SHRUTI MALLAVOLU

 Portfolio Website ‖ linkedin.com/in/shruti-mallavolu/ ‖ ✉ shrutiramesh67@gmail.com ‖ 📱 +1 (413) 210-3675

Summary

Data Analyst with 2+ years of experience in Software Development and Masters in Computer Science (focus on Data Science). Have experience developing and deploying machine learning models, analyzing large datasets, and delivering actionable insights. Proficient in Python, R, and SQL, with a track record of reducing operational costs by 25% and improving prediction accuracy by 20%.

Education

University of Massachusetts Amherst   **Jan 2022 - Dec 2023**

Master of Science in Computer Science (Concentration in Data Science), GPA: 3.9/4.0   Amherst, MA
- **Coursework:** Machine Learning, Statistics, Information Retrieval, Reinforcement Learning, DS Algorithms, Distributed Systems

SRM Institute of Science and Technology   **Jun 2015 - May 2019**

Bachelor of Technology in Computer Science, GPA: 8.67/10.0   Chennai, India
- **Coursework:** Data Structures and Algorithms, Database Systems, Natural Language Processing, Data Mining, AI

Technical Skills

Programming languages: Python, R, Java, C++ ; **DBMS** : MySQL, PostgreSQL, No-SQL ; **OS** : Linux, Windows
Tools and Frameworks: Tableau, Docker, AWS, Spark, Airflow, Kubernetes, FlaskAPI, VS Code, Eclipse, Google Colab, Jupyter
ML/AI: Scikit-learn, XGBoost, TensorFlow, Numpy, Pandas, Matplotlib, Seaborn, Pytorch, AutoML, Beautiful Soup, RNN, LSTM

Professional Experience

Servicenow   **Aug 2024 - Current**

Data Quality Analyst | *Python, SQL, Excel, Snowflake, PowerBI*   San Jose, CA
- Engineered and monitored Data Quality (DQ) reconciliation processes for consistency, accuracy, completeness, and uniqueness across systems (CPQ, CRM, SURF) with Master Data Management (MDM).
- Implemented Python-based frameworks for automated DQ checks, enhancing data validation efficiency by 30%.
- Developed PowerBI dashboards to track and visualize DQ metrics, streamlining decision-making across four business verticals.
- Built exception reporting with automated email alerts, improving response time to DQ anomalies and mitigating integrity risks.

The Coding School (Data Science Research Program)   **July 2024 - Aug 2024**

Teaching Assistant | *Data Science, Statistics*   Remote, US
- Mentored students in DS, ML (PCA, Statistics, K-NN) and guided projects analyzing housing, ecological footprint datasets.

Microsoft (ALICE team)   **Feb 2023 - May 2023**

Graduate Student Researcher | *Python, Scikit-learn, JupyterHub, EconML*   Amherst, MA
- Automated the selection of first-stage models and hyperparameters for CATE estimators used in Causal Inference enhancing user input flexibility for the EconML package catering to use cases like Customer Segmentation, Recommendation A/B testing.
- Enhanced R^2 score by 6% through advanced feature engineering which includes feature selection through importance scores and strategic model selection, leveraging regression and classification algorithms on semi-synthetic and synthetic datasets.
- Achieved 62% Runtime and 90% Tao Risk improvement by opting Bayesian over Grid Search for hyperparameter optimization.

Salesforce   **Aug 2021 - Nov 2021**

Support Analyst | *Salesforce Marketing Cloud, SQL, A/B Testing*   Hyderabad, India
- Analyzed and debugged over 7% of incidental data generated from all customer reports within my team to promptly resolve cases.
- Provided support for A/B Testing in Email Studios and integrations between Salesforce CRM and Marketing Cloud.
- Extracted relevant insights from Email Subscriber data and leveraged SQL queries to identify segmented audience, customer trends.

Bank of America   **Jun 2019 - Jul 2021**

Software Engineer, Data | *Salesforce CRM, Sales Cloud, Apex, OOPs, REST API, Bitbucket, Agile, JIRA*   Chennai, India
- Spearheaded ARM portal enhancement, optimizing logic using Apex Triggers, Batch jobs, SQL queries, and test classes achieving atleast 80% coverage enabling Financial Advisors' role requests, achieving 50% efficiency boost.
- Implemented Twilio interface for streamlined Financial Advisor-Customer communication, integrating via Mulesoft & REST APIs.
- Refactored application components, reducing codebase by 1000+ lines and improving system performance by 30%.
- Executed daily jobs on Mulesoft to extract data from multiple external systems to synchronize the Accounts data in Salesforce.
- Configured Hadoop cluster to query Salesforce historical data when limits exceeded and stored results in an on-premise SQL database for enhanced data management.

Projects

Review of Machine Learning algorithms on various datasets | *Numpy, Pandas, Matplotlib*
- Achieved 90%+ accuracy using Neural Networks, Random Forests, Decision Trees, and K-NN on datasets like Parkinson's, Digits, Titanic, Loan, and Telecust. Evaluated performance using Accuracy, Precision, Recall, and F1 Score.

Text Data Analysis and Ranking Framework Development | *Information Retrieval, pyltr, LambdaMART, Google Colab*
- Applied k-means clustering on the BBC dataset and LambdaMART for MS Marco, achieving a 0.33 test score.

Predictive Analysis on Stroke Dataset | *R, Statistical Analysis, Kaggle, Linear Regression*
- Developed lr model to predict BMI from glucose level, age (R^2 of 0.28). Conducted chi-square tests on stroke vs gender, residence.

Data Visualization and Exploratory Data Analysis for Social Impact | *Python, Streamlit, Seaborn, Tableau, Excel*
- Analyzed college donation dataset, leveraging Tableau and Excel, uncovering donation patterns and academic donor trends.

Build a Toy Store using Microservices | *Docker, Kubernetes, AWS*
- Architected and deployed a microservices-based system with Docker, Kubernetes, and AWS, enabling concurrent request handling, multi-threading, caching, and fault tolerance across distributed services.