CS 725- Foundations of Machine Learning: Credit Card Reward Maximization

Datasets:  
credit\_card\_transactions (from kaggle)  
cleaned\_dataset\_for\_model (preprocessed)

Code: dataset\_cleaning.ipynb

Preprocessing Steps:

1. Handled missing values.

2. Encoded categorical variables using (Label Encoding).

3. Standardized numerical variables using (tandardScaler).

Key Features:

Categorical Features: - `first`, `last`, `category`.

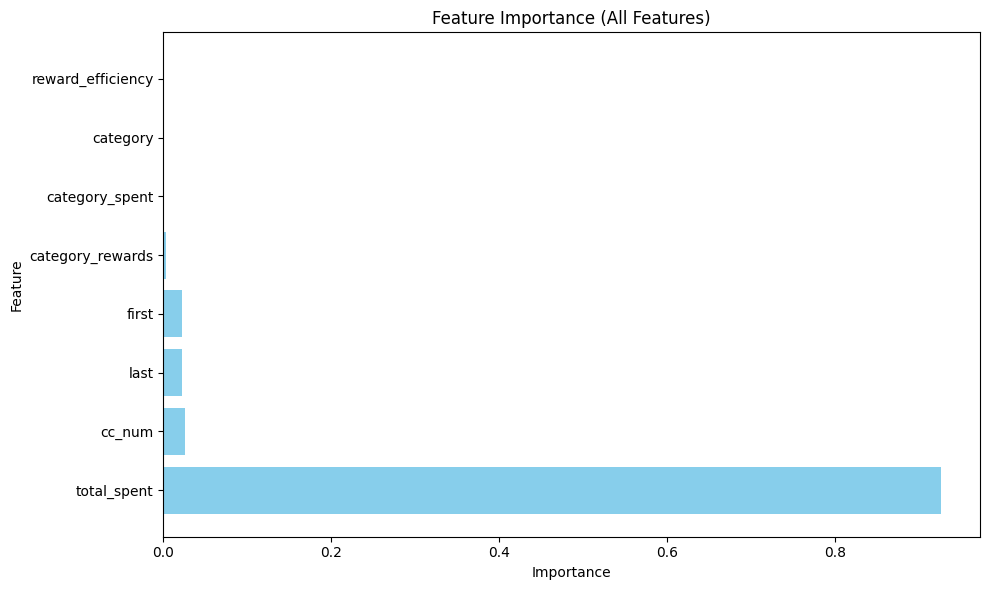
Numerical Features: - `cc\_num`, `category\_spent`, `reward\_efficiency`, `total\_spent`.

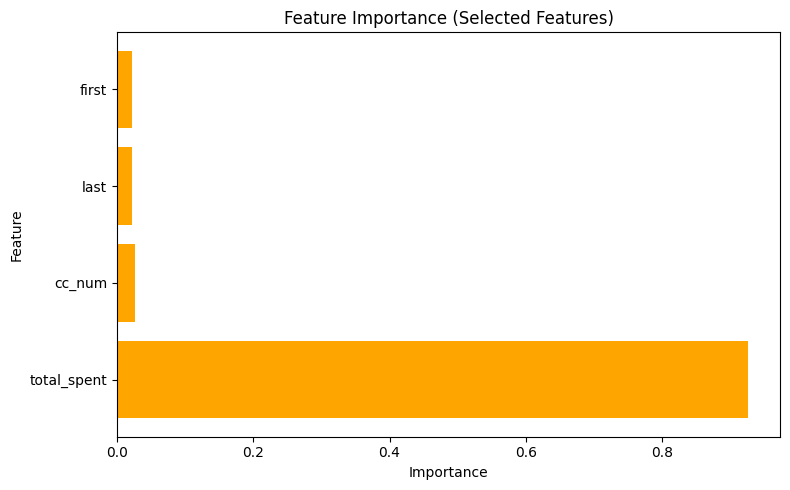
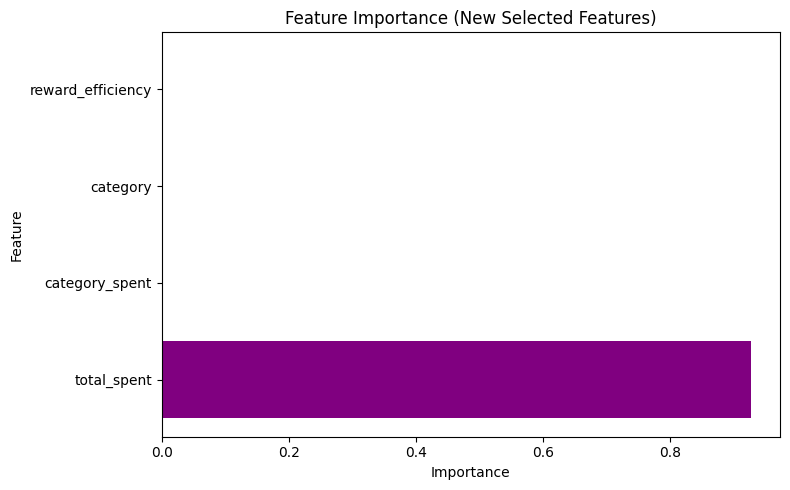
Target Variable: - `total\_rewards` (predicted rewards for users).

Feature Importance:

most impactful features influencing the model's predictions.

Code: feature\_importance.ipnb





Model Training:

Code: model\_training.ipynb

Features= ['first', 'last', 'cc\_num', 'category', 'category\_spent', 'reward\_efficiency', 'total\_spent']

Target= data['total\_rewards']

Split: Divided the cleaned dataset into training (80%) and validation/test (20%) sets.

Model Selection: Random Forest Regressor, LightGBM, XGBoost.

Each model was trained on the prepared training set.

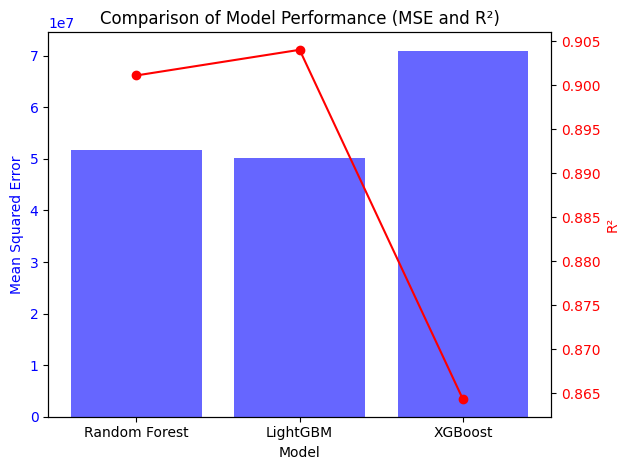
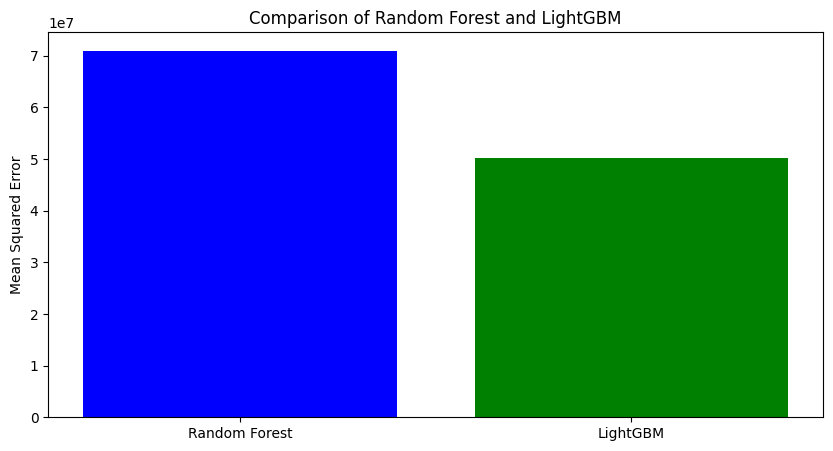
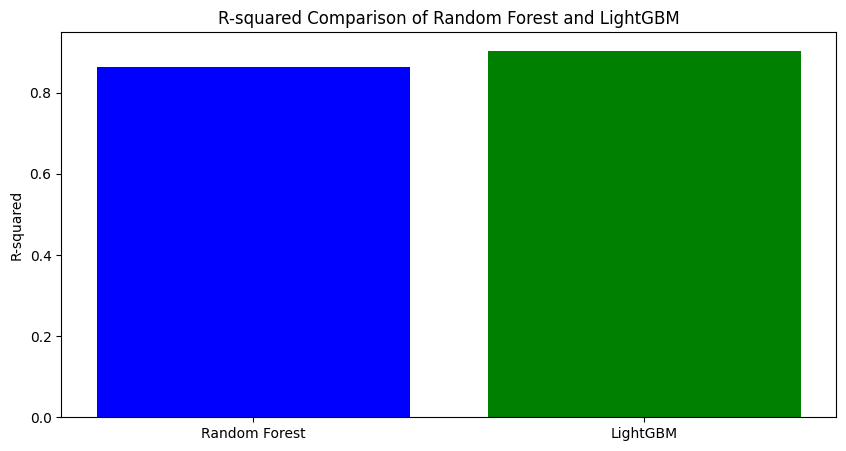
Computed predictions on the validation/test set.

Cross-validation reduced the risk of overfitting and confirmed that the models generalized well.

Hyperparameter tuning significantly improved LightGBM’s performance, leading to the lowest MSE and highest R² values.

Evaluation Metrics: Compared model performance using the following metrics:

Mean Squared Error (MSE), R² (Coefficient of Determination)



LightGBM was the optimal model for predicting credit card rewards due to its balance of speed and accuracy.