Sanjana Rao

DATA SCIENCE STUDENT | MACHINE LEARNING DEVELOPER

Mangalore, Karnataka | +91 8296586865 | raosanjana0007@gmail.com
linkedin.com/in/raosanjana07 | github.com/raosanjana07

PROFESSIONAL SUMMARY

Second-year B.Sc. Data Science student with hands-on experience in supervised and unsupervised machine learning, predictive modeling, customer analytics, and deployment. Built and deployed K-Means customer segmentation with a 0.554 silhouette score, tuned XGBoost house-price prediction, and Random Forest churn prediction with 77.9% accuracy. Seeking a Data Science or Machine Learning internship.

TECHNICAL SKILLS

Programming: Python, SQL
Machine Learning: Supervised Learning, Unsupervised Learning, Classification, Regression, Clustering, K-Means, Random Forest, XGBoost
Data & Model Development: Data Cleaning, EDA, Data Preprocessing, Feature Engineering, Cross-Validation, Hyperparameter Tuning, Elbow Method, Silhouette Analysis, Model Evaluation
Libraries & Tools: Pandas, NumPy, scikit-learn, XGBoost, Matplotlib, Plotly, Joblib, Streamlit, Git, GitHub, VS Code

PROJECTS

Customer Segmentation | Deployed Unsupervised Learning Application

Technologies: *Python, Pandas, scikit-learn, K-Means, Plotly, Joblib, Streamlit, Git, GitHub*

- Built an end-to-end unsupervised learning pipeline that segmented mall customers by annual income and spending score using K-Means.
- Used the Elbow Method and silhouette analysis to evaluate cluster quality, achieving a silhouette score of 0.554, and translated the segments into targeted marketing insights.
- Deployed a Streamlit business-intelligence dashboard with Plotly cluster visualizations, a segment explorer, new-customer prediction, CSV batch segmentation, and downloadable executive HTML reports.

GitHub: https://github.com/raosanjana07/Customer-Segmentation
App: https://customer-segmentation-t2enf56jsxpguapplse6ktz.streamlit.app/

House Price Prediction | Deployed Regression Application

Technologies: *Python, Pandas, NumPy, scikit-learn, XGBoost, Plotly, Joblib, Streamlit, Git, GitHub*

- Built a regression pipeline on 1,460 Ames Housing records and 79 predictive variables, covering EDA, missing-value treatment, categorical encoding, feature alignment, and reusable preprocessing.
- Compared Linear Regression, Random Forest, XGBoost, and Tuned XGBoost using cross-validation and hyperparameter tuning; evaluated with R2, MAE, RMSE, residual analysis, and feature importance.
- Deployed a Streamlit dashboard with Plotly visualizations, single and CSV batch predictions, and downloadable reports.

GitHub: https://github.com/raosanjana07/House-Price-Prediction
App: https://house-price-prediction-bpwgp8gyrepgv9sqh5cappn.streamlit.app/

Customer Churn Prediction | Deployed Classification Application

Technologies: *Python, Pandas, NumPy, scikit-learn, Joblib, Streamlit, Git, GitHub*

- Built an end-to-end telecom churn pipeline covering data cleaning, preprocessing, exploratory analysis, feature engineering, and Random Forest training across 30 features, achieving 77.9% accuracy.
- Deployed a Streamlit application with churn predictions, probability scores, model metrics, feature-importance insights, and reusable model, scaler, and feature-schema artifacts.

GitHub: https://github.com/raosanjana07/Customer-Churn-Prediction
App: https://customer-churn-prediction-z8vtrr2kbxca4yfmeqihmj.streamlit.app/

EDUCATION

St. Aloysius (Deemed to be University), Mangalore
Bachelor of Science (B.Sc.) in Data Science | Expected Graduation: 2028
Academic Performance: Semester 1 SGPA 9.32 | Semester 2 SGPA 9.04

CERTIFICATIONS

IBM Cloud Essentials - IBM | March 2026
Python for Mathematics - St. Aloysius (Deemed to be University) | February 2026