ABHISEK CHATTERJEE

Master of Science, Department of Statistics
University of Calcutta
GitHub

Contact Information

ADDRESS	F30, Anamika Apartment, 26/1/11, Narendra Nagar, Kamarhati, Kolkata, West Bengal, India, Pincode: 700056
EMAIL	abhisekchatterjee108@gmail.com
CONTACT NO.	(+91) 7278668746

Education

08/2019 - 08/2021

Master of Science (M.Sc.) in STATISTICS
Department of Statistics, University of Calcutta, India
Specialization: Nonparametric Inference and Big Data Analysis
CGPA: **6.949** on a scale of 10 (equivalent to **69.7**%)
1st Semester SGPA : 6.64
2nd Semester SGPA : 6.68
3rd Semester SGPA : 7.00
4th Semester SGPA : 7.48

06/2016 - 06/2019

Bachelor of Science (B.Sc.) in STATISTICS (HONOURS)
Asutosh College, University of Calcutta, India
Percentage Obtained: **77.125**%
1st Year Percentage (Hons): 68.5%
2nd Year Percentage (Hons): 79.5%
3rd Year Percentage (Hons): 80.25%
General Subjects: Computer Science and Mathematics

2016

Higher Secondary Education
(West Bengal Council of Higher Secondary Education)
Percentage Obtained: **90**%
BT Road Govt Sponsored HS School

2014

Secondary Education
(West Bengal Board of Secondary Education)
Percentage Obtained: **89**%
Baranagore Ramakrishna Mission Ashrama High School

Professional Experience

09/2024 - Present

Senior Data Scientist (Full-time)
Big Data & Computations
Nielsen

Audience Measurement: As part of Nielsen Media, contributing to the Data Enrichment process of the Panel Crediting method for audience measurement. The Panel Data, which has high coverage, is collected through probabilistic sampling and is representative of the target population complements the Big Data, which is collected by third parties and despite having a much larger sample size, may suffer from limited coverage and may include systematic exclusion or bias. The objectives are to coalesce the two datasets properly to make sure that the resultant dataset remains representative of the target population while illogical tuning events and redundancy is removed. To solve the problem, building robust querying scripts and automating the checking mechanism. Part of the solution also includes visualization of specific audience behaviors. *Tools & Tech: PySpark, PrestoDB, Airflow, Tableau*

| 06/2023 - 09/2024 | AI/ML Computational Science Senior Analyst (Full-time) |
| | **Accenture** |

Generative AI proof-of-concept: The objective of this project was to use open-source Large Language Models to extract business themes from a huge corpus of customer reviews. Leveraged Huggingface library and LLMs like Falcon-7B, Llama-8B to successfully extract themes from customer reviews to produce word clouds and further analysis.

Demand Forecasting & Explainability: Was engaged in sales forecasting for a FMCG market leader client using various demographic, product information and promotion features. Sketched an initial solution to the forecast problem through feature engineering, data analysis and model building. Implemented ML algorithms like Lasso, Adaboost, Random Forest, XGBoost. The first iteration of the forecasts of supermarket chains' sales across 5-6 categories resulted in 75-80% accuracy. *Tools & Tech: PySpark, Azure Databricks*

Aircraft Surrogate Modeling: Worked for an aircraft manufacturing giant to create prediction models in the context of Structural Optimization of different aircraft components. Built end-to-end ML applications for two different components that can predict the reserve factors. The ML models, primarily propelled by residual networks are used as surrogates to bypass a structural analysis tool that generates the reserve factors with simulation but takes a much longer time. Each application can save the client a projected 600K Euros per year in cost and 80% of simulation time. *Tools & Tech : Linux, VS Code, Git, Jupyter Lab, Python, Tensorflow, MLflow, Tensorboard, Dash, REST API*

| 09/2021 - 05/2023 | Business Analyst & Data Scientist Analyst (Full-time) |
| | **BRIDGEi2i Analytics Solutions (acquired by Accenture) & Accenture** |

Marketing Mixed Modeling: Worked for an AlcoBev market leader client in Marketing Mixed Modeling. The objectives of the project were to explain the contribution of different traditional and digital media campaigns along with trade contributions for target & non-target brands in two countries and optimizing spends for further budget allocation. To solve the problem, worked on data preprocessing and used mainly hierarchical linear models to model around 100M+ units of volume across all the brands and helped increasing the ROI of modern digital channels and generated insights towards a more effective budget allocation decision. *Tools & Tech : Jupyter Notebook, Python, MS Excel*

Academic Experience

| 10/2021 - 12/2021 | Research Project - Deep Learning in collaboration with Avirup Das, MSc in Data Science, Chennai Mathematical Institute & Instructor: Dr. Sushma Kumari |
| | **Gaussian Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning** view pdf |

The objective of this project was to investigate the possible advantage of Gaussian Dropout in neural networks over Monte Carlo Dropout in the realm of regularization and representing model uncertainty. With experiments on CO_2, MNIST & monkey classification datasets, we were able to reduce the model training time by 50% on an absolute basis and discovered that the distribution of outputs had less uncertainty, which indicates the vast applicability of Gaussian Dropout in production level ML systems. *Tech: Python, Colab, TensorFlow*

06/2021 - 07/2021	Master's Dissertation Research at *Department of Statistics, University of Calcutta*
	Association of Different Genres of Music and Human Behaviour
	Supervisors: Dr. Bhaswati Ganguli and Dr. Gaurangadeb Chattopadhyay

The idea of the project was to find any association between the most frequent genres of music people listen to and the mental health of them in a particular age group of 18-25 years. Collected data about musical preferences of individuals and answers to mental health questions with the help of two widely used mental health questionnaires, the PHQ-9 & the Perceived Stress Scale. The analyses led to some interesting conclusions on the association of musical preferences and depression & stress levels. *Tools: R Studio, Classification Trees, Inference and Contingency Tables*

Scholarship

2016 - 2020	Department of Science and Technology, Government of India **INSPIRE Scholarship for Higher Education (SHE)**, Stipend Amount: ₹ 80,000 per year

Major Achievements

2021	Ranked 14th (among 38 students) in MSc Examination, Department of Statistics, University of Calcutta
2019	Ranked 8th (among 218 students) in BSc 3 year Honours Examination, University of Calcutta
2016	Marks secured in top 0.6% of the total candidates (7.89*lacs*) in *West Bengal Higher Secondary Examination*

Skills

Languages: Python, SQL, R, C/C++
Libraries: Scikit, TensorFlow, Keras, ggplot2, Matplotlib, NumPy, Pandas, Streamlit, Dash, PySpark, NLTK, Huggingface
Software/Platforms: Microsoft Office, Power BI, Databricks
Others: Agile, Git, LaTeX

Academic Interests

1. Bayesian Statistics
2. Nonparametric Inference
3. Machine Learning
4. Survey Methodology

Standardized Test Scores

TOEFL	Reading: 30, Listening:27 Speaking: 24, Writing: 27 Total: 108
GRE	Quantitative: 170, Verbal: 157, Analytical Writing: 5 Total (Quantitative & Verbal): 327