Vengaiah Madineni

Data Scientist

Arlington, TX | 706-908-7565 | vengaiahchowdarymadineni@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

Data Scientist with 4+ years of experience developing machine learning and statistical solutions across financial services, customer analytics, behavioral analytics, and business intelligence. Strong background in Python, SQL, Scikit-learn, XGBoost, Random Forest, and Gradient Boosting, with hands-on expertise in exploratory data analysis, feature engineering, predictive analytics, and model validation. Builds analytical datasets and predictive solutions for customer behavior, risk, fraud, segmentation, and retention, and partners with stakeholders to turn complex data into actionable business insight and production-ready models.

TECHNICAL SKILLS

Programming & Data: Python, SQL, Pandas, NumPy, Data Cleaning, Data Wrangling, Data Preprocessing, Data Transformation
Machine Learning: Scikit-learn, XGBoost, Random Forest, Gradient Boosting, Supervised & Unsupervised Learning, Classification, Clustering, Predictive Modeling
Statistical Analysis: Statistical Modeling, Hypothesis Testing, Correlation Analysis, Trend Analysis, Cohort Analysis, Segmentation, Anomaly Detection, Behavioral Analysis
Feature Engineering: Feature Selection, Feature Transformation, Encoding, Scaling, Dimensionality Reduction, Missing Value Treatment, Outlier Analysis, Data Quality Analysis
Model Development & Evaluation: Model Training, Validation, Cross-Validation, Hyperparameter Tuning, Error Analysis, Model Comparison, Model Calibration; Accuracy, Precision, Recall, F1-Score, ROC-AUC, RMSE, MAE
Analytics: Customer, Transaction, Risk, Fraud, Retention, Portfolio, and Operational Analytics; Customer Segmentation
Data Science Workflows: EDA, Analytical Dataset Development, Reusable Feature Pipelines, Statistical Experimentation, Model Monitoring, Model Retraining, Production Validation
Visualization: Matplotlib, Seaborn, Analytical Reporting, Data Visualization
Development & Delivery: Model Deployment, Data Validation, Testing, Code Reviews, Production Support, Technical Documentation, Agile, Scrum
Collaboration: Business Requirements, Stakeholder Communication, Business Analysis, Cross-Functional Collaboration, Analytical Storytelling

PROFESSIONAL EXPERIENCE

Financial Data Scientist **Feb 2025 - Present**
Synchrony USA

- Develop and deploy machine learning and statistical models in Python and Scikit-learn across customer, account, transaction, and operational datasets of 300K+ records.
- Perform EDA, feature engineering, feature selection, and statistical analysis, improving analytical data quality by 18% through structured preprocessing, missing-value treatment, and validation.
- Build predictive models for customer behavior, risk, credit, fraud, and retention, supporting 8+ business use cases across financial and customer analytics.
- Apply XGBoost, classification, clustering, and supervised/unsupervised learning across 12+ modeling experiments, using behavioral and transactional features to identify predictive patterns.
- Evaluate models using precision, recall, F1-score, ROC-AUC, RMSE, and MAE, improving selected model performance by approximately 15% through hyperparameter tuning, cross-validation, and error analysis.
- Use SQL to create reusable analytical datasets and feature pipelines, reducing recurring data preparation effort by 20% across 6+ machine learning workflows.
- Collaborate with risk, product, engineering, analytics, and business stakeholders across 5+ cross-functional initiatives, presenting model findings and supporting validation, deployment, monitoring, and production activities.

AI/ML Data Scientist **Jan 2021 - Jul 2023**
Cybage Software India

- Developed machine learning and statistical models for business, customer, and operational analytics on datasets of 150K+ records across multiple analytical initiatives.
- Performed data cleaning, preprocessing, EDA, and feature engineering in Python and SQL, reducing recurring data preparation time by 15% through reusable workflows.
- Built supervised and unsupervised models using Scikit-learn, XGBoost, Random Forest, and Gradient Boosting, evaluating 10+ model configurations across business use cases.
- Applied statistical analysis, feature selection, transformation, encoding, scaling, and dimensionality reduction across 20+ candidate variables to improve model inputs.
- Conducted hyperparameter tuning, cross-validation, model evaluation, and error analysis, improving selected model performance by approximately 12% across iterative experiments.
- Used SQL, Python, Matplotlib, and Seaborn to prepare analytical datasets, build visualizations, perform customer segmentation, and communicate findings across 6+ recurring analytical workflows.

EDUCATION

Master of Science in Applied Statistics and Data Science **Dec 2024**
University of Texas at Arlington, Arlington, TX

Bachelor of Technology in Electronics and Communication Engineering **May 2023**
Dhanekula Institute of Engineering and Technology, Vijayawada, AP