Problem Statement: Statlog Vehicle Silhouettes

https://www.kaggle.com/datasets/patriciabrezeanu/statlog-vehicle-silhouettes

To classify a given silhouette as one of four types of vehicle, using a set of features extracted from the silhouette. The vehicle may be viewed from one of many different angles. The task is a multi-class classification problem, where each vehicle must be classified into one of four categories: Bus, Opel, Saab, or Van.

Data Description

Number of instances: 846

Number of features: 18 numerical attributes

Target variable: Vehicle class (4 classes)

Problem type: Multi-class classification

Feature type: Continuous numerical values

The dataset contains geometric and shape-based features extracted from vehicle silhouettes. The objective is to classify each vehicle correctly based on these measurements.

Gitrepo

https://github.com/tusharmulkar/ml-model-app/

Streamlit App

https://ml-model-app-tusharmulkar.streamlit.app/

Model Used

Model	Accuracy	AuC	Precision	Recall	FI	MCC
Logistic Regression	0.6850	0.8715	0.6852	0.6889	0.6679	0.5918
Decision Tree	0.7244	0.8154	0.7216	0.7229	0.7198	0.6335
KNN	0.7165	0.9211	0.7062	0.7171	0.7087	0.6239
Naive Bayes	0.5118	0.7914	0.5650	0.5205	0.4906	0.3780
Random Forest	0.7165	0.9418	0.6973	0.7172	0.7035	0.6242
XGBoost	0.7717	0.9476	0.7640	0.7726	0.7668	0.6963

Observations

Model	Observation
Logistic Regression	Achieved moderate performance with an accuracy of **68.5%**. Relatively High AuC indicates good class separation. Low F1 score shows that model is not able to capture non-linear patterns. Not very effective for chosen dataset
Decision Tree	Improved performance over Logistic Regression, achieving **72.44% accuracy** It effectively captures nonlinear relationships within the dataset. However, its AUC (0.8154) is lower than other models, indicating comparatively weak class separation.
KNN	Both AuC and Accuracy is high makes it very good model for chosen dataset
Naïve Bayes	Naive Bayes showed the weakest performance among all models. Its quite probable that dataset has features which are correlated
Random Forest	Achieved strong and stable performance. indicating reliable overall classification
XGBoost	XGBoost achieved the best overall performance with an accuracy of **77.17%**, F1-score of **0.7668**, and MCC of **0.6963**. It also recorded the highest AUC (**0.9476**), demonstrating superior class separation capability. Boosting allowed the model to sequentially correct errors from previous trees