JANGA SASHI HARI KRISHNA

+91-7995272311 sashiharikrishnajanga@gmail.com GitHub | LinkedIn

PROFESSIONAL SUMMARY

Data Scientist experienced in building, evaluating, and deploying machine learning models focused on risk and fraud detection use cases. During my internship, I worked with real-world datasets to build fraud-style classifiers, optimize pipelines, and deploy models using joblib and Pickle. I've led two end-to-end projects: a heart disease prediction system (Random Forest, SHAP, AUC 0.89) and a bulldozer price forecasting model (XGBoost, RMSE < 0.25). Skilled in SQL for data extraction, Python for ML, and Power BI for storytelling. Strong foundation in EDA, statistical analysis, root cause detection, experimental design, A/B testing logic, and stakeholder-driven reporting. Focused on delivering scalable, interpretable solutions for fraud mitigation, operational efficiency, and business metrics improvement.

EDUCATION

B.Tech – Civil Engineering
Rajeev Gandhi Memorial College of Engineering & Technology | 2019–2023 | CGPA: 7

PROFESSIONAL EXPERIENCE

MACHINE LEARNING INTERN

Unified Mentor | Remote | DEC 2024 – JAN 2025

CERTIFICATE

- Designed fraud-style classification models to detect anomalies using Logistic Regression and Random Forest.
- Performed exploratory data analysis (EDA) on real-world datasets to uncover risk patterns and data drift.
- Extracted large volumes of data using SQL and optimized features for model accuracy and stability.
- Simulated A/B testing setups to evaluate fraud rules and model effectiveness.
- Deployed models with joblib for reuse in production-like setups, monitoring performance with F1-score and recall.

PROJECTS

HEART DISEASE PREDICTION USING MACHINE LEARNING

JAN 2024

Tools Used: Python, Pandas, Scikit-learn, SHAP, Matplotlib, Seaborn, Jupyter Notebook

GITHUB: ML PROJECT

- Built a binary classifier (Random Forest, Logistic Regression) to detect potential heart disease in patients.
- Performed deep EDA to understand feature importance and risk signals (e.g., cholesterol, age, blood pressure).
- Used SHAP values and feature importance to explain model predictions and reduce false positives.
- Achieved AUC of 0.89 and F1-score of 0.84, optimizing recall to minimize high-risk misclassifications.
- Designed the pipeline for A/B-style testing to compare rule-based vs ML-based alerts.

BULLDOZER PRICE REGRESSION MODEL

OCT 2023

Tools Used: Python, XGBoost, Pandas, Scikit-learn, Matplotlib, Seaborn, joblib, Excel

GITHUB: ML PROJECT

- Created a regression model using XGBoost to forecast auction sale prices of used bulldozers.
- Cleaned and processed over 50,000+ rows, handling missing values and outliers affecting model quality.
- Engineered date, category, and text-based features to improve prediction and model generalization.
- Achieved RMSE < 0.25 and improved operational efficiency by reducing price forecast variance.
- Packaged model with joblib and documented a monitoring framework for pricing risk threshold detection.

TECHNICAL SKILLS

- Languages & Libraries: Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- Machine Learning: Classification, Regression, Supervised Learning, SHAP, XGBoost, Random Forest, Logistic Regression
- Data Handling: SQL (MySQL), Excel, Power Query
- Data Visualization & Dashboards: Power BI, Matplotlib, Seaborn
- Model Deployment: joblib, Pickle, Pipeline Serialization
- Concepts: Risk Modeling, Fraud Detection, Root Cause Analysis, Experimental Design, A/B Testing, EDA, False Positive Reduction, Data Storytelling

CERTIFICATIONS

- Microsoft Certified: Power BI Data Analyst Associate (PL-300)
- AI & Machine Learning Data Science Bootcamp – Udemy | Sept 2024
- ChatGPT Complete Guide: Learn Midjourney, ChatGPT 4 – Udemy | Mar 2024
- Web Design with HTML5 & CSS3 – Udemy | Sept 2023

LANGUAGES KNOWN

- English | Hindi | Telugu
- Availability: Immediate (Remote / On-Site)
- Interests: Risk Analytics, Real-Time ML, Fraud Mitigation, Data-Driven Experiments