Harshal Geete

Boston, MA | +1-857-313-5526 | geete.h@northeastern.edu | LinkedIn | GitHub

Data Scientist with **3+ YOE** in statistical analysis, predictive modelling, and data visualization. Adept at strategic problem-solving, cross-functional collaboration, and transforming complex data into actionable business insights

EXPERIENCE

Data Scientist, Bright Mind Enrichment, US   **May 2024 – Present**
- Orchestrated development of a **conversational AI assistant**, leveraging LLMs (GPT-4, LLAMA), RAG architecture, Hugging Face, and Langchain to provide context-aware guidance to volunteers, resulting in **25% increase** in user experience
- Collaborated with support team to identify key challenges and compile 100+ documents to build a knowledge base for volunteer information retrieval, utilized SBERT for semantic search capabilities, achieving **91% information accuracy**
- Streamlined **end-to-end deployment** by implementing **CI/CD pipelines** with Streamlit, **Docker**, Heroku, and **GitHub Actions** for seamless integration and version control, enhancing efficiency and reliability across development platforms

Data Science Teaching Assistant, Northeastern University, Boston, MA   **Jan 2023 - Apr 2024**
- Led data science lab sessions & office hours, providing guidance to **50+ students** for course "Programming for Data Science"
- Utilized Python skills to create 20+ engaging coding exercises, data analysis projects, and interactive Jupyter notebook tutorials, covering topics such as pandas, scikit-learn, and data visualization libraries, **enhancing student learning**

Data Scientist, Tata Consultancy Services, India   **Jan 2021 - Jul 2022**
- Spearheaded optimization of a credit risk assessment model by applying advanced machine learning algorithms, including Random Forest, ElasticNet that resulted in a **18% increase** in accuracy, enhancing decision making for finance team
- Partnered with product development team to build a recommendation system to personalize product suggestions by implementing hybrid collaborative & content-based filtering algorithm, leading to **10% boost** in customer engagement
- Deployed **10+ A/B tests** to targeted user groups, testing variations in navigation, transaction efficiency, and layout adjustments to enhance the mobile banking app's user experience, resulting in a **22% improvement** in overall user satisfaction
- Implemented a Risk Analysis dashboard on Tableau and performed data analysis in python, incorporating 10+ key performance indicators (KPIs) to monitor and analyze credit risk patterns across 50,000+ loan applications, **reducing default rates by 15%**

Data Science Intern, Affirmity (previously Emphasis Corp), India   **Nov 2019 - Jan 2020**
- Utilized Google Speech-to-Text API to convert customer audio feedback into text transcript and performed topic modeling using LDA to identify 11 different customer intent resulting in coherence score of **0.65,** improving marketing strategies
- Engineered an LSTM based sentiment classifying model to segregate positive and negative conversations within diverse contexts (identified using topic modeling) that resulted in weighted recall rate of **97%**

PROJECTS

Air Quality Forecasting, Time Series Analysis & Predictive models
- Performed large-scale time series processing on quantitative dataset with **2M+** data points, involving missing value imputation, and feature engineering (datetime features, lag generation), to train and compare ARIMA, XGBoost, and LSTM
- Fine-tuned hyperparameters such as learning rate, dropout layer, and batch size using randomsearchCV that resulted in LSTM neural network outperforming XGBoost, and statistical ARIMA model with an RMSE score of **0.61**

SKILLS

Programming Languages: Python, SQL, R, MATLAB, SAS, C++
Database and Tools: SQL, MySQL, MongoDB, SQL Server, Jupyter, git, Tableau, Microsoft Excel, Power BI, Google Analytics
Python Libraries & AI/ML Techniques: Sklearn, PyTorch, keras, numpy, tensorflow, Machine Learning (Regression, Classification, Clustering, Predictive modeling, Decision Trees), Deep Learning, NLP, Computer Vision, A/B Testing
Parallel Processing: ETL, Kinesis, Redshift, Hadoop, Apache Spark, PySpark, S3, Kafka, Hive, Cassandra, GCP, AWS
Statistics: A/B Testing, Causal Inference, Multivariate Testing, Non-Parametric tests, Survival Analysis, Probability

EDUCATION

Northeastern University, *Boston, MA*   **Sep 2022 - May 2024**
Master of Science in Data Science
Coursework: Supervised & Unsupervised Learning, NLP, LLMs, Time Series Analysis, Data management, Data Mining, Mathematics
Certification: Oracle Cloud Infrastructure 2024 Generative AI Certified Professional (2026)
Rajiv Gandhi Technical University, *India*   **Aug 2016 - Jun 2020**
Bachelor of Engineering in Computer Science and Engineering