■ Machine Learning Project: Cake Price Prediction

Objective:

The goal of this project is to predict the price of cakes based on features such as size, ingredients cost, design complexity, and other factors.

Dataset Overview:

The dataset contains 4000 records with 8 columns: Sold_On, Size, Ingredients_Cost, Design_Complexity, Time_Taken, Price, Amount, Gender.

Data Preprocessing:

- Encoded categorical features (Sold_On, Size, Design_Complexity, Gender) using LabelEncoder. - Removed extreme outliers using the 95th percentile of Price. - Split the data into 80% training and 20% testing sets.

Model Selection:

A Random Forest Regressor was chosen for its robustness and high performance on structured data. Hyperparameters were optimized using GridSearchCV to find the best combination of: n_estimators, max_depth, and min_samples_split.

Model Evaluation:

- Evaluation Metrics: • Mean Absolute Error (MAE): 5.68 • Root Mean Squared Error (RMSE): 7.39 These values indicate that, on average, the model predicts cake prices within approximately ±5.7 units of the true value.

Prediction Example:

For a medium-sized, simple cake costing 50 in ingredients and taking 3 hours to make, predicted price = 72.26.

Challenges:

- Handling categorical encoding correctly with LabelEncoder. - Ensuring model generalization and avoiding overfitting. - Dealing with data imbalance across cake sizes and design complexity.

Recommendations for Improvement:

- Try advanced models such as Gradient Boosting or XGBoost. - Use One-Hot Encoding instead of Label Encoding for better categorical handling. - Collect more data on cake features such as flavor, decoration materials, and seasonality.