Sathwik Bollepalli

860-781-1558 | sathwik238@gmail.com| LinkedIn | sathwik-bollepalli.github.io | GitHub

TECHNICAL SKILLS

- **Languages & Tools:** Python (NumPy, pandas, matplotlib, scikit-learn, seaborn), R, SQL, Oracle, AWS, Azure, Git
- **Visualization & Reporting:** Tableau, Power BI (DAX), Excel (Power Query, Pivot Tables, Macros, VBA, VLOOKUP)
- **Machine Learning & AI:** Generalized Linear Model (GLMs), Decision Tree, Random Forest, XGBoost, Neural Network
- **Statistical Analysis:** Hypothesis Testing, A/B Testing, Statistical Modeling

EDUCATION

University of Connecticut – Storrs, CT August 2023 - December 2024
Master of Science (M.S) in Data Science

Lovely Professional University, Punjab, India August 2019 - May 2023
Bachelor of Technology (B. Tech) in Computer Science

PROFESSIONAL EXPERIENCE

Plymouth Rock Assurance - Boston, USA May 2024 – August 2024
Data Scientist Intern

- Cleaned ISO neighborhood property loss data, creating 800+ new features and ensuring data consistency and quality through a rigorous checking process.
- Optimized code and data structures, reducing processing time by 70% and streamlining the feature creation pipeline for future data ingestion.
- Conducted extensive exploratory data analysis (EDA) to review feature distributions and identify risk patterns.
- Built robust home insurance risk prediction models using Tweedie XGBoost with Bayesian search for hyper-parameter optimization, among other machine learning techniques, achieving high performance evaluated by Gini coefficient, lift chart, and Lorenz curves.
- Demonstrated the additional value of newly created features by improving existing models and data sources.

MAQ Software Pvt Ltd - Hyderabad, India
Machine Learning Engineer *May 2023 – August 2023*

- SQL Query Optimization: Played a pivotal role in developing and implementing complex SQL queries, boosting data retrieval and analysis efficiency by 50%.
- Machine Learning Lifecycle Management: Streamlined machine learning lifecycle using Azure ML, facilitating a no-code approach for model development and deployment, remarkably minimizing manual intervention.
- Code Optimization: Achieved a significant reduction in code execution time from 2 hours to 27 minutes, a 77% improvement, through efficient optimization techniques.
- Complex Problem Solving: Utilized advanced data structures and algorithms, including trees and recursion, to effectively tackle complex business challenges, thus enhancing the usefulness and precision of solutions.

Data Engineer (Internship) *May 2022 – April 2023*

- Data Visualization and Reporting: Developed dynamic, interactive Power BI reports, greatly increasing user engagement and enriching the data visualization, leading to more informed decision-making.
- Data Management: Implemented Azure Data Factory pipelines and SQL stored procedures to simplify the extraction and organization of data from diverse sources into Azure Data Lake Storage (ADLS), automating the data extraction process, and leading to optimized access and increased data reliability.
- Data Transformation with Azure Databricks: Leveraged Azure Databricks for performing complex data transformation using SQL and Python, improving business intelligence and analytics.

ACADEMIC PROJECTS

- **Bank Customer Churn Prediction | R | ML** – Utilized logistic regression, decision trees, and random forests for predictive analysis of customer churn in the banking sector, incorporating various performance metrics to identify high-risk customers, contributing to the development of effective retention strategies, showcasing skills in data analysis and problem-solving.
- **Email Spam Detection | Python | ML** – Developed an advanced email spam detection system using Python and ML techniques, achieving 95% accuracy in classifying emails into 'spam' or 'ham'. Boosted data management and classification efficiency, enhancing workflow processes and minimizing false positives.
- **Energy Demand Forecasting | Python | Deep Learning**– Developed an RNN-based energy demand forecasting model using weather data, leveraging Conv1D, LSTM layers, dropout, and activation functions in Keras. Achieved high accuracy with feature engineering, walk-forward validation, early stopping, and validated using mean absolute error.

ACHIEVEMENTS AND CERTIFICATIONS

- Modern Big Data Analysis with SQL Specialization
- Google Data Analytics
- Hacker Rank certified Python Programmer and Advanced SQL Programmer