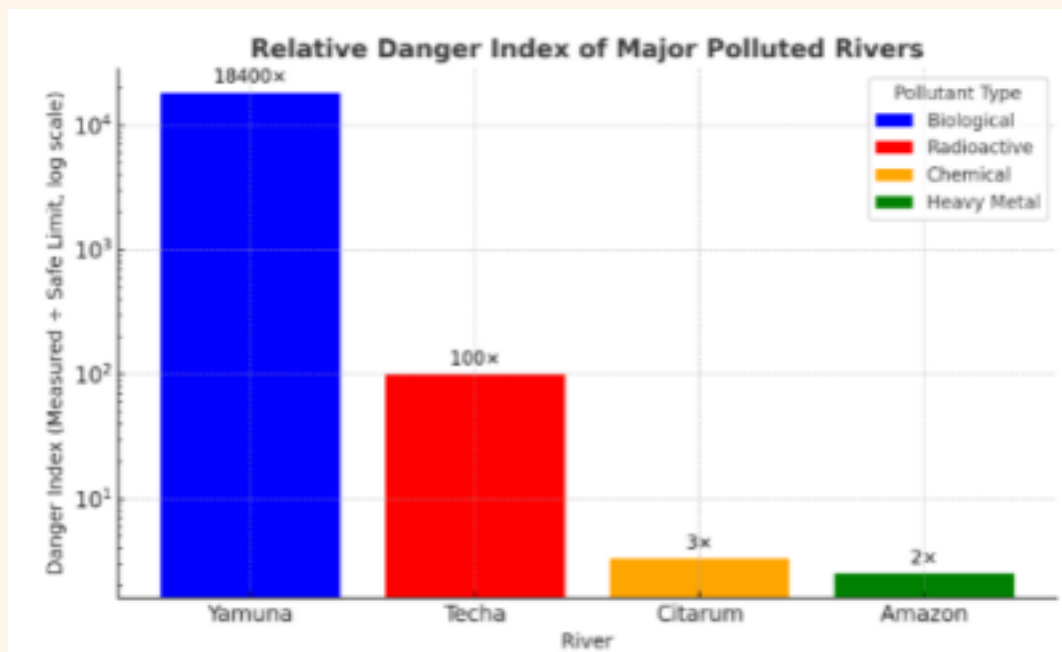
So, we calculated the logarithmic values of the danger indices:

River	Exceedance Ratio (\times Safe Limit)	Calculation	$\text{Log}_{10}(\text{Value})$
Yamuna	18,400	$\text{log}_{10}(18,400)$	4.26
Techa	12	$\text{log}_{10}(12)$	1.08
Citarum	3	$\text{log}_{10}(3)$	0.48
Amazon	2	$\text{log}_{10}(2)$	0.3

Table 3.0 Showing the calculations of the log values for the plotting of the table

Logarithmic scale is a fundamental tool when data spans across multiple orders of magnitude. In this case, we put the logarithmic values in the Y axis, which implies that for each unit increase in the Y-axis, there has been a tenfold increase in the actual value of the danger index.

The resulting logarithmic bar chart:



Graph Showing Relative danger index of four major rivers

Graph Details:

X-axis: Rivers (Yamuna, Techa, Citarum, Amazon)

Y-axis: Relative Danger Index = Measured Pollution/Safe Limit

Scale: Logarithmic (base 10) on the Y-axis

Results: When Algorithms Meet Reality

Finding 1: Metric Selection Shapes Perception

It might appear that Yamuna is the most dangerous and polluted out of the four mentioned rivers, however, it must be noted that Techa remains the most dangerous to living organisms and Citarum the most polluted of them all.

Our logarithmic bar chart revealed a stark hierarchy with Yamuna at the highest level: Data visualisation strongly influences interpretation, since taller bars appear more dangerous even when they mainly reflect algorithmic choices.

The Yamuna's extreme score results from the dominance of biological pollution over other types of pollution in Yamuna. The reason this does not represent the objective truth but appears to is that:

- Radioactivity, even in smaller amounts, can be more lethal than biological pollutants, hence making Techa more dangerous than any of the other rivers mentioned here
- The Citarum is more polluted than the Yamuna or any of the other mentioned rivers because its chemical pollution makes it toxicologically more severe, and because the Citarum contains a wide variety of pollutants, making it difficult to determine a single danger index based on one type of pollutant unlike the Yamuna, where pollution is dominated by sewage and bacterial contamination

Finding 2: Dimensional Reduction Creates Information Loss

The Citarum River presents our algorithm's most significant limitation. In our danger score ranking, Citarum appears to have moderate danger, yet when we look into our exploratory pie chart, it is revealed that Citarum suffers from the most diverse pollution profile of the four rivers in question with significant contamination across multiple categories. This causes a misrepresentation of the Citarum in the final bar chart. Hence, it is observed that a single danger score enables easy comparison and decision making through simplicity, but falls behind due to a lack of preservation of the full complexity that tends to resist simple interpretation.

Finding 3: Logarithmic Visualisation Distorts Intuition

The logarithmic transformation was a necessary portion of our project, but it risks systematic misinterpretation due to the visual perception of those unfamiliar. A bar twice as tall is considered twice as important, but in the case of a logarithmic bar chart, a bar of height 4 units is 100 times a bar of height 2 units.

Statistical Implications and Algorithmic Trade-offs

Our project serves as a case study in how statistical and algorithmic choices shape our understanding of reality. Each decision point in our analysis represented a trade-off:

While normalisation enables comparison, it requires the choosing of a reference point: We chose to make a danger index using safety limits; other choices, like historical values and ecosystem thresholds, shall produce different results.

Dimensional reduction may provide simplicity, but causes information loss: Our most-dominant-pollution-type algorithm misses multi-dimensional complexity. Alternative methods may preserve more information but resist simple interpretation.

Logarithmic scaling reveals patterns across magnitude ranges but creates issues in visual interpretation: The transformation of the

danger index into logarithmic values promotes easier representation but risks a wrong impression of the values to those not familiar.

Decoding Reality Through Statistical Choices

This project started as an attempt to compare pollution across rivers that could not be directly compared, and later evolved into how statistical choices affect the final interpretation. Each step of the analysis involved decisions that affected which results became visible.

From choices about where to collect our data from, how to normalise the values, which aspects to consider in analysis to the choices about how to represent our data, every choice encoded assumption about what mattered, leading to the final clean, simple, and apparently objective bar chart that was actually a compressed representation of these layered decisions.





THE VETERAN VOICES

*“The teachers’ heart is a cradle of knowledge,
where young minds learn to dream, question
and grow.”*

— Rabindranath Tagore

The Veteran Voices section brings forward the wisdom and experiences of two highly respected senior professors whose years of dedication have shaped generations of students. Through their insightful interviews, readers gain a glimpse into their academic journeys, personal reflections, and the lessons they have gathered over decades of teaching and research. Their words offer guidance, inspiration, and a deeper appreciation for the values that define true scholarship.

*“Years of wisdom, shared for the minds of
tomorrow.”*



Prof. (Dr.) Manisha Pal

Professor (Dr.) Manisha Pal is a distinguished statistician and academic leader whose illustrious career spans over four decades of excellence in research, teaching, and institutional service.

She served the Department of Statistics at the University of Calcutta for more than 39 years, including two terms as Head of the Department, and was the Convener of the Ph.D. Research Advisory Committee in Statistics for five years. She later continued at St. Xavier's University, Kolkata, serving as Senior Professor until July 2024 and currently as an Adjunct Professor. Her academic presence extends internationally through invited lectures, collaborative research, and conference participation across India, the United States, the United Kingdom, Thailand, Bangladesh, and Nepal.

Professor Pal has authored over 135 research publications in reputed national and international peer-reviewed journals, including *Statistics*, *Communications in Statistics*, *Statistics and Probability Letters*, *Journal of Applied Statistics*, *Opsearch*, *Fertility and Sterility*, and the *Journal of Medical Virology*, reflecting the interdisciplinary depth of her work. She is a co-author of the Springer monograph *Optimum Mixture Experiments* in the *Lecture Notes in Statistics* series and has contributed chapters to edited volumes published by Springer honoring Professor C. R. Rao and commemorating P. C. Mahalanobis. Her research spans optimal experimental design, mixture experiments, reliability theory, operations research and applied statistical modeling.

Her professional service includes roles as an Expert Member of UPSC recruitment boards and association with the Staff Selection Commission. She has received the Siksha Ratna Award from the Government of West Bengal and the International Distinguished Research Excellence Award, and currently serves as President of the Calcutta Statistical Association, editor of leading statistical journals, Fellow of the West Bengal Academy of Science and Technology, and Executive President of the Society of Statistics and Computer Applications in New Delhi.

In Conversation with Prof. (Dr.) Manisha Pal

Interviewer:

Ma'am, when you first began teaching, data was limited and carefully curated. Today, data is abundant and often messy. How has this shift changed the way we should think about statistics?

Prof. Pal:

When I began my career, data was indeed limited. Because of that, we relied heavily on theory. The statistical tools we used were grounded firmly in theoretical foundations and we applied them carefully to the small datasets available.

Today, the situation is quite different. With the abundance of data, exploratory data analysis (EDA) has become extremely important. At the very first stage, EDA provides deep insight into the structure and nature of the data. Earlier, we often had to make assumptions such as normality without having enough data to thoroughly verify them. Now, with larger datasets, we can check these assumptions much more effectively.

This shift makes hypothesis testing and estimation comparatively simpler because we already have a clearer idea of the data from the exploratory analysis itself. It also helps us decide what to test and where to place emphasis. That clarity was not always possible earlier. Additionally, the development of AI and online platforms has made data widely accessible. Websites like Kaggle provide numerous datasets for practice and research. However, while such sources are useful, one must always be cautious about reliability and data quality.

Interviewer:

Statistics deals with uncertainty rather than certainty. Why do people struggle to accept probabilistic answers in a world that expects definite outcomes?

Prof. Pal:

I would not say people completely struggle with uncertainty. Everyone understands that there is always some chance associated with events. However, in practical situations, people expect definite answers. For example, if someone is told that a machine will function for a certain number of years, they interpret that as a guarantee. If it fails earlier, they feel dissatisfied. But statistical statements always include an element of probability. There is always a chance that outcomes may differ. Whether people like it or not, uncertainty is inherent in life. Statistics simply quantifies that uncertainty.

Interviewer:

Many modern models produce impressive numerical results, yet their assumptions often go unquestioned. How important is it to understand what lies beneath a statistical model?

Prof. Pal:

It is extremely important. If the assumptions underlying a model are not satisfied, then the results obtained will not be reliable. Assumption verification is essential.

However, in reality, especially when dealing with human subjects, it is not always easy. People may refuse to participate, may not provide accurate information or sampling methods may not be perfectly followed. In such cases, deviations occur. Nevertheless, checking assumptions remains crucial. Without that, the “most probable” answer we provide may not truly be dependable.

Interviewer:

From your academic experience, where do errors in statistical reasoning most commonly arise—data collection, analysis or interpretation?

Prof. Pal:

Errors most commonly begin at the stage of data collection. If the personnel collecting data are not properly trained, mistakes will occur at the source and those errors will propagate throughout the analysis. The second stage is analysis. If assumptions are not checked before applying statistical methods, incorrect conclusions may result. The third stage is interpretation. If one is not well-versed in the tools being used, misinterpretation is likely. So, errors can arise at all three stages, but faulty data collection is particularly critical because it affects everything that follows.

Interviewer:

As data increasingly influences public narratives, how can statisticians ensure that numbers inform rather than mislead society?

Prof. Pal:

Statistics always contains an element of uncertainty. Therefore, one cannot guarantee absolute correctness. However, the way results are presented matters greatly. People tend to believe in numerical evidence: historical data, performance records, continuous data. So, if findings are presented responsibly and transparently, they can positively influence understanding. Ultimately, it depends on how carefully data is collected, analysed and interpreted. While we cannot “fully ensure” that numbers will never be misused, responsible statistical practice can significantly reduce the risk of misleading narratives.

Interviewer:

What mindset should students develop to use statistics not just as a tool, but as a way of thinking about reality?

Prof. Pal:

Statistics is the backbone of data science and data analytics. While students today can use advanced tools and software without deeply understanding the theory, interpretation becomes difficult without statistical knowledge. Students should develop a mindset that values statistical foundations. They must understand why a particular tool is used, what assumptions it relies on and how results should be interpreted. Statistics is not merely a computational technique; it is a structured way of reasoning about data and uncertainty. That perspective is essential for anyone working with data.

Interviewer:

We are aware that you have more than 135 research publications. Among your many interests of research areas, which one have you enjoyed the most?

Prof. Pal:

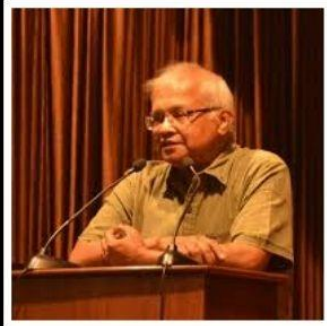
Yes, I have more than 135 publications. The area I particularly enjoyed working in is the design of mixture experiments. Mixture experiments are especially relevant in pharmaceutical research. Consider an oral drug capsule composed of several components. Instead of focusing on the absolute quantities, we are interested in the proportions of these components. Since proportions lie between 0 and 1 and are uncountably infinite, it is impossible to test every possible combination. Therefore, we select a finite number of optimal sets of proportions using specific optimality criteria. Experiments are then conducted using these selected combinations. The results help in building predictive models and determining the best formulation before releasing the product into the market. I found this area fascinating because it focuses on how to collect data optimally, which directly impacts the quality of conclusions drawn.

Interviewer:

Thank you for sharing your insights and experiences with us today. Your reflections remind us that statistics is not merely about numbers or models, but about disciplined thinking, intellectual honesty and responsible interpretation in an increasingly data-driven world. Your journey as a teacher, researcher and academic leader continues to inspire students and scholars alike. We are grateful for your time and for the clarity you have brought to some of the most important questions facing the field of statistics today.

◆

Interview conducted by: Ms. Ushosee Mukherjee & Ms. Swastika Hui
M.Sc. Statistics (2024-2026), Sem-IV



Prof. (Dr.) Sugata Sen Roy

Professor Sugata Sen Roy is a distinguished statistician and academic leader whose career has significantly shaped modern statistical research and education.

He joined the Department of Statistics at the University of Calcutta in April 1989 and served the institution for over three and a half decades, superannuating as Professor in August 2025. During his tenure, he also held the position of Head of the Department, contributing not only through scholarship but also through academic leadership and institutional development.

His areas of specialization include Time Series Analysis, Regression Analysis, Survival Analysis, Econometrics, Development Statistics, Functional Data Analysis, and Applied Multivariate Analysis. His work on asymptotic theory in autoregressive and ARMA models has strengthened the theoretical foundations of time series analysis, while his contributions to survival and recurrent event modeling have found meaningful applications in health sciences and epidemiology.

With publications in reputed international journals such as *Communications in Statistics*, *Journal of Time Series Analysis*, *Statistics in Medicine*, and the *Journal of the Royal Statistical Society, Series C*, his research reflects both rigor and relevance. His collaborative projects span diverse domains—from environmental health and arsenic exposure studies to advanced forecasting models using deep learning—demonstrating a rare ability to connect mathematical theory with real-world challenges.

He has also held visiting positions at the University of Nebraska–Lincoln, USA, and Monash University, Australia, and has been associated with premier institutions in India such as the Indian Statistical Institute and the Indian Institute of Management. As a mentor and researcher, his influence extends beyond publications, shaping generations of statisticians who value both analytical depth and practical impact.

In Conversation with Prof. (Dr.) Sugata Sen Roy

Interviewer:

The theme of our inaugural issue is '*Decoding Reality*.' From a theoretical standpoint, do you believe statistics discovers truth, approximates it or constructs it?

Prof. Roy:

Okay, to be honest, let me distinguish between Mathematics and Statistics. So, mathematics is deductive by nature, so, everyone talks about mathematics being the mother of all sciences. Mathematics doesn't need any other sciences, it can work on its own, other sciences need mathematics. Whereas for statistics, it is necessary to have some other science. Basically, it serves all the other sciences and, in this case, it is imperative that there is some other science from which the data comes, because statistics is related to the data and science actually does the experimentation and you get data from a particular branch and then you analyse it. As a Statistician, you should be careful. Or in other words, you should be acquainted with the subject from which the data comes. That is one more difficulty that we have in Statistics. Thus, if you are a bio-statistician you need to know their problems, if you are an econometrician, you need to know the problems of economics, etc. It has to be related to the given subject and that way, Statistics is more data-oriented and more inductive by nature rather than deductive as in the case of mathematics.

Interviewer:

As someone who has witnessed the evolution of statistics from classical inference to algorithmic modelling, what do you think is the next fundamental shift in the discipline?

Prof. Roy:

It's not fundamental; it's rather a reorientation of the roles in the sense that statistics is somewhat more esoteric. Most statisticians are more involved in doing theoretical research which is very often devoid of any relation to data. So, it has to be data-driven and if you permit me to go a little bit beyond that, this is where data science has taken over. If you actually look at statistics itself, it is basically concerned with data science, it's concerned with data and how to analyse data. Data science is taking over simply because statisticians are not doing their jobs. Statisticians have become more theoretical and are more concerned with mathematics. As I said, mathematics is deductive so if you have a theorem and you prove it then that's fine, but if you are analysing data, each data has its own story. There is no uniform or one single methodology and it doesn't sound so esoteric in that sense. Therefore, in my opinion, probably statisticians are forgetting that they are primarily data analysts.

Interviewer:

Algorithms are designed to optimize specific goals. How does the choice of what to optimize influence real-world outcomes, often in unintended ways?

Prof. Roy:

In a sense, exactly what I was saying! Obviously, it has to be related to a subject you are studying. So, related to the data that is generated from a given area, first of all, you need to know the area very well. Next, you need to relate the basic statistical methods that you have studied and improvise on it. As I always say, that there is no one statistical rule. So, when you face a data, you need to have a basis and you need to change it according to the given situation, according to the data or according to the experiment in hand and then maybe analyse the data, so that actually is an improvisation. As statisticians, we need to improvise a lot more if there is no one single result.

Interviewer:

Many people interact with algorithms daily without realizing it. Do you think this invisibility increases their power or their risk?

Prof. Roy:

It is obviously a risk, because you don't really understand what you are doing, that is to say, you know a technique but you also need to know when and where to apply the technique. This is what I always keep saying: when you make a shirt, you don't first make the shirt and then try to wriggle into it, it's the other way around. You always buy the shirt according to your figure! So, what we try to do in most cases is that we have a methodology and then we want to fit an algorithm and we try to fit known algorithms for the data that we're facing, we cannot do that.

Interviewer:

For students entering the field today, should they focus more on mathematical rigor or computational skills, and how should they integrate the two?

Prof. Roy:

I think it should be a combination of both. Mathematical rigor is required but the algebra is not required. Because, I think when students come after class XII, they know enough algebra to do the mathematics by themselves. The important thing to learn is the ideas, it is important to understand how the subject has been developed, get an idea about what exactly are you trying to do. Because you can always find the formulas, they are always there in the books.

There is one thing that I believe in, what is there in the book you need not memorize it, because once you pass out the books will always be available. Hence what is written in them, you don't have to mug them up. What you actually need to do is to understand what the results are and how you interpret the results, as these are things which will not be written in the book, so that is the part that you should really focus on, to understand the ideas and how and where you should apply those.

Interviewer:

Algorithms learn from past data, but society constantly evolves. How do we prevent systems from being anchored to outdated or biased patterns?

Prof. Roy:

You have to constantly keep evolving. There is always a basic standard methodology to proceed but as statisticians, as you go along you have to keep changing that, building on it or you have to make some sort of small changes to accommodate the dataset that you have or the study that you are conducting. You have to be continuously flexible in your approach because as I said every data has its own story to tell; and if the story keeps changing, you have to modify your results accordingly. There exists no one liner, that's where the difference with mathematics will come in, there is no single hard and fast rule that you have to follow, you have to keep improvising!

Interviewer:

From your perspective, what is the most misunderstood aspect of algorithmic decision-making among non-technical audiences?

Prof. Roy:

The problem primarily is that the non-technical audience always come to statisticians *after* they have collected the data. So, the problem should be to consult a statistician from the very beginning. An important aspect that even statisticians miss out: when we teach in class, we talk a lot about inference, right? But to infer well, you need very good data, that is, a very representative sample from the population that you are studying. The two subjects that students almost don't look at are **Sample Survey** and **Design of Experiments**; but before you plan out an experiment or before you actually do the analysis you have to have a good data. The non-statisticians generally come to us after collecting the data and very often it is not collected in a manner which would actually yield the results that they are seeking and that, in my experience, is the major problem that I've seen. It is a poorly collected data and you can't get any answers out of it.

Interviewer:

As algorithms begin to assist or replace human judgment, where should the boundary between automation and human oversight lie?

Prof. Roy:

Automation can only lead to a given extent. Beyond that, the interpretation obviously has to be by a human and I would more importantly emphasize on how you apply the automation. Whether it is at the beginning or the end, because when you use an algorithm, which algorithm to use? That discretion itself is important. The chosen algorithm should be according to the problem and to match this perfectly, you need human expertise. Once you have got the results, you have to relate it back to the problem, so that again needs human expertise. I think that would be the difference. Automation can do the middle part but to start the problem, to actually identify and engage in the problem and finally, to interpret the problem, these two areas cannot be automated.

Ultimately, remember one thing: there would always be changing situations, to account for that human judgement can't be automated. They talk about scenario forecasting these days but it can't be all possible from automation and human intervention is absolutely needed.

Let me put it in two separate issues. First issue is who is making the black box, obviously some statistician that you don't see. The algorithms like PCA, somebody has to build it in, so there is some human input into the black box, that's one part of it. The other part is when the black box gives you the results. So, who is to decide which one you go for when doing the PCA and analysing the components, that is human decisions, and that can't and shouldn't be decided by algorithms.

Interviewer:

If you were to redesign the statistical curriculum today to prepare students for an algorithm-driven world, what core theoretical pillars would you insist on preserving?

Prof. Roy:

I wouldn't actually go by that. The theoretical pillars are fine, what I had been thinking for the last 15-20 years that we haven't been able to employ is that the teaching has to be different. In the sense that I would rather prefer that you don't have practical classes separately, it should be integrated with the theoretical classes. If I want to teach a particular theory, I should be able to show the results through data; that can actually help students visualize the concept and it is easier to grasp. This enforces a better relationship between what you are learning theoretically with what you are doing in practice. Look at statistics these days for example, things can be shown and it can be done in the classroom. Teaching has to be more oriented towards practical applications; the theoretical distinction has to be lessened and if the two can be merged that's the best possible way in my opinion.

To be honest, the transitions we have witnessed has been huge! Our generation probably has seen a lot of changes, but your generation is much more stable and are more equipped with these sorts of things. When we were students in the '90s, we used to use the facet machines for calculations, so it has come from that to scientific calculators and all kinds of algorithms now!

Interview conducted by: Ms. Aditi Shah, Mr. Akashdeep Das & Mr. Anirban Nath
M.Sc. Statistics (2024-2026), Sem-IV



FUN ZONE

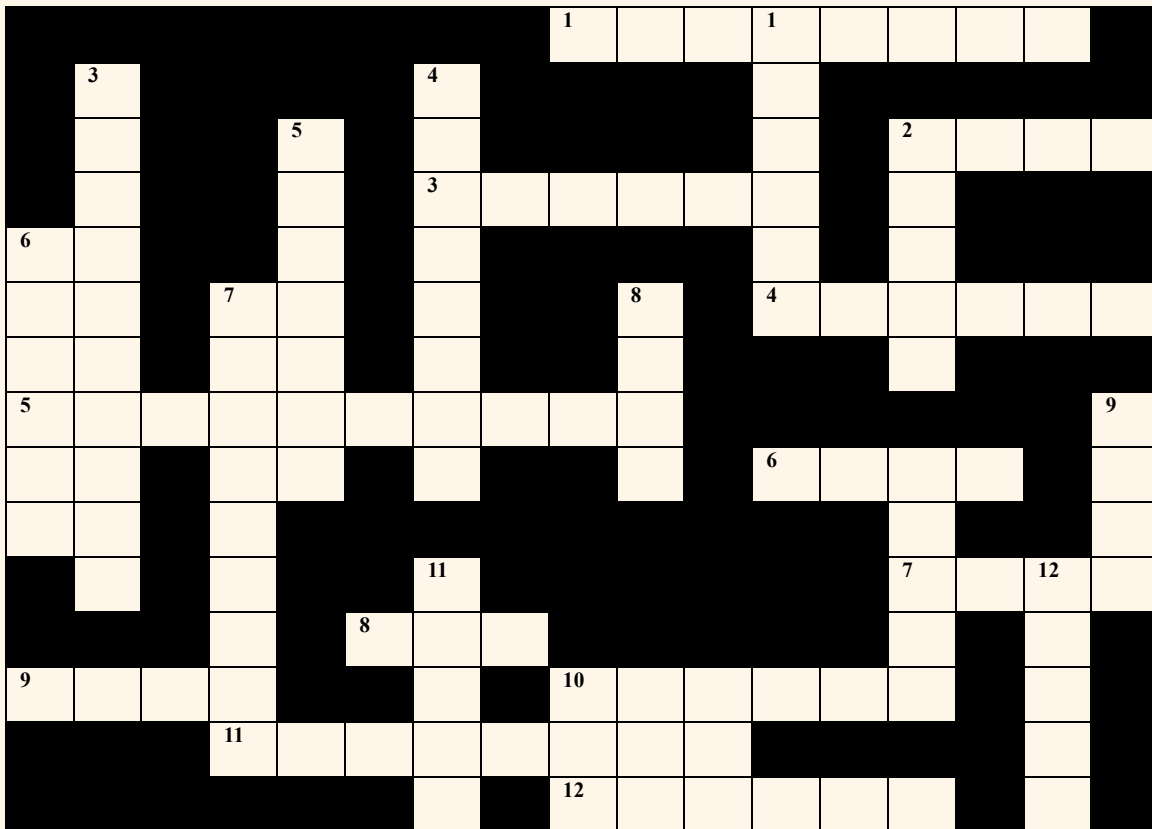
“Laughter is timeless, imagination has no age, and dreams are forever.”

— Walt Disney

The Fun Zone adds a touch of laughter and lightheartedness to the magazine. This section features funny memes, word searches and engaging puzzles designed to entertain and refresh the mind. It offers readers a playful break while encouraging creativity, curiosity and a little bit of fun along the way.

“A little fun makes every page brighter.”

Crossword



ACROSS

1. The process of selecting a sub-group from a larger group/population
2. The average value of a set of numbers
3. Without a fixed pattern
4. Patterns that show which way things are going
5. A statistical method used to study relationships between variables
6. When data leans unfairly in one direction
7. The chance of loss or uncertainty
8. A “worker” that never gets tired
9. A connection point in a network or system
10. A world with only two choices
11. What’s left after the model has its say
12. A small technical problem or error

DOWN

1. What you give an AI before it answers
2. A system built to explain or predict
3. An educated statement waiting to be tested
4. Data’s best guess about the future
5. Opposite of cluster
6. Brain-inspired computing helpers
7. A value that controls how a system behaves
8. A position or order in a list
9. A quick or clever technical fix (sometimes informal)
10. A flaw that breaks code; makes software misbehave
11. Rule-based reasoning
12. A data structure where the last item added is removed first

- Ms. Aditi Shah
M.Sc. Statistics, Sem – IV

Word Search

Find 15 words related to Machine Learning.

X R W G M S A T L D O F I N T K L G O P
V U Q H K M E A N S L D R E R H N M I R
K N J L Q E W R L U Z E W L A G H A K E
B D B Y M J H K B P X C P D I I N C L C
U E F K O H L J Q E W I Q T N M L H G I
E R A N D O M F O R E S T K I J X I L S
G F P Q E N Q T L V H I G B N I N N L I
I I P L L W H M T I N O K G G N Q E H O
M T K X L F S S H S K N L J B K W L B N
B T S D I S P B K E T T G M G L P E L M
Z I N A N K Y M W D F R A C C U R A C Y
J N J T G M T G E K M E D P N P N R K K
K G L A D B H A R R T E M T H W K N F O
O J I B H S O S G L L W M L T G L I H T
M L O A M T N X P R E D I C T I O N B E
H B L S K M L T K N S Z Q Q M H Q G C T
M D S E K P M H L H G F J X Y K F D X R
I G G V O V E R F I T T I N G L H F E G
R E G R E S S I O N M B R Z L D R V W H
O H A Q E A Y P S L R L M N V H B W Q N

- Ms. Mukta Mahato
M.Sc. Statistics, Sem – IV

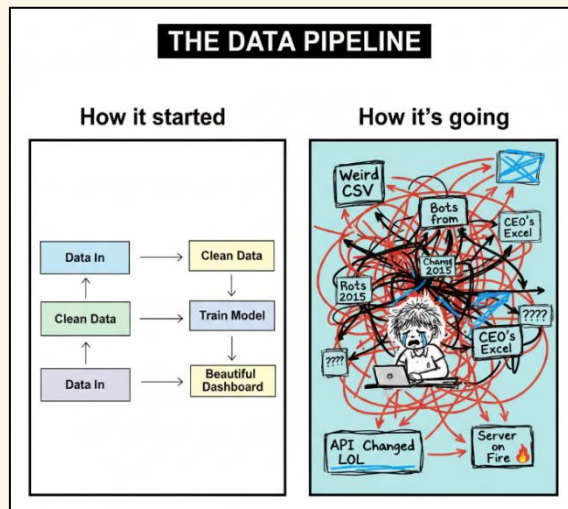
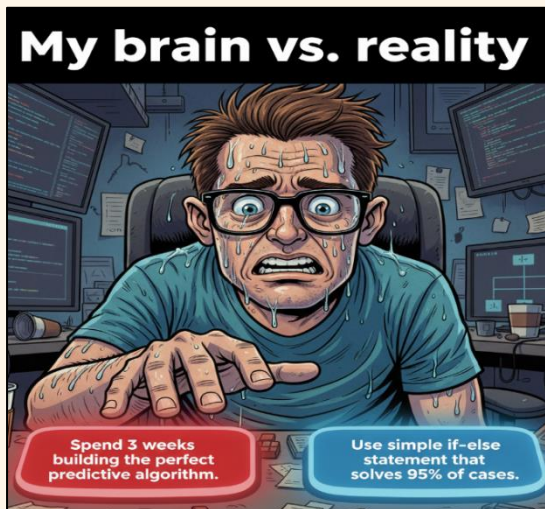
Underfitting
Overfitting
Machine learning
Prediction
Random forest

Precision
Accuracy
Regression
K-means
Decision tree

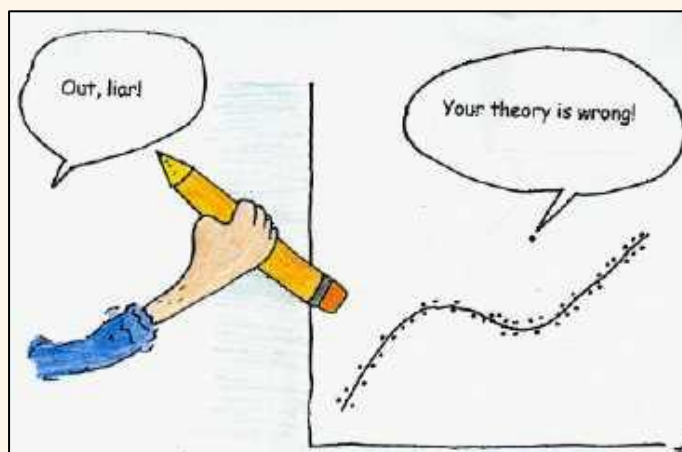
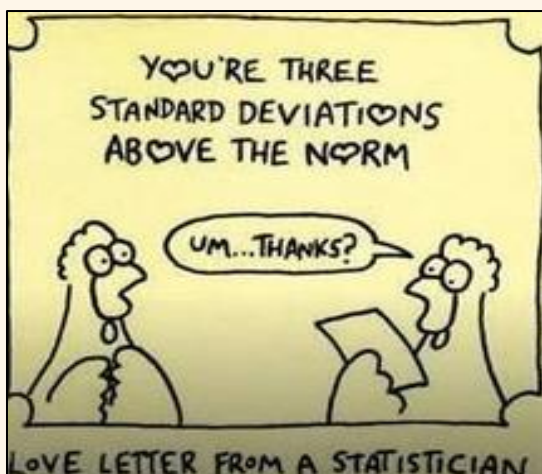
Modelling
Database
Python
Supervised
Training



Memes



Mr. Sreesant Saha
B.Sc.(Hons.) in Statistics & Data Science, Sem – II



CROSSWORD ANSWERS:

1. Prompt, 2. Model, 3. Hypothesis, 4. Forecast, 5. Scatter, 6. Neural;
7. Parameter, 8. Rank, 9. Hack, 10. Bug, 11. Logic, 12. Stack

DOWN

1. Sampling, 2. Mean, 3. Random, 4. Trends, 5. Regression, 6. Bias;
7. Risk, 8. Bot, 9. Node, 10. Binary, 11. Residual, 12. Glitch

ACROSS



ART & PHOTOGRAPHY

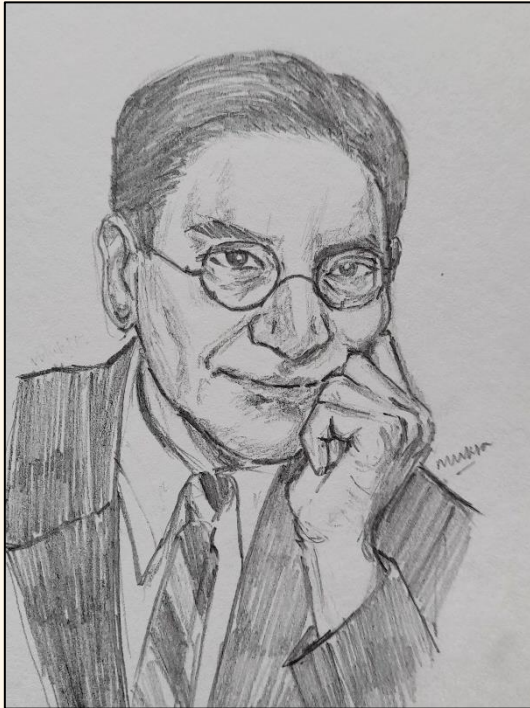


*“Art is not what you see, but
what you make others see.”*

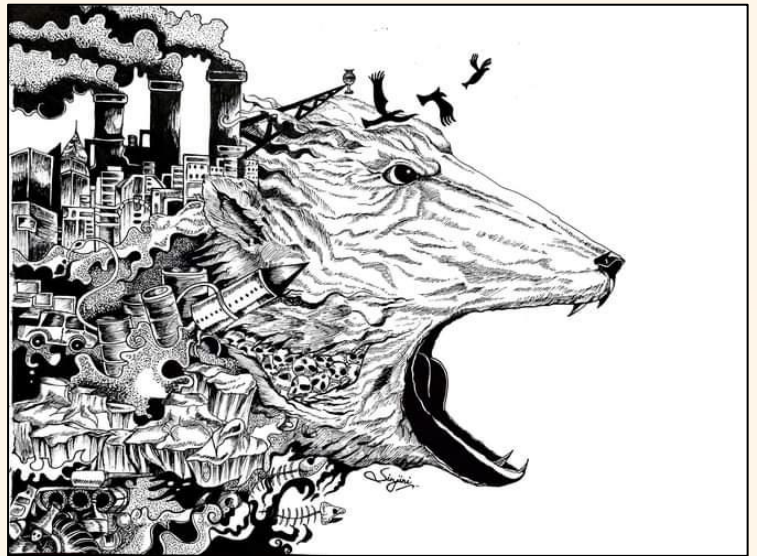
— Edgar Degas

The Art and Photography section celebrates the visual creativity of our professors, alumni and students. Each piece reflects a unique vision and artistic voice, offering readers a glimpse into the imagination, talent and creativity.

“Every artwork tells a story beyond words.”



*A Sketch of P. C. Mahalanobis
Ms. Mukta Mahato
M.Sc. Statistics, Sem-IV*



*Innovation in Its Fierce Nature
Ms. Sinjini Pal
M. Sc. Statistics, Sem-II*



*The Defiant Smile
Mr. Anish Barik
M.Sc. Statistics, Sem-IV*



*The Algorithmic Blueprint
Ms. Aditi Shah
M.Sc. Statistics, Sem-IV*



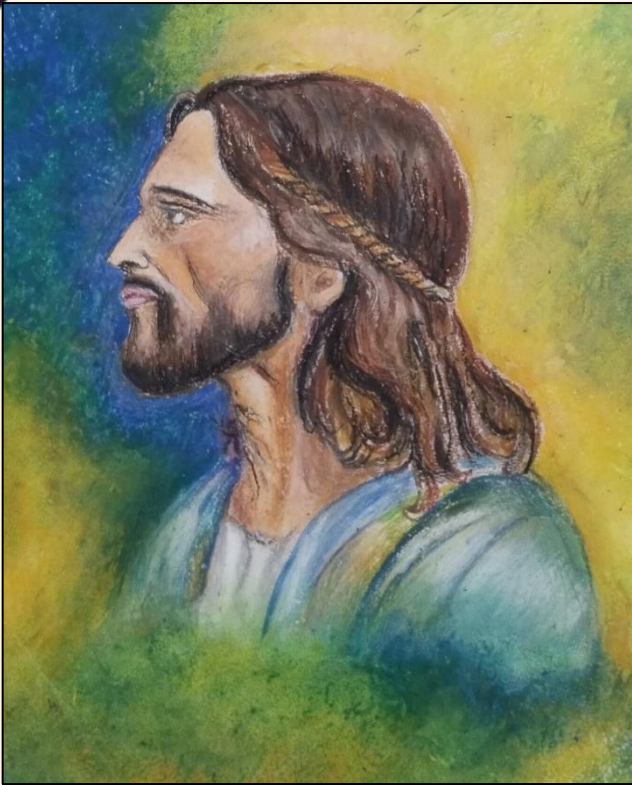
The Starry Nights
Ms. Paulami Dutta
B.Sc. Statistics & Data Science, Sem-II



The God of Football
Ms. Ushosee Mukherjee
M.Sc. Statistics, Sem-IV



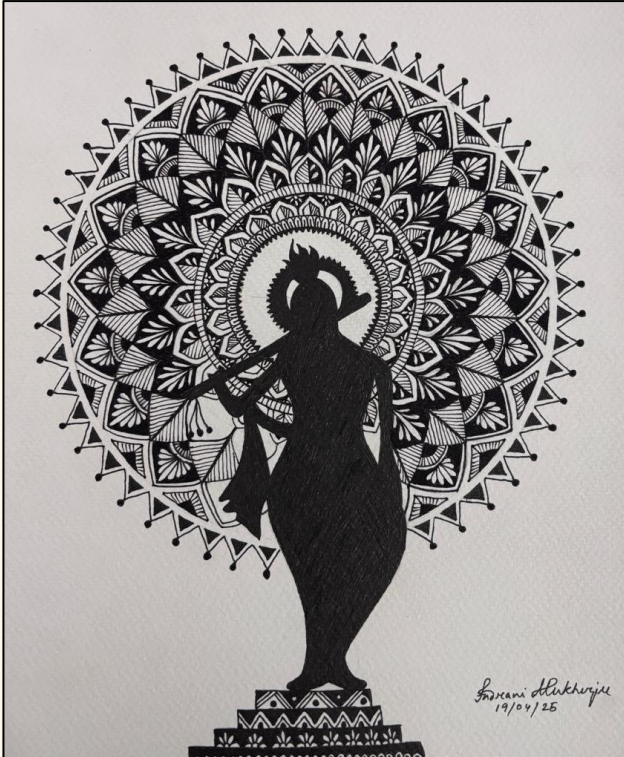
Nanami Kento
Mr. Anugraha Bar
B.Sc. Statistics & Data Science, Sem-II



A Portrait of Jesus
Ms. Soumili Chowdhury
M.Sc. Statistics, Sem-IV



Pencil Sketch: Frog
Mr. Anugraha Bar
B.Sc. Statistics & Data Science, Sem-II



The Radial Balance
Dr. Indrani Mukherjee
Assistant Professor, Department of Statistics



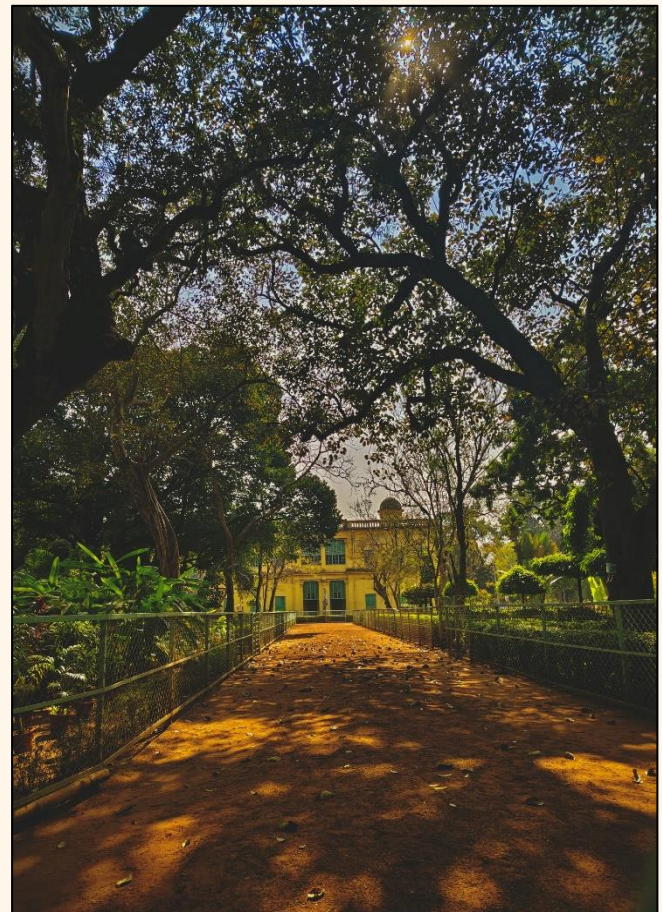
Kinnara: The Celestial Bard
Surath Chakraborti
M.Sc. Statistics (2022-2024)



Serenity
Mr. Soham Manna
M.Sc. Statistics, Sem-IV



Poetic & Mystical
Ms. Trisha Patra
B.Sc. Statistics & Data Science, Sem-II



Sun-lit Morning
Ms. Swagata Barman
B.Sc. Statistics & Data Science, Sem-II



The Fibonacci Circuit
Ms. Smriti Chauhan
M.Sc. Computer Science, Sem-II



Architecture Photography, Fatehpur Sikri
Ms. Trisha Patra
B.Sc. Statistics & Data Science, Sem-II



Howrah Bridge
Ms. Aditi Shah
M.Sc. Statistics, Sem-IV



GALLERY

“Photography is the story I fail to put into words.”

— Destin Sparks

The Gallery section captures the memorable moments and vibrant experiences from various department events. Through these photographs, readers can revisit celebrations, achievements, and the shared memories that bring the academic community together. Each image reflects the spirit of participation, collaboration, and the lively atmosphere that defines our department.

“Moments captured, memories preserved.”

CORE COMMITTEE



USHOSEE MUKHERJEE
SECRETARY



DIPANJAN GANGULY
ASSISTANT SECRETARY



AKASHDEEP DAS
TREASURER

WORKING COMMITTEE



ADITI SHAH



ANIRBAN NATH



JIGESAA CHAKRABORTY



SWASTIKA HUI



RUPAK BERA



SMRITI CHAUHAN



SIDDHARTHA DAS



KRITEE MUKHERJEE

DEPARTMENTAL PHOTOGRAPHS



DEPARTMENTAL EVENTS



CYBER MYTHS OF KOLKATA NIGHT
(AI ART COMPETITION)



DESI DATA DASH
(DATA STORYTELLING
&
VISUALIZATION COMPETITION)



THE COSMIC CONFERENCE
(SCIENTIFIC ROUND - TABLE
CONFERENCE)



CODEMANTHAN
(CHURNING IDEAS INTO
INNOVATION -HACKATHON)

DEPARTMENTAL EVENTS



DEPARTMENTAL EVENTS



DEPARTMENTAL SEMINARS



ST. XAVIER'S UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

INDUSTRY TECH TALK
CONVENTION AND TECHNOLOGY

A story of Humanity's Journey from Classical Limits to Quantum Horizons
DIFFICULTY - DESTINY - DISCOVERY

Lecture

FEBRUARY 2, 2026
Amphitheatre 200 - From 14:30 Onwards

Guest Speaker
Dr. Tamal Chakraborty

St. Xavier's University, Kolkata
Department of Computer Science

Santa Clara UNIVERSITY

4TH INDUSTRY TECH TALK

FROM AUTOMATION TO AUTONOMY
DATA SCIENCE IN THE ERA OF ROBOTS

Tue, 04 Nov 2025
From 11:00 AM Onwards
Room 200 Amphitheatre

About
Prof. Maria Kyrarini (SCU) Assistant Professor directs the HMC Group, specializing in Human-Robot Interaction, Robot Learning, and Assistive Robotics to enhance human-robot performance.

SPEAKER
Prof. Maria Kyrarini

For more information Scan QR

Industry Tech Talk
On the Topic

Data Analytics
Trends & Opportunities

Mr. Sourabh Mukherjee
Head of Data-AI Practice
accenture

4 Sept, 2025
Venue: Amphitheatre 500
From 2:30pm onwards

ANNUAL SPORTS DAY



2025



2026

CAMPUS CHRONICLES





**XAVIER'S SOCIETY FOR
SCIENCE & INNOVATION (XSSI)**

-----PRESENTS-----

XENESIS 2026

FROM THE FUNDAMENTAL STRUCTURES OF ALGORITHMS TO THE COMPLEX LAYERS OF NEURAL NETWORKS, XENESIS SERVES AS A CURATED COLLECTION OF INTELLECTUAL CURIOSITY. THIS VOLUME SHOWCASES THE RESEARCH, INSIGHTS, AND TECHNICAL BREAKTHROUGHS OF THE ST. XAVIER'S UNIVERSITY COMMUNITY—BRIDGING THE GAP BETWEEN THEORETICAL DATA SCIENCE AND REAL-WORLD APPLICATION. WITHIN THESE PAGES, WE DOCUMENT THE EVOLUTION OF CODE INTO CONDUCT AND DATA INTO DISCOVERY.

**"INFORMATION IS THE RESOLVING OF
UNCERTAINTY." — CLAUDE SHANNON**