**Varun Balle**

**Maryland|**[**ballevarun55@gmail.com**](mailto:ballevarun55@gmail.com)**|**[**Portfolio**](https://varun570.github.io/%20)**|+1(443)-851-1426|**[**LinkedIn**](http://www.linkedin.com/in/varun-balle)**|**[**GitHub**](https://github.com/varun570)**|**

**EDUCATION**

***Masters in Professional Studies in Data Science May 2024***

University of Maryland Baltimore County,

***Bachelor of Technology in Computer Science and Engineering April 2022***

CVR College Of Engineering, Hyderabad

**SKILLS**

***Programming:*** SQL, Python, R, JAVA, C++

***Data Tools:*** MS SQL Server, Hadoop, Spark, Scala, Oracle Apache, Tableau, GCP, AWS, Azure, BigQuery

***Database Systems:*** MySQL, MongoDB, Relational Databases

***Coursework:*** Data Mining, Machine Learning, Project Management, Process Improvement, Agile Methodology

***Software:*** JIRA, GitHub, MS Excel, MS Excel, Informix, MS Access, MS Power BI, MS PowerApps

**PROFESSIONAL EXPERIENCE**

***Teaching Assistant, Baltimore City Public Schools System August 2023 -Present***

* Led intensive teaching programs for Baltimore students in grades 2 to 9, boosting average student scores in math by 10%.
* Tailored individualized teaching strategies for 25+ students, aligning with their unique learning styles and strengths; this approach resulted in a 40% improvement in test scores and renewed motivation for academic success.

**PROJECTS**

***Unraveling India-China Trade Relations from an NLP Perspective December 2023***

* ***Technologies: Google Colab, 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 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, increased 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.

***Digit Classification May 2023***

* ***Technologies: Google Colab, Sklearn, Matplotlib***
* Trained and evaluated Machine Learning models using the Sklearn Library to classify digits 0 to 9 with an accuracy of 97%.
* Applied algorithms, including LR, SVM, KNN, and GridSearchCV to fine-tune hyper-parameters and achieved a 17% improvement in model efficiency.
* Determined SVM with GridSearchCV as the top-performing algorithm with 97% accuracy in classifying digits based on their numerical significance.

***Data Analysis on Maryland Vehicular Crashes December 2022***

* ***Technologies: SQL, Google Colab, Git, Sklearn, Tableau***
* 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.

***Regression Model for Predicting Stock Price December 2022***

* ***Technologies: Jupyter, Sklearn, TensorFlow, LSTM***
* Implemented data cleaning and preprocessing on a dataset of 5030 stock market records encompassing 20 years to prepare it for modeling. Employed sci-kit-learn and TensorFlow libraries for advanced data processing and model development.
* Conceptualized Linear Regression and LSTM models to predict future stock prices for TSLA(Tesla) and TM (Toyota Motors) Tickers.
* Final Model predicted stock prices with an RMSE of 1.97 and LSTM provided better accuracy figures at 92%.

**CERTIFICATIONS**

* ***APICS CPIM***  ***Ongoing 2024***
* ***Google Cloud Computing Certificate,*** Google Cloud ***May 2024***
* ***Tableau Data Specialist*** Certification, Udemy ***October 2023***
* ***Data Governance 101,*** Oval Edge Academy ***August 2022***