AMARTYA KUMAR MAULIK

+1 (970) 889-9924
amartyakumarmaulik@gmail.com ⋄ akmaulik@colostate.edu

EDUCATION

Ph.D. in Statistics, Colorado State University — Aug 2022 - Aug 2027(Expected)
GPA: 4.00/4.00

M.S. in Statistics, Indian Institute of Technology, Kanpur, India — Aug 2019 - May 2021
CGPA: 9.2/10

B.S. in Statistics, University of Calcutta, Kolkata, India — Aug 2016 - May 2019
CGPA: 8.58/10 | Minor in Mathematics and Computer Science

EXPERIENCE

Graduate Teaching Assistant — Aug 2022 - Present
Colorado State University — *Fort Collins, CO*

- Teaching and grading for STAT 201 Statistics students.

Data Analyst Intern, Intelligent Customer Interactions Team — May 2020 - Jul 2020
Ford Motors (GDIA) — *Chennai, India*

- **Problem :** Addressing of the classification problem of predicting the likelihood of buying various Ford Nameplates at customer level.
- **Major Tasks performed :**
  - Imputing of missing values and binning of categories of categorical variables with high cardinality to keep the dummy variables in check.
  - Performed Under-Sampling, Over-sampling and SMOTE to address the imbalance in the dataset.
  - Concluded that Random Forest Classifier outperformed Multinomial Logistic, Gradient Boost and Decision Tree using Balanced Accuracy Score (55%).
  - Inferred that the separability of the model was better at segment level compared to nameplate level with an increase of 33% in Balanced Accuracy Score which was further improved by Hyper Parameter Tuning.

PROJECTS & RESEARCH

MCMC in Mixture Models | Guide: Dr. Dootika Vats, IITK

- Modeled a simulated statistical distribution by a mixture (or weighted sum) of other distributions using the infinite Dirichlet mixture model without the assumption of a finite number of clusters.
- Methods used : Collapsed Gibbs, Blocked Gibbs and Sliced Gibbs for updating the parameters in R.

Feature Selection and Comparison of classifiers | Guide: Dr. Amit Mitra, IITK

- Used the Mushroom Classification data to see which model and feature selection method gave a better prediction for poisonous mushrooms.
- Methods used: Logistic Regression, Linear Discriminant Analysis classifier, Random Forest classifier , Chi-Square Feature Selection and Mutual Information Feature Selection in Python.
- Concluded that Linear Discriminant Analysis classifier was better in comparison to other models based on both the computational time (37.1 ms) and accuracy (98.64

Time Series Modelling of the Apple Stock Price | Guide: Dr. Amit Mitra, IITK

- Performed removal of stationarity by differencing and testing it's presence using Augmented Dickey-Fuller test

- Analysed the ACF and PACF plots for better insights.
- Obtained the optimal (p,d,q) value for an ARIMA(p,d,q) process by comparing AIC values for a range of models which were shortlisted by checking the p-values of the Ljung-Box test statistic.
- Forecasted with a low Mean Squared Error

Martingales in Discrete Time | Supervisor: Dr. Suprio Bhar, IITK

- Introduced to the measure theoretic approach for Conditional expectation.
- Studied the topics of Martingales, Martingale Convergence Theorem, Martingales bounded in $\mathcal{L}^2$, Uniform Integrability and UI Martingales.
- Learnt about a Lemma regarding Martingale Representation and discrete Black-Scholes formula.

PCA on Directional Data | Guide: Dr. Minerva Mukhopadhyay, IITK

- Focused on principal component analysis (PCA) of directional data to investigate the the direction or angle which best represents the data.
- Discussed some basic concepts related to directional statistics and some distributions namely Uniform Distribution and Von Mises-Fisher Distributions for spherical data as well as Bingham Distributions and Angular Central Gaussian Distribution for axial data.
- Explored the nature of principal components on directional data using simulation studies.

A Psychological Survey among College Students | Undergraduate Project

- Used Likert Scale for ordinal scaling and K-sample rank tests for Umbrella alternatives to deal with the alternative hypothesis in the non-parametric setup.
- Failed to reject the null hypothesis of the first-year and third-year students of undergraduate honours courses in Calcutta tending to be more depressed than the second-year students using method of Combined Hypothesis Testing.

RELEVANT COURSEWORK

Probability Theory, Regression Analysis, Statistical Inference, Analysis of Variance, Time Series Analysis, Applied Stochastic Process, Statistical Simulation and Data Analysis, Statistical and AI Techniques in Data Mining, Asymptotic Statistics, Non-parametric Inference, Bayesian Analysis,Multivariate Analysis, Markov Chain Monte Carlo, Sampling Theory

TECHNICAL SKILLS

Programming Languages	R,Python
Other	LaTeX, JMP

OTHER SKILLS

Painting and playing chess are my favorite leisure activities.