Machine Learning Project Report

# 1. Problem Statement

The goal of this project is to classify mobile phones into four price range categories: Low, Medium, High, and Very High. The classification is based on several phone features such as RAM, battery power, screen size, and more.

# 2. Dataset Overview

The dataset was sourced from Kaggle:  
https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification  
  
It contains 2000 records and 21 columns, with the target column 'price\_range' indicating the price class (0: Low, 1: Medium, 2: High, 3: Very High).

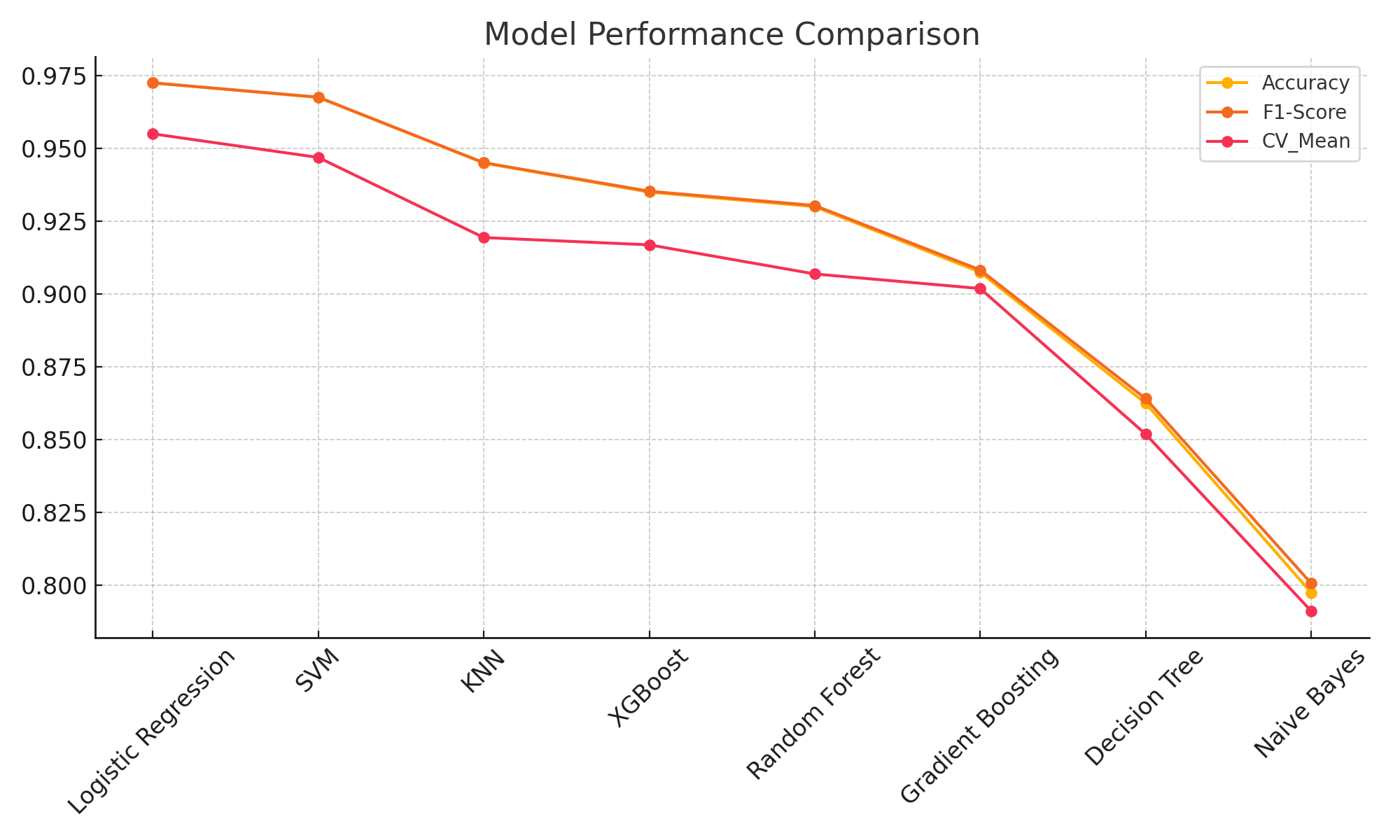
# 3. Preprocessing Steps

- Loaded the dataset and performed exploratory data analysis  
- No missing values were found  
- Features were scaled/normalized where required  
- Used train\_test\_split to divide data into 80% training and 20% testing sets

# 4. Model Training and Evaluation

The following models were trained and evaluated:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | F1-Score | CV\_Mean | Training Time (s) |
| Logistic Regression | 0.9725 | 0.9725 | 0.955 | 0.3265 |
| SVM | 0.9675 | 0.9676 | 0.9469 | 0.2677 |
| KNN | 0.945 | 0.9451 | 0.9194 | 0.0051 |
| XGBoost | 0.935 | 0.9353 | 0.9169 | 0.4212 |
| Random Forest | 0.93 | 0.9304 | 0.9069 | 0.8154 |
| Gradient Boosting | 0.9075 | 0.9083 | 0.9019 | 2.9674 |
| Decision Tree | 0.8625 | 0.8641 | 0.8519 | 0.0226 |
| Naive Bayes | 0.7975 | 0.8007 | 0.7912 | 0.0048 |



# 5. MLflow Tracking

All experiments were logged using MLflow, which tracked parameters, metrics, and models. Artifacts like confusion matrices and performance metrics were logged for easy comparison.

# 6. Cloud Storage Integration

The best-performing model (Logistic Regression) was saved and uploaded to Azure Blob Storage. This enables model deployment in a scalable production environment. Azure Blob was accessed using the official Azure SDK for Python.

# 7. Conclusion and Learnings

- Logistic Regression was the most effective model with the highest accuracy and CV score  
- Simpler models can outperform complex ones with the right feature set  
- MLflow improved tracking and experimentation significantly  
- Azure Blob provides a secure and scalable way to store ML models for deployment