

Sathwik Bollepalli

860-781-1558 | sathwik238@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)

SUMMARY

Data Scientist skilled in predictive modeling, machine learning, and data visualization, with a proven track record of improving model accuracy by 15% and reducing data processing time by 70%. Experienced in Python, SQL, and Tableau to deliver actionable business insights and drive strategic decisions.

PROFESSIONAL EXPERIENCE

Plymouth Rock Assurance - Boston, USA

Data Scientist

May 2024 – December 2024

- Achieved a **70% reduction in data processing time** by restructuring pipelines, using data structures and automating the generation of 800+ features, enabling the scalability of predictive models and reducing deployment delays.
- Developed Tweedie XGBoost models with Bayesian hyperparameter tuning, achieving a 15% improvement in Gini coefficient and more accurate home insurance premium predictions.
- Enhanced risk assessment by applying localized adjustment factors, increasing prediction accuracy for high-risk areas like flood zones and high-crime neighborhoods.
- Improved precision for predicting large-loss fire events by **30%** in highly imbalanced datasets, leveraging advanced resampling techniques and XGBoost-based logistic regression.
- Influenced key strategic decisions by delivering insights to senior management, resulting in the integration of two high-value third-party data sources, improving model performance and decision-making processes.

MAQ Software Pvt Ltd - Hyderabad, India

Machine Learning Engineer

May 2022 – August 2023

- Achieved a 50% improvement in data retrieval efficiency and **reduced processing time by 77%** (from 2 hours to 27 minutes), directly enhancing operational workflows and enabling faster decision-making across the organization.
- Improved model deployment efficiency through the implementation of a no-code approach in Azure ML, reducing manual errors and accelerating the time-to-production for machine learning models.
- Automated data extraction and organization processes, resulting in a **40% reduction in data preparation time** and increasing data accessibility for analytics teams, improving workflow reliability.
- Increased **stakeholder engagement by 35%** and improved decision-making by providing clear, actionable insights through visually compelling dashboards that synthesized complex data into easily digestible formats.
- Enhanced analytics and reporting accuracy by 20%, enabling the delivery of more precise insights to stakeholders and reducing data inconsistencies across reporting platforms.

EDUCATION

University of Connecticut – Storrs, CT

August 2023 - December 2024

Master of Science (M.S) in Data Science, **GPA : 3.7**

Relevant Courses – Predictive Modeling, Machine Learning, Statistical Analysis, Causal Inference, Data Visualization, Deep Learning.

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.

TECHNICAL SKILLS

Python (Pandas, Keras, NumPy, PySpark, Scikit-learn, Matplotlib, Seaborn), SQL (MySQL, PostgreSQL, Snowflake), R, Data Analysis, Data Wrangling, ETL Pipelines, Data Visualization (Tableau, Power BI, Matplotlib, ggplot2), Statistical Modeling, A/B Testing, Hypothesis Testing, Bayesian Analysis, Transformers (BERT, GPT), Tableau, Microsoft Office Suite (Excel, PowerPoint, Word), Machine Learning (XGBoost, Random Forest, SVM, K-Means, KNN), Predictive Modeling, Deep Learning (CNN, RNN, LSTM), Negative Binomial Regression, Linear & Logistic Regression, AWS (S3, EC2, SageMaker), Azure (Data Factory, Databricks).

ACHIEVEMENTS AND CERTIFICATIONS

- Modern Big Data Analysis with SQL Specialization
- Google Data Analytics
- Hacker Rank certified Python Programmer and Advanced SQL Programmer
- JP Morgan & Chase CADP Certification