Kapil Gupta

PhD Candidate, Decision Sciences Area, IIM Bangalore, India
kapil.gupta20@iimb.ac.in • kapil21gupta.github.io/ • Kapil Gupta

Research Interests

House Price Dynamics, Spatial statistics, Spatio-temporal Modelling, Clustering and classification, Sports Analytics, Applications of variable selection methods.

Education

- **Indian Institute of Management Bangalore** — **Bangalore, India**
  Doctoral Candidate in Decision Sciences Area (Cumulative GPA: 3.66/4) — *(Expected) Mar 2025*
  Recipient of the Director's Merit List Award in both the years during the coursework
  PhD Thesis : Spatial and Spatio-Temporal Models in analysing the House Price Dynamics
  PhD Advisor : Dr. Soudeep Deb
  Other committee members : Dr. Venkatesh Panchapagesan, Dr. Kunal Dasgupta, Dr. Rishideep Roy
- **Indian Institute of Technology Delhi** — **Delhi, India**
  Master of Science in Mathematics (Cumulative GPA: 7.52/10) — *Jul 2016–May 2018*
  Master's Thesis : Numerical Solutions of Singularly Perturbed Linear Problems in One Dimension
  Supervisor : Dr. S. Chandra Sekhara Rao
- **Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram** — **Chennai, India**
  Bachelor of Technology in Computer Engineering (Cumulative GPA: 7.55/10) — *Aug 2012–May 2016*
  Bachelor's Thesis : The cd-coloring of bipartite graphs
  Supervisor : Dr. Shalu M A

Experience

- **Indxx, LLC** — **Gurgaon, India**
  Senior Data Analyst, Engineering Division — *Jun 2018–Jul 2020*
  - Responsible for data analyzing and cleansing.
  - Developed and calculated In-house and client based indices using SQL.
  - Created the SSIS packages for loading daily financial data from ICOS (Data provider).
  - One of two project team members chosen to visit the client's office (Qontigo) in London.
- **Indian Institute of Science Education & Research** — **Bhopal, India**
  Summer Research Intern, Department of Mathematics — *May 2014–Jun 2014*
  - Reading project in ordinary differential equations under the guidance of Dr. Ashish Gupta.

Awards and Honors

- **Session Chair (Spatial statistics session)**
  6th International Conference on Econometrics and Statistics, Tokyo, Japan — *Aug 2023*
- **Director's Merit List Award (2020–21 and 2021–22)**
  Indian Institute of Management Bangalore — *2020–2022*
  Best academic performance during the course work of the Doctoral Programme in both the years.
- **Employee of the Year - Entrepreneurship**
  Indxx, LLC — *Nov 2019*

Publications & Submitted Articles

- **Gupta, K.**(2022). An integrated batting performance analytics model for women's cricket using Principal Component Analysis and Gini scores, Decision Analytics Journal. [publication]
  Abstract : Over the years, women cricketers have garnered little media or societal attention. This study aims to quantify women's batting performances in one-day international (ODI) matches. The average batting score has traditionally been used in cricket to determine a batter's worth. However, ranking by simple average ignores two important factors in ODI matches: consistency and strike rates. To address these issues, we propose a novel methodology for calculating ranks

that incorporates Gini-based average scores rather than simple averages and considers batting strike rates in a Principal Component Analysis (PCA) framework. As a result, PCA combined with Gini scores can assist cricket fans, coaches, and managers in better understanding a player's performance.

- Bag, S., **Gupta, K.**, Deb, S. (2022+). A review and recommendations on variable selection methods in regression models for binary data. Under revision in International Statistical Review. [Pre-print]
  Abstract : The selection of essential variables in logistic regression is vital because of its extensive use in medical studies, finance, economics and related fields. In this paper, we explore four main typologies (test-based, penalty-based, screening-based, and tree-based) of frequentist variable selection methods in logistic regression setup. Primary objective of this work is to give a comprehensive overview of the existing literature for practitioners. Underlying assumptions and theory, along with the specifics of their implementations, are detailed as well. Next, we conduct a thorough simulation study to explore the performances of sixteen different methods in terms of variable selection, estimation of coefficients, prediction accuracy as well as time complexity under various settings. We take low, moderate and high dimensional setups and consider different correlation structures for the covariates. A real-life application, using a high-dimensional gene expression data, is also included in this study to further understand the efficacy and consistency of the methods. Finally, based on our findings in the simulated data and in the real data, we provide recommendations for practitioners on the choice of variable selection methods under various contexts.
- **Gupta, K.**, Krishnamurthy, V., Deb, S. (2023+). What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data. Under revision in IIMB Management Review.
  Abstract : This study aims to examine the importance of the game elements of the first set in Wimbledon matches. Our initial exploratory analysis reveals that winning the first set is often the key to winning the match eventually and the behaviour of game elements also change across the rounds. Motivated by our findings, we identify what factors may have the most significant impact in deciding the match outcome based on the first set data only. We propose a least absolute shrinkage and selection operator (LASSO) induced logistic regression model to identify the effects of the variables. The primary analysis suggests that the service points and average distance travelled by a player turn out to be the most significant factors in deciding the match outcome. Further, we find that the association of tennis professionals (ATP) rating points have a significant impact on the match outcome across the tournament. Finally, we show that our proposed LASSO-induced logistic model with a suitable random effect structure can be used effectively for within-match forecasting during the first set of a match. It is found to provide superior accuracy than other popular statistical and machine learning approaches.

Work in progress

- **Gupta, K.**, Deb, S. Efficient Divide-and-Conquer Approach for Spatio-Temporal Modeling of Real Estate Data. Manuscript in preparation.
  Abstract : Statistical research in real estate markets have garnered attention in recent times from the perspective of understanding spatio-temporal dynamics of the house prices. Markov chain Monte Carlo (MCMC) is generally used for Bayesian inference in spatio-temporal modeling. However, standard techniques of MCMC are usually slow for large datasets such as the real estate data, due to the requirement of multiple passes through the entire data in each iteration. This paper proposes a divide-and-conquer spatio-temporal modeling approach to tackle this problem. The method involves partitioning the data into multiple subsets of sufficient locations and utilizing an appropriate Gaussian process model for each subset in parallel. The parameters corresponding to each subset are then combined to obtain the global parameters for the original problem. The proposed methodology allows us to assess the spatially varying impact of various factors on the house-price dynamics. It is also proved to be much faster than a conventional Bayesian approach. As a real life application of the proposed model, we analyze house price data from London from January 2011 to October 2019, covering 53 bi-monthly time points and 906 middle layer super output areas (MSOAs). The results furnish insightful analysis, and render good predictive accuracy, as demonstrated from a cross-validation study. In summary, this work provides a solution to the problem of slow computation in Bayesian inference for spatio-temporal modeling with massive datasets, paving the way for promising future research in this area.
- **Gupta, K.**, Deb, S. A Novel Spatio-Temporal Statistical Model to Analyze Real Estate Market in Bangalore. Manuscript in preparation.
  Abstract : Recent attention has been drawn to statistical research in real estate markets, aiming to comprehend the dynamic patterns of real estate prices in terms of space and time. Our study contributes to this field by presenting compelling evidence of the spatial and temporal dependence in real estate prices. We introduce a novel statistical model that captures the intricate spatial and temporal dependencies of the real estate prices efficiently. Our proposed model employs a separable Gaussian spatio-temporal process, incorporating an additive mean structure and a random error process. We implement our model through a bayesian setup as it gives flexibility and computational advantage over other approaches. As a real life application of the proposed model, we analyze house price data from Bangalore from January 2015 to March 2020, covering 62 monthly time points and 76 micro-markets. The analysis of residual diagnostics confirms that the model effectively captures the spatio-temporal dependence pattern present in the data. Furthermore, the model has been demonstrated to outperform the linear, spatial and temporal models in terms of its predictive capabilities.

Presentation and Talks

- **JOINT EVENT: 2023-ORSI & 2023-ICBAI, IISC Bangalore, India**
  Kapil Gupta, *Soudeep Deb* *Dec 2023*
  "A Novel Spatio-Temporal Statistical Model to Analyze Real Estate Market in Bengaluru, India".
- **Indo-German workshop on Data Mathematics and Scientific Computing , IIT Tirupati, India**
  Kapil Gupta, *Soudeep Deb* *Sep 2023*
  "Efficient Divide-and-Conquer Approach for Spatio-Temporal Modeling of Real Estate Data".
- **Invited Talk: 6th International Conference on Econometrics and Statistics, Tokyo, Japan**
  Kapil Gupta, *Soudeep Deb* *Aug 2023*
  "Efficient Divide-and-Conquer Approach for Spatio-Temporal Modeling of Real Estate Data".
- **IMR Doctoral Conference, 2023, IIM Bangalore, India**
  Kapil Gupta, *Vijayshankar Krishnamurthy and Soudeep Deb* *Feb 2023*
  "What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data".
  - One of the 10 papers accepted out of 98 submissions.
- **Management Doctoral Colloquium Shodh Samagam, IIM Visakhapatnam, India**
  Kapil Gupta *Dec 2021*
  "Measuring Batting Performance in Women's Cricket - An In-Depth Analysis of One-Day International Matches".
- **8th MathSport International Conference, University of Reading, UK**
  Kapil Gupta, *Vijayshankar Krishnamurthy and Soudeep Deb* *Jun 2021*
  "Does the outcome of a tennis match hinge on the opening set? An in-depth analysis of the Wimbledon data".

Teaching Experience at IIM Bangalore

Course Instructor

- *Quantitative Techniques (PGP preparatory course)* *Jun 2023*
  - Overall Rating: 4.5/5 (No. of respondents: 26)
- *R for Data Science (PGPBA preparatory course)* *Jun 2023*
  - Overall Rating: 4.3/5 (No. of respondents: 65)
- *Calculus (PhD preparatory course)* *Jun 2023*
  - Overall Rating: 3.6/5 (No. of respondents: 11)

Teaching Assistant

- *Sports Analytics (PGP)* *Jun–Aug 2023*
- *Multivariate Data Analysis (PGPBA)* *Jun–Aug 2023*
- *Data Science Doctrines: Prediction, Inference, and Causality (PGPBA)* *Oct–Dec 2022*
  - Grade: Excellent
- *Decision Sciences II (PGP)* *Oct–Dec 2022*
  - Grade: Good
- *Decision Sciences I (PGP)* *Jul–Sep 2022*
  - Grade: Excellent

Summer School & Workshop

- **Workshop on Trustworthy AI** **Bangalore, India**
  Microsoft Research, University of Pennsylvania, and Wadhwani AI *Jan 2023*
  - One of the 35 participants selected from a pool of of 150+ student applicants from India.
- **The Summer Institutes in Computational Social Science (SICSS)** **Delhi, India**
  Ashoka University & CSIR-Central Scientific Instruments Organisation *Jun 2022*
  - One of the 25 participants chosen from a pool of applications from scholars all over the world.

Scholastic Achievements

- Secured All India Rank 235 in GATE Mathematics 2018.
- Secured All India Rank 30 in CSIR-NET Mathematics 2017.
- Secured All India Rank 60 in IIT JAM Mathematics 2016.
- Secured All India Rank 268 in GATE Mathematics 2016.

Positions of Responsibility

- **Indian Institute of Management Bangalore** — **Bangalore, India**
  PhD Students Academic Representative — *Nov.2020–Oct.2021*
  - Represented the whole PhD student community at IIM Bangalore.
- **Indian Institute of Technology Delhi** — **Delhi, India**
  MSc Mathematics 2016 batch class representative — *Jul 2016–May 2018*
  - Represented the class of 60 MSc students at IIT Delhi.
- **Indian Institute of Information Technology Kancheepuram** — **Chennai, India**
  Quality Management Service core in annual techno-cultural fest Samgatha — *Jul 2014–Jun 2015*
  - Worked as an organiser for the Samgatha, and led the team of 40 volunteers.

Computer Skills

R, SQL, LaTeX, Python, MATLAB, C, C++, Microsoft Office.

References

Prof. Soudeep Deb
Assistant Professor
Decision Sciences Area
Indian Institute of Management, Bangalore
Email: soudeep@iimb.ac.in

Prof. Venkatesh Panchapagesan
Associate Professor
Finance & Accounting Area
Indian Institute of Management, Bangalore
Email: Venky@iimb.ac.in

Prof. Kunal Dasgupta
Associate Professor
Economics Area
Indian Institute of Management, Bangalore
Email: kunal.dasgupta@iimb.ac.in

Prof. Rishideep Roy
Assistant Professor
Decision Sciences Area
Indian Institute of Management, Bangalore
Email: rishideep.roy@iimb.ac.in