JANGA SASHI HARI KRISHNA

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

PROFESSIONAL SUMMARY

Data Scientist with strong skills in Python, SQL, and Exploratory Data Analysis (EDA), experienced in building, optimizing, and deploying supervised and unsupervised machine learning models. Completed internship focused on solving real-world business problems through classification, regression, clustering, and ML model deployment. Applied MLOps workflows to package models using Pickle and Joblib, and explored deployment using Google Colab and AWS. Delivered end-to-end projects such as heart disease prediction using Random Forest and SHAP, and bulldozer price forecasting using XGBoost and regression pipelines. Skilled in using data visualization tools like Power BI, Seaborn, and Tableau to communicate insights to non-technical stakeholders. Recently explored Generative AI with ChatGPT for use cases in EDA summarization and automated reporting.

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

- Developed classification and regression ML models using Python, Scikit-learn, and SQL on business-relevant datasets.
- Conducted EDA to identify trends and patterns, and presented results visually using Matplotlib and Seaborn.
- Built and validated models using metrics like accuracy, precision, F1-score, RMSE, and confusion matrix.
- Deployed models using Pickle and Joblib; tested pipeline integration via Google Colab and basic AWS setup.
- Explored Generative AI tools like ChatGPT for summarizing EDA reports and producing insights in natural language.

PROJECTS

HEART DISEASE PREDICTION USING MACHINE LEARNING

JAN 2024

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

GITHUB: ML PROJECT

- Built a binary classification model to detect heart disease using clinical dataset of patient metrics.
- Performed data cleaning, outlier removal, and feature normalization as part of preprocessing.
- Trained and optimized Logistic Regression, Random Forest, and SVM using GridSearchCV.
- Explained model decisions using SHAP values and feature importance for non-technical interpretation.
- Saved final model using Pickle and demonstrated real-time use case via Google Colab API mockup.

BULLDOZER PRICE REGRESSION MODEL

OCT 2023

Tools Used: Python, XGBoost, Scikit-learn, Pandas, Matplotlib, Joblib, Power BI

GITHUB: ML PROJECT

- Developed a regression model to forecast bulldozer auction prices using 400,000+ records from Kaggle.
- Performed EDA to analyze price trends by year, machine age, and category using Power BI & Matplotlib.
- Engineered time-based and categorical features to improve model performance.
- Tuned XGBoost hyperparameters using RandomizedSearchCV and achieved RMSLE < 0.25.
- Exported trained pipeline using Joblib for deployment and created dashboard for business interpretation.

TECHNICAL SKILLS

- Languages & Libraries: Python (Pandas, NumPy, Scikit-learn, XGBoost, Seaborn, Matplotlib), SQL (MySQL)
- Machine Learning: Classification, Regression, Clustering, EDA, Feature Engineering, Model Validation, MLOps
- Visualization Tools: Power BI, Tableau, Excel, Seaborn, Matplotlib
- Tools & Frameworks: Jupyter Notebook, VS Code, Git, GitHub, SHAP, Pipelines, GridSearchCV, Joblib, Pickle
- Cloud & GenAI: AWS (basic), Google Colab (model deployment), Generative AI (ChatGPT for EDA/reporting automation)
- Soft Skills: Business Analysis, Insight Presentation, Communication, Team Collaboration

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