

Mayukh Bhattacharjee

+1-343-558-5762 / mbhat053@uottawa.ca / [LinkedIn](#) / [Google Scholar](#) /

EDUCATION

University of Ottawa

Master of Computer Science (MS); GPA: 9.0/10

Ottawa, Ontario, Canada

Sept. 2022 - Present

University of Engineering & Management

Bachelor of Technology (B.Tech); GPA: 8.98/10

Kolkata, India

Jun. 2018 - Jul 2022

EXPERIENCE

Data Science Intern

BuyerFolio Inc.

Nov 2023 - Present

Montreal, Canada

- Researched Matchmaking models (Matrix Factorization Matchmaking Models) for co-ownership recommendations. Refined & integrated them, improving match accuracy by 70%
- Leveraged Wide & Deep Learning for baseline experiments, achieving 1.5X compared to traditional models.

Machine Learning Research Assistant

University of Ottawa

Sep. 2022 - Dec. 2023

Ottawa, Canada

- Integrated Anomaly Detection models (Local Outlier Factor, etc.) with Active Learning strategies to mitigate the Class Imbalance problem, achieving a Precision score of 92%
- Developed high-performance anomaly detection in large log datasets using LLMs (BERT, etc.) with word embeddings and Mix-Up algorithm, achieving 80% precision

Graduate Teaching Assistant

University of Ottawa

Fall '22, Fall '23, & Winter '24

Ottawa, Canada

- Engaged students through dynamic lab sessions and customized C++ tutorials (data structures, algorithms, design patterns) resulting in enhanced proficiency with a student rating of 4.8/5
- Supported Fundamentals of Data Science course facilitating labs, guiding individual projects, and aiding learning for 50+ students

Data Science Intern

Cognizant Technology Solutions Corp

May. 22 - Jul. 22

Kolkata, India

- Optimized ETL processes with Informatica PowerCenter, streamlining data integration and boosting operational efficiency
- Led data analytics projects and research, utilizing statistical analysis and visualizations to generate data-driven insights and inform strategic decisions

Research Intern

MITACS Globalink

May. 21 - Jul. 21

Ottawa, Canada

- Developed data-driven strategies for large time-series dataset to rectify imbalanced class issues, enhancing predictions and decision-making in Machine Learning Regression models

PROJECTS

Active Learning for Imbalanced Classes / Outlier Detection, Active Learning, Resampling Techniques

- Proposed two novel algorithms (ALOD & ALOD-RE) boosting model performance by 82% & mitigating class imbalance in real-world datasets
- Presented the research at top conference- IEEE DSAA

Sarcasm Detection in Tweets / NLTK, Scikit-Learn, Tensorflow, Keras, SMOTE

- Leveraged NLP & research to identify sarcasm detection precision & recall by 20% in social media data.
- Utilized SMOTE to improve sentiment analysis accuracy by 85% on Twitter data by addressing class imbalance

Neural Network Based Forecasting / RNN, LSTM, DNN, Forecasting, Renewable Energy

- Selected by University Dean to analyze campus solar data, leading a DNN study on LSTM networks for improved PV power generation forecasting
- Led DNN study: LSTMs improve PV power generation forecasting accuracy by minimizing RMSE to 0.02, yielding data-driven renewable energy insights using state of the art time-series analysis technique

PUBLICATIONS

Effects of Class Imbalance Problem in Convolutional Neural Network Based Image Classification / *CNN, Neural Networks, Pytorch*

- Led study on CNNs & imbalanced classes in image classification, highlighting their impact on model performance
- Optimized metrics (Precision, Recall, etc.) for imbalanced data, boosting results. Demonstrates understanding of model dynamics and potential for improvement

Variational Autoencoder (VAE) Based Imbalanced Covid-19 Detection Using Chest X-Ray Images / *SVM, KNN, Random Forest, Logistic Regression, Decision Trees*

- VAE with DL (CNNs) & resampling (SMOTE/ENN) improved imbalanced COVID-19 detection in chest X-rays
- Utilized SMOTE to improve sentiment analysis accuracy by 85% on Twitter data by addressing class imbalance

Multilabel Sentiment Prediction by addressing Imbalanced Class Problem using Oversampling / *NLTK, Tensorflow, Tweepy, SVM, Decision Trees, Natural Language Processing (NLP)*

- Developed a novel four-step approach to address class imbalance in Multi-Label Sentiment Classification, particularly for limited sentiment samples, resulting in substantial enhancement of sentiment prediction accuracy on real-world Twitter data
- Oversampled minority class, boosting the classifier sensitivity score by 91%

TECHNICAL SKILLS

Languages: Python, C/C++, Java, SQL, R (basic)

Developer Tools: Git, Docker, VS Code, PyCharm, Notebook Environments, Kafka, Docker, Spark, Hadoop

Libraries: Scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, LightGBM, NLTK, Pandas, Numpy, Matplotlib

Applications: Power BI, Tableau, L^AT_EX

Supervised Learning Algorithms: Linear Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gradient Boosting

Deep Learning Models: Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNNs), Generative Adversarial Neural Network (GANs)

Soft Skills: Communication, Analytical, Self-Motivating, Detail-Oriented, Organized, Ethical, Leadership, Personal Development