Harsh Malashetti

9019856807 harshmalashetti@gmail.com harsh-malashetti github.com/132905 harsh-malashetti.vercel.app

Technical Skills

Programming: Python, SQL, Git, Docker, Java
ML & Analytics: Supervised and Unsupervised Learning (Regression, Classification, SVM, XGBoost), Hyperparameter Tuning (GridSearchCV), Model Evaluation (Cross-Validation, AUC, F1-score), Predictive Analytics, A/B Testing, Statistical Analysis
Libraries: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, LightGBM, Matplotlib, Seaborn
Tools: Jupyter Notebook, VS Code, Power BI, Tableau, Excel, Hugging Face
Cloud & MLOps: GCP (cloud exposure), Docker (containerization), ML-pipeline deployment

Experience

EffiGo **Jan 2024 – Dec 2024**
Product Engineer *Bengaluru, KA*

- Revamped backend systems by migrating legacy architecture to Spring Boot and PostgreSQL, improving query efficiency by 45% and enhancing backend scalability for analytics-driven applications.
- Automated CI/CD and testing pipelines using GitHub Actions, Docker, and Selenium, cutting deployment time by 55% and enabling faster release cycles for data products.
- Developed demand forecasting models using Python (scikit-learn) and Excel, achieving 85% accuracy on retail customer churn data and reducing false positives by 20% demonstrating strong predictive analytics and structured modeling.
- Designed SQL-integrated data pipelines with Excel-based reporting dashboards, streamlining ingestion, transformation, and visualization of retail demand signals and improving stakeholder decision-making.
- Authored detailed technical documentation, including Jupyter notebooks, product release notes, and troubleshooting guides—demonstrating structured problem solving, stakeholder communication.

Projects

Binary Classification ML Pipeline | *Python, Scikit-learn, SVC, Neural Network, Pandas, NumPy*

- Built an end-to-end binary classification workflow on a dataset of 10,000+ records, boosting AUC by 15% through hyperparameter-tuned SVC and Neural Network models.
- Executed full-scale exploratory data analysis and robust data cleaning—including outlier detection, preprocessing of 4 string and 6 mixed-type columns—using Pandas and NumPy.
- Engineered 20+ new features and normalized mixed-data columns (e.g. currency, percentages), increasing model recall by 12%.
- Compared performance of 6+ supervised models using GridSearchCV with 10-fold stratified cross-validation, selecting top 2 and deploying them for class-probability prediction on unseen test data.
- Authored a detailed comparative writeup explaining why Neural Network and SVC outperformed others, improving reproducibility and stakeholder understanding by documenting decision rationale.

Airbnb Listing Price Prediction | *Python, Pandas, Scikit-learn, XGBoost, LightGBM, Jupyter*

- Developed a regression model on a dataset of 40,000 Airbnb listings, achieving a 20% improvement in RMSE, by creating a stacked ensemble of XGBoost and LightGBM.
- Executed thorough EDA and preprocessing—including missing-value imputation, log transformation of skewed variables, and encoding of categorical features—using Pandas and NumPy.
- Engineered 30+ features, capturing location, host traits, and seasonal effects, improving model R^2 by 25%.
- Tuned model hyperparameters using GridSearchCV and evaluated through 5-fold cross-validation; final ensemble delivered reliable nightly price predictions.
- Compiled a detailed report with data visualizations (Matplotlib, Seaborn), model diagnostics, and recommendations for pricing strategy—enhancing stakeholder interpretability and adoption.

Education

M S Ramaiah Institute Of Technology **Sep 2020 – May 2024**
Bachelor of Engineering in Electronics and Communication *CGPA: 9.02/10*

Certifications

Supervised Machine Learning: Regression and Classification – *DeepLearning.AI*
Explore Generative AI with the Vertex AI Gemini API – *Google*
Prompt Design in Vertex AI – *Google*