SOWMYA Yalavarthi
Data Scientist

Jersey City, NJ 07307 | 201-885-8579 | sowmyayalavarthi98@gmail.com | LinkedIn | GitHub

Professional Summary

Data Scientist with **3.6+ years of experience** conducting **exploratory data analysis (EDA), building interactive dashboards in Tableau,** and delivering actionable insights for product and business teams. Skilled in **Python (Pandas, NumPy), SQL, and data visualization tools.** Strong foundation in **statistics, machine learning, and business communication**. Passionate about continuous learning and active contributor in agile environments.

Technical Skills

- **Programming:** Python (Pandas, Numpy), SQL
- **Machine Learning & Deep Learning:** Supervised, Un-Supervised, Semi-Supervised, Random Forest, Boosting (XGBoost, LightGBM), Transformers, NLP, Recommendation Systems, Text Summarization
- **Mathematical Modeling & Statistical Analysis**: Regression, Classification, Probability, Optimization, Hypothesis Testing
- **Big Data & Cloud Technologies:** Spark, Hadoop, Snowflake, AWS (S3), GCP (BigQuery).
- **ML & Data Science Libraries:** Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow
- **Data Analysis & Visualization:** Exploratory Data Analysis (EDA), Tableau, Matplotlib, Seaborn, Power BI, Plotly
- **Methodologies**: Agile/Scrum, Ad-hoc Analysis, Insight Generation

Professional Experience

Capgemini, Hyderabad, India
Associate Consultant (Data Scientist)   **12/2019 – 05/2023**

- Built predictive models (**RF, Logistic Regression, Clustering)** to forecast KPIs, improving accuracy by 20%.
- Deployed real-time pipelines in **PySpark & AWS (Lambda, EMR) for 5M+ records/day with 99.9% success.**
- Developed **NLP pipelines using BERT, SpaCy, and rule-based logic for BKMs and sentiment analysis**.
- Automated **ML workflows using Python & SQL**, reducing manual effort and boosting efficiency.
- Conducted **EDA and delivered Tableau dashboards** to support product and business decisions.
- Handled **ad-hoc analysis and collaborated with cross-functional teams** to drive insights.
- Participated in **agile ceremonies**: daily stand-ups, sprint planning, and reviews.
- Monitored model performance using **MLflow** and dashboards to ensure continuous improvement.

Academic Projects

Stock Prediction using Machine Learning   **05/2024**
- Developed **Random Forest, SVM, and Logistic Regression** models to classify stock price movements.
- Improved model accuracy by **15%** using feature engineering and hyperparameter tuning.
- Evaluated models using **AUC and F1-score** and presented actionable insights to guide trading decisions.

Insurance Complaints Analysis with NLP   **12/2024**
- **Cleaned and analyzed** complaint data from the Texas Department of Insurance to uncover trends and drive improvements in customer service.
- Built a **BERT**-based pipeline to classify sentiments and summarize insurance complaints.
- Created visualizations in **Tableau and Matplotlib** to display sentiment trends and summarize complaint volumes.

Chatbot using Gemini API   **09/2024**
- Developed a basic chatbot leveraging **Gemini API** to answer FAQs related to primary **healthcare** during a hands-on university workshop.
- Explored **text understanding and automation** via natural language APIs for healthcare use cases, simulating intelligent summarization and response generation based on user intent.

Education

Pace University, New York, NY   **05/2025**
Master of Science in Data Science

Koneru Lakshmaiah University, Vijayawada, India   **05/2019**
Bachelor of Science in Electronics and Communication Engineering

Awards / Certifications

- **Awards**: Received Customer delight award for excellence in project work delivery & customer satisfaction at Capgemini.
- **Certifications: AWS Cloud Practitioner**