Tharun Boddu

+1-(517) 528-7611 | boddutha@msu.edu| LinkedIn | Tableau | GitHub

EDUCATION

Michigan State University - ***Master of Science, Business Data Science and Analytics*** **2024**

BML Munjal University - ***Bachelor of Technology, Computer Science*** **2020**

SKILLS

- **Languages & Libraries**: SQL, Python **[**NumPy, Pandas, Seaborn, Matplotlib**]**, R, JavaScript.
- **Machine Learning:** Predictive Modeling, SVM, KMeans, Decision Trees, Naive Bayes, Random Forest, XGBoost, PCA.
- **Technical:** Data Analysis**,** ETL, Linux, Amazon Redshift, S3 Bucket, OOP, Data structures & Algorithms, HTML5, CSS3.
- **Tools/IDE**: Visual Studio Code, Jupyter Notebook, Anaconda, Power-BI, GIT, D-Beaver, R Studio, MS-Excel, Power point.

PROFESSIONAL EXPERIENCE

Research Assistant, Department of Civil Engineering **Present**

- Currently working on drone video stabilization using the computer vision library **OpenCV** and template matching concept

Qside Institute, Research Fellow **MA**

June 2024-Nov 2024

- Extensively utilized Python scripts, Pandas, and Matplotlib to retrieve and clean funder data from the IRS API, creating a data lake and generating insightful visualizations for analysis.
- Developed and refined a Tableau dashboard that visualizes pre-processed data to display the distribution and various indicators of opportunity youth between 16-24 years (Disconnected, Foster, Youth Arrests) in the U.S.
- Enhanced the dashboard's user interface to ensure it is intuitive for non-data experts, incorporating features such as filtering by location and age, and toggling between different indicator variables.

Alshaya Group, Software Engineer **Bengaluru, India**

Oct 2021-Nov 2023

- Proficient in Python fundamental concepts, crafting scripts for automation and data processing. Utilized Python for exploratory data analysis (EDA) to derive actionable insights.
- Implemented all SQL Commands including DDL, DML, TCL, DQL, DCL, and Aggregate functions, JOINS, sub-queries, Rank, Dense rank, Row number, and constraints for analyzing complex datasets. Extensively used Oracle Redshift DB.
- Scheduled ETL jobs for smooth transformation of raw data into structured, usable formats. Extensive use of Redshift for data access. Loading Raw Data from S3 bucket, Data Cleansing and creating KPIS for Business users.
- Demonstrated understanding of data warehousing concepts: star schema, Snowflake schema, facts, dimensions, cardinality, and granularity for efficient data modeling and analysis.
- Led the release of software into production, including documentation, code reviews, coordination with IT release teams, raising Snow and Jira tickets, and execution on production servers.

PROJECTS

1.Predictive Analytics in KPI Forecasting (Academic-Capstone)

- Utilized historical data and applied statistical and machine learning methods, including ensemble and neural networks, to improve KPI prediction accuracy, assess forecast reliability, and communicate forecast uncertainty with confidence intervals.

2.Market Data Segmentation Through Credit Card Analysis in R – Clustering (Unsupervised ML)

- Segmented credit card users by spending habits, demographics, and real estate trends for targeted marketing, using data wrangling, outlier detection, and visualizations.
- Applied **PCA** for dimensionality reduction and clustered users with KMeans, Fuzzy, and CLARA (achieving a silhouette score of 0.8) to identify distinct groups.

3.Forecasting Consumer Credit Card Balances for Large Banks: A Time Series Analysis

- The research explores the quarterly trends of consumer credit card balances for large banks and aims to forecast future balances using advanced time series methods. Insights from this study can help banks in strategic decision-making, risk assessment, and resource allocation. Applied simple to complex time series forecasting techniques and analysis of the model diagnostics outputs.