

SHRUTI MALLAVOLU

Portfolio Website || [in linkedin.com/in/shruti-mallavolu/](https://www.linkedin.com/in/shruti-mallavolu/) || [✉ shrutiramesh67@gmail.com](mailto:shrutiramesh67@gmail.com) || [📞 +1 \(413\) 210-3675](tel:+14132103675)

Summary

Data Analyst with 3 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 improving system efficiency by 25% and increasing prediction accuracy by 20%.

Education

University of Massachusetts Amherst

Jan 2022 - Dec 2023

Master of Science in Computer Science (Concentration in Data Science)

- **Coursework:** Statistics, DS Algorithms, Information Retrieval, Machine Learning, Reinforcement Learning, Distributed Systems

SRM Institute of Science and Technology

Jun 2015 - May 2019

Bachelor of Technology in Computer Science

- **Coursework:** Natural Language Processing, Data Mining, AI, Data Structures and Algorithms, Database Systems

Technical Skills

Programming languages: Python, R, Java, C++ ; **DBMS :** MySQL, PostgreSQL, No-SQL ; **OS :** Linux, Windows, Mac

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 - Feb 2025

Data Quality Analyst | *Python, Pandas, SQL, Snowflake, PySpark, PowerBI, DAX, Excel*

San Jose, CA

- Designed, implemented, and monitored Data Quality (DQ) reconciliation processes to ensure consistency, accuracy, completeness, and uniqueness across Sales, CRM (Customer Relationship Management), CPQ (Configure, Price, Quote), and SURF systems, improving data integrity and compliance with Master Data Management (MDM) standards.
- Developed and optimized ETL pipelines using PySpark and SQL, improving data ingestion, transformation, and processing efficiency by 40%, ensuring seamless integration of structured and semi-structured data from multiple sources.
- Built an automated Python-based framework for data quality checks, reducing manual effort and increasing validation efficiency by 30%, identifying and resolving anomalies in high-volume datasets.
- Created PowerBI dashboards and visual reports, leveraging DAX functions to track DQ metrics, anomaly trends, and data completeness levels, leading to an improvement in decision-making across four business verticals.
- Implemented exception reporting with real-time automated email alerts, reducing response time to data quality anomalies by 35% and mitigating risks related to incorrect sales forecasts and revenue leakage.
- Conducted root cause analysis (RCA) on data inconsistencies, identifying systemic issues in upstream data sources and providing actionable insights to business teams, leading to a 25% improvement in data accuracy.
- Collaborated with cross-functional teams, including Data Engineering, Sales Operations, and Business Intelligence, to define data quality KPIs, build scalable monitoring solutions, and optimize data governance strategies.
- Automated ad hoc data extraction and transformation tasks using Python and SQL, reducing manual intervention by 50% and improving overall productivity.
- Implemented advanced statistical methods and anomaly detection techniques to identify patterns in missing or inconsistent data, ensuring higher reliability of business-critical reports.
- Optimized Snowflake query performance by implementing clustering techniques, partitioning strategies, and query tuning, reducing reporting latency by 30% for high-volume datasets.
- Led data reconciliation projects, comparing ERP, CRM, and Sales pipeline data to detect discrepancies, ensuring accurate reporting for leadership and financial analysis.

Microsoft (ALICE team)

Feb 2023 - May 2023

Graduate Student Researcher | *Python, Numpy, Scikit-learn, JupyterHub, EconML*

Amherst, MA

- Developed an automated pipeline for model selection and hyperparameter tuning of CATE (Conditional Average Treatment Effect) estimators, reducing manual effort and improving reproducibility in Causal Inference workflows.
- 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.
- Optimized hyperparameter tuning by implementing Bayesian Optimization over Grid Search, reducing model training runtime by 62% and improving Tao Risk by 90%.
- Benchmarked multiple Causal Inference models on semi-synthetic and real-world datasets, evaluating stability and generalization across different data distributions.
- Integrated Explainable AI techniques such as SHAP values and Counterfactual Explanations to improve interpretability of treatment effect predictions, ensuring transparency for business stakeholders.
- Refactored and contributed improvements to the EconML package, enhancing flexibility for user-defined model inputs and expanding usability for real-world business applications.

Salesforce

Aug 2021 - Nov 2021

Support Analyst | *SQL, A/B Testing, Tableau, Salesforce Marketing Cloud*

Hyderabad, India

- Analyzed and debugged over 7% of incidental data generated from all customer reports within my team to promptly resolve cases.
- Facilitated seamless execution of A/B Tests on two subscriber lists while monitoring highest unique open rate and click-through rate to optimize campaign performance.
- Assisted in data integration efforts between Salesforce CRM and Marketing Cloud, ensuring seamless data flow across platforms.
- Extracted insights from Email Subscriber data, leveraging SQL and Tableau to identify segmented audience trends.

Bank of America

Jun 2019 - Jul 2021

Software Engineer, Data | *SQL, Python, Salesforce CRM, Apex, REST APIs, Hadoop, JIRA, OOPs, Bitbucket, Agile*

Chennai, India

- Designed, optimized, and automated SQL queries and Apex logic to streamline role request processing for financial advisors, reducing manual intervention by 60% and improving workflow efficiency by 50%.
- Developed automated data extraction workflows using MuleSoft, integrating customer and account data across multiple external systems, ensuring seamless data synchronization and minimizing data discrepancies.
- Integrated Twilio API for real-time customer communication, enabling financial advisors to automate appointment scheduling and client interactions, leading to a 25% increase in customer engagement.
- Implemented a Hadoop-based data pipeline to store and query historical Salesforce data, ensuring that data retrievals do not hit Salesforce platform limits, improving query performance and operational efficiency by 40%, and enabling smoother access to historical data for analytical purposes.
- Refactored and modularized application components, eliminating 1,000+ redundant lines of code, reducing system load times by 30%, and improving overall application maintainability and performance.
- Built SQL-driven reports and automated dashboards, leveraging Python and PowerBI to enable real-time monitoring of data synchronization, system performance, and user activity trends, improving data accessibility for business stakeholders.
- Developed and optimized REST APIs for seamless data exchange between Salesforce CRM and external banking applications, enhancing data interoperability and enabling real-time access to customer profiles and SMS logs.
- Engaged in regular Scrum and PI meetings throughout the Agile SDLC to discuss relevant feedback and project goals. Participated in code reviews and used Bitbucket to keep track of production ready changes.

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 nDCG score.

Fraud Detection in Financial Transactions | *Scikit-learn, SMOTE, xgboost*

- Handled class imbalance with SMOTE, engineered features and built XGBoost and Gradient Boost models to detect fraudulent transactions. Evaluated model performance using AUC-ROC curves and analyzed key fraud indicators.

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.