

Varun Balle

Maryland | ballevarun55@gmail.com | [Portfolio](#) | +1(443)-851-1426 | [LinkedIn](#) | [GitHub](#)

EDUCATION

Masters in Professional Studies in Data Science

May 2024

University of Maryland Baltimore County, Maryland

Bachelor of Technology in Computer Science and Engineering

April 2022

CVR College of Engineering, Hyderabad

SKILLS

Programming: SQL, Python, R, JavaScript, VBA

Data Tools: AWS, Azure, MS SQL Server, Tableau, GCP, Scikit-learn, NLTK, SpaCy, CI/CD

Database Systems: MySQL, MongoDB, Relational Databases, GIS

Coursework: LLMs, Langchain, Data Modeling, BI, DLM, Agile Methodology,

Software: MS Excel, Power BI, JIRA, Confluence, Git, STATA, SAS, MS PowerApps, Collibra, EBX Tibco

PROFESSIONAL EXPERIENCE

Data Scientist, GLAFPOL LLC

June 2024 -Present

- Constructed predictive models that analyzed 40,000+ data entries on Somalia's fragility spanning over the last 6 years, to identify patterns; efforts resulted in a 30% increase in data-driven predictions.
- Enhanced data integrity by cleaning, and validating 3 databases, reducing inaccurate records by 35%, and increasing reporting accuracy by 20% through a validation system addressing 3,500 incidents.
- Produced 10 detailed reports featuring data visualizations that highlighted key research outcomes; these reports facilitated crucial discussions in policy-making conferences, leading to new research-driven strategies.

Operations Assistant, University of Maryland

August 2023 -May 2024

- Engineered data-driven strategies to monitor academic performance, pinpointing 4 major areas for focused support; optimized student outcomes by 25% over two academic terms.
- Boosted student math scores by 5% over a semester through targeted tutoring sessions; 20 students exhibited a 15% improvement in test results.

PROJECTS

Liquor Demand Forecasting a ML Approach

January 2024-May 2024

- Technologies:** ARIMA, SVR, Scikit-learn, Streamlit, Sklearn, Power BI
- Developed and trained machine learning models LR, ARIMA, and LSTM using 5GB+ data on Iowa State liquor sales data.
- Created a Streamlit-based web tool for users to engage with forecasting models, resulting in a 35% boost in user satisfaction and a 20% reduction in support queries.
- Evaluated KPIs and CDEs on liquor sales including store location, product category, and bottle volume; for an interactive forecasting model, enhancing data-driven decision-making by 30%.

Unraveling India-China Trade Relations from an NLP Perspective

August 2023-December 2023

- Technologies:** LLM, NLTK, Seq2Seq, LongFormer, Text Summarization,
- The research outputs highlight the effectiveness of NLP and ML techniques in uncovering insights from India-China trade articles.
- Implemented a data extraction process utilizing Latent Dirichlet Allocation (LDA) and Seq2Seq Model to summarize articles from credible news sources, enhancing content curation efficiency by 40% in the media industry.
- Leveraged a pre-trained LongFormer LLM to achieve up to 98% confidence in sentiment analysis for individual article chunks.

Flight Delay Analysis and Prediction

January 2023-May 2023

- Technologies:** SQL, Hadoop, MapReduce, Tableau, Apache Spark, MongoDB
- Designed an advanced flight analysis and prediction system that leverages historical flight data to predict flight delays, increasing delay-predicting accuracy by 32%; from 63% to 95% compared to existing methods.
- Architected a scalable data processing system to manage large datasets, handling over 15GB of flight delay data utilizing MongoDB and Apache Spark.

Data Analysis on Maryland Vehicular Crashes

August 2022-December 2022

- Technologies:** SQL, Google Colab, Git, Sklearn, Tableau, Airflow
- Extracted meaningful insights from Maryland Crash Data Resources, identifying the nature and reasons for over 15,000 accidents.
- Built a pipeline to collect and clean data from datasets encompassing over 7 years; 2015-2022 of data and prepared an RCA.
- Performed data analysis and visualization to communicate insights to stakeholders, including Maryland State Police, resulting in a 20% increase in understanding of crash patterns.

CERTIFICATIONS

- Google Cloud Computing Certificate**, Google Cloud

May 2024

- Tableau Data Specialist** Certification, Udemy

October 2023

- Data Governance 101**, OvalEdge Academy

August 2023