AARYA BAGDE

+1 (551) 804-2755 | ab11304@nyu.edu | LinkedIn | Github

EDUCATION

New York University – Center for Data Science **September 2023 – May 2025**
Masters of Science in Data Science ***GPA: 3.8/4.0***
Relevant Coursework: Probability and Statistics for Data Science, Linear Algebra. Big Data, Machine Learning, Natural Language Processing

Dwarkadas J. Sanghvi College of Engineering **August 2019 – June 2023**
Bachelor of Technology in Electronics and Telecommunication ***GPA: 9.55/10***
Relevant Coursework: Data Structures and Algorithms, Database Management System, Neural Networks and Fuzzy Logic, Computer Networks

SKILLS

Programming	C, C++, R, Java, Python, SQL, TensorFlow, PyTorch, Scikit-Learn, MATLAB, Numpy, Pandas
ML Framework	Regression, SVM, k-means, Xg-boost, Random Forest, k-NN, Bayesian ML
Statistics	Statistics, Probability distributions, Multivariate analysis, Time series analysis, A/B testing
Data Analytics	Tableau, MySQL, PowerBI, Matplotlib, Microsoft Excel
Big Data	Hadoop, Apache Spark, Hive, ETL

WORK EXPERIENCE

Amazon | *Data Scientist Intern* | *Seattle, USA* **June 2024 – August 2024**
- Increased operational efficiency of Fulfillment Centers (FCs) by 12% by developing a recommendation system that identified missing and misconfigured settings across over 1,000 configurations, reducing manual setup time for new FCs.
- Automated and optimized identification of configuration issues in Fulfillment Centers by implementing Multi Label Classification, achieving an F1 score of 0.88 with a Random Forest model.
- Independently designed the approach and delivered a Minimum Viable Product for configuration recommendation system

Healthysure | *Growth and Partnership Intern\ Mumbai, India* **November 2022 – January 2023**
- Conducted market analysis to identify potential clients, delivering tailored outreach campaigns that resulted in onboarding 3 new clients
- Implemented multi-channel digital marketing strategies on LinkedIn and Instagram, optimizing content with SEO and targeted ads, which led to a 10% increase in brand engagement.

TCSiON | *Data Science Intern| Remote* **August 2022 – October 2022**
- Collaborated in a team of 4 to complete my week-long Data Science training workshop and its capstone project which enhanced my skills in analyzing data and building Deep Learning models and applying NLP techniques.
- Used Twitter API to scrape 2,500 raw tweets, followed by manual annotation and labeling for sentiment polarity, ensuring high-quality data for model training and evaluation.
- Developed a sentiment analysis application leveraging transformer-based LLMs (BERT) achieving an F1 score of 0.82.

PROJECTS

Enhancing Text Coherence with Mixture of Experts Model (Manuscript)
- Achieved a 13% increase in accuracy and a 19% improvement in F1 score by developing a custom Mixture of Experts model using Mixtral, outperforming BERT's baseline performance in next sentence prediction task, with dataset sourced from Hugging Face.
- Enhanced model interpretability by implementing gating networks and expert specialization within MoE framework, which demonstrated 10% faster convergence and superior text coherence compared to traditional models.

Stock Movement Prediction using State-of-the-Art Transformer Model (Code)(Paper)
- Leveraged Amazon.inc dataset from Yahoo Financial Services of over 2,000 data points, featuring date, opening, closing, high, low, volume traded, and adjusted closing prices from January 12, 2011, to July 9, 2021.
- Developed a Transformer model for stock price prediction, which outperformed LSTM and ARIMA models, achieving a Mean Squared Error (MSE) of 0.0004091, demonstrating enhanced accuracy and prediction reliability
- Presented the paper at the "International Conference on Advances in Computer Engineering, Communication Systems and Business Development" and is awaiting publication in the corresponding journal

Heartbeats: A Real Time Emotion-based Music Recommendation System (Code)(Paper)
- Developed a music recommendation system using a public domain dataset – DREAMER that predicts a user's mood from ECG sensor data, classifying emotions into happiness, sadness, anger, and calmness.
- Trained recognition model after careful exploratory analysis and feature extraction using Decision Trees. The output of the emotion recognition model was used while making a call to Spotify API to recommend the music based on the mood.
- Presented the project at a national level competition DJ Spark and published a paper in the corresponding journal with ISBN number: 978-93-5777-300-3

CERTIFICATIONS

- Machine Learning specialization by IBM – Coursera
- Practical Data Science on AWS Cloud Specialization offered by DeepLearning.AI and AWS