Credit Risk Behavior Score Prediction – Final Report

Overview:

The task was to predict the next month default of some credit card users given in the validate_dataset_final.csv which has been successfully done by using Extreme Gradient Boost classifier.

The Exploratory Data Analysis was done and then feature engineering and feature selection process was done on the train_dataset_final.csv to improve the efficiency of the models finally best classification model was chose on the basis of their F-Beta(F2) metric which is LGBM Classifier.

EDA findings and visualizations:

- To start the exploratory data analysis, the Pandas Profiling was used to see the profile report which showed various insights such as the imbalance structure of the data and null rows etc.
- Only nearly 0.1% of the data was null so it was dropped from the dataset to proceed further.
- Only nearly 22% of the customers defaulted and rest had 0 default value in the training data, so SMOTE oversampling was performed to make the data balanced.
-

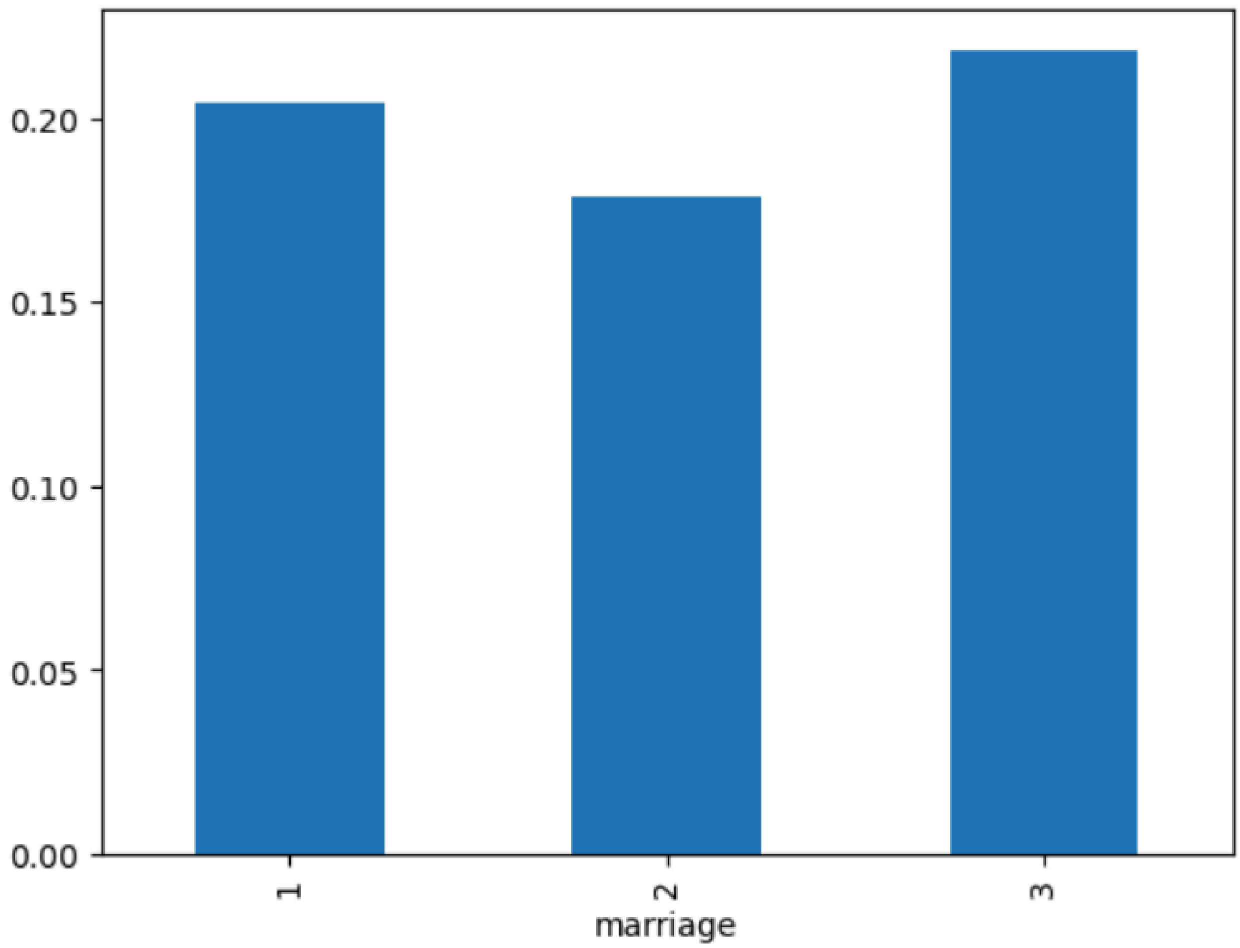


Figure 1

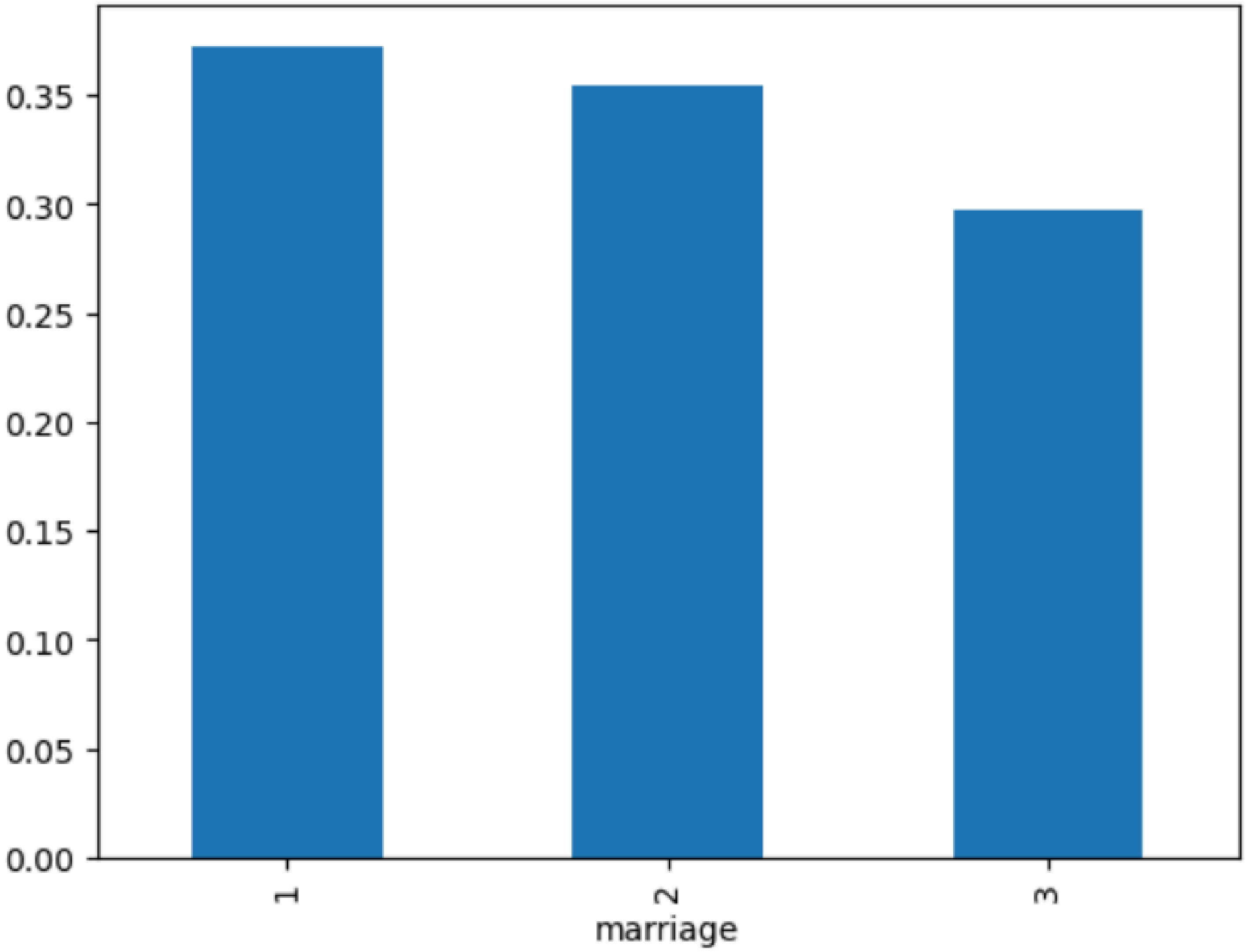


Figure 2

The behaviour of married and unmarried people for default is same but rest (divorced, etc.) have both high default count and low pay to bill ratio.

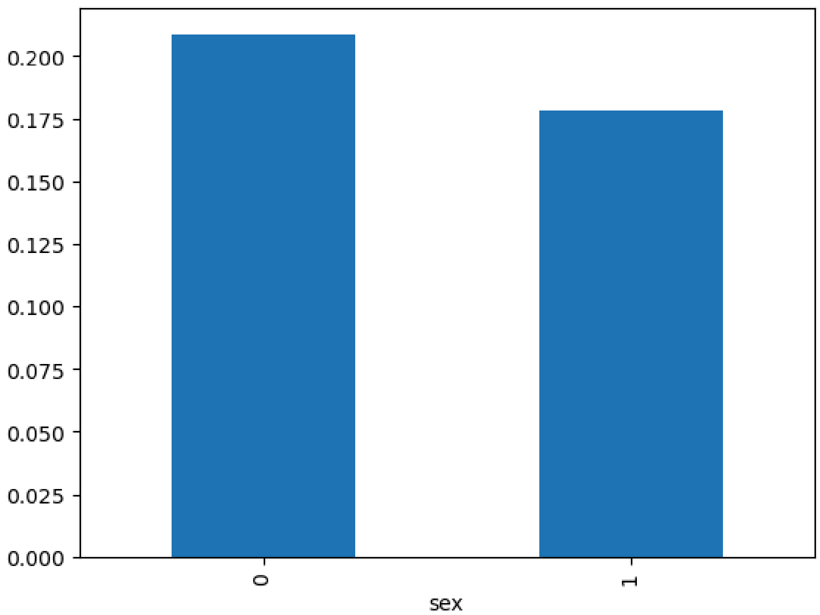
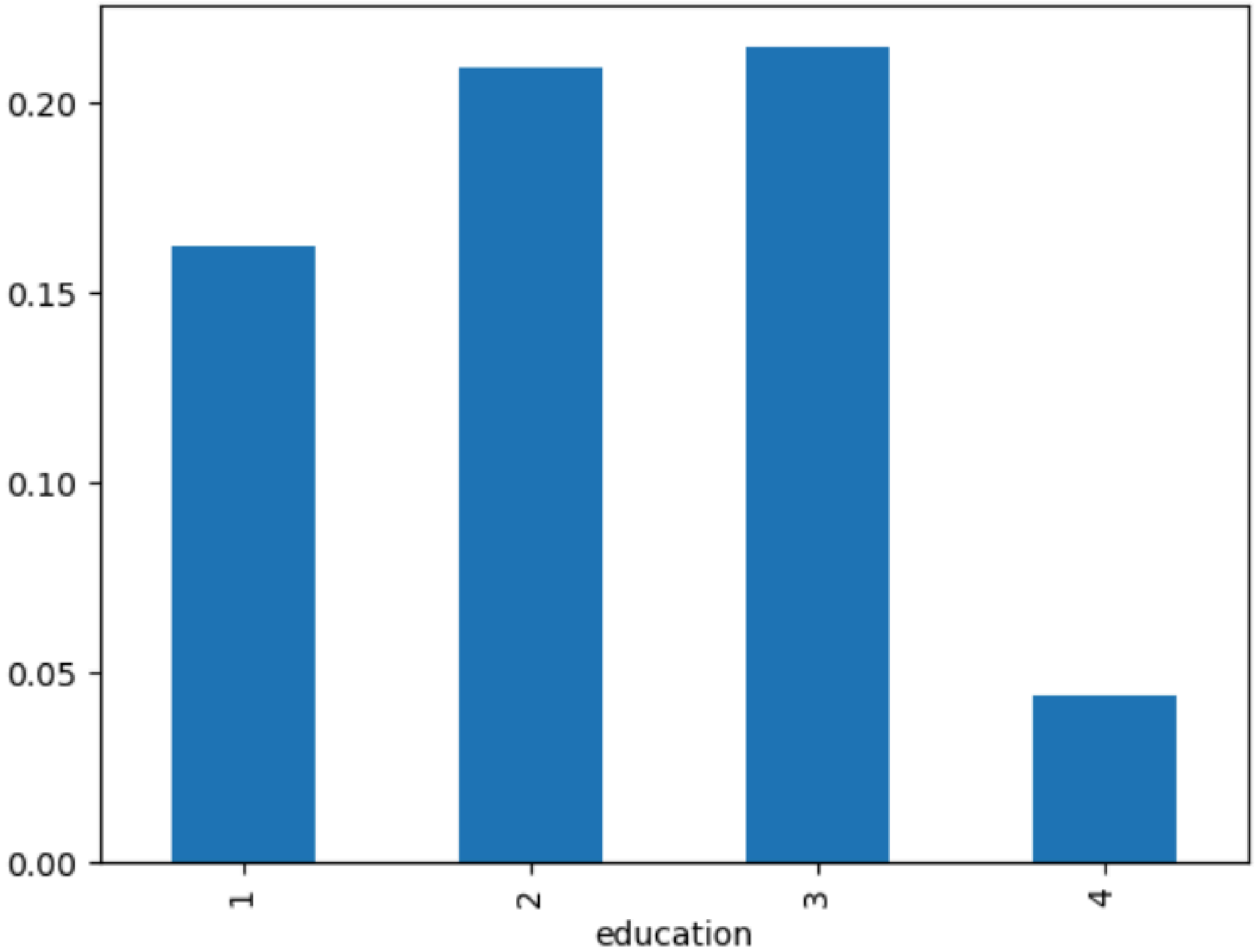
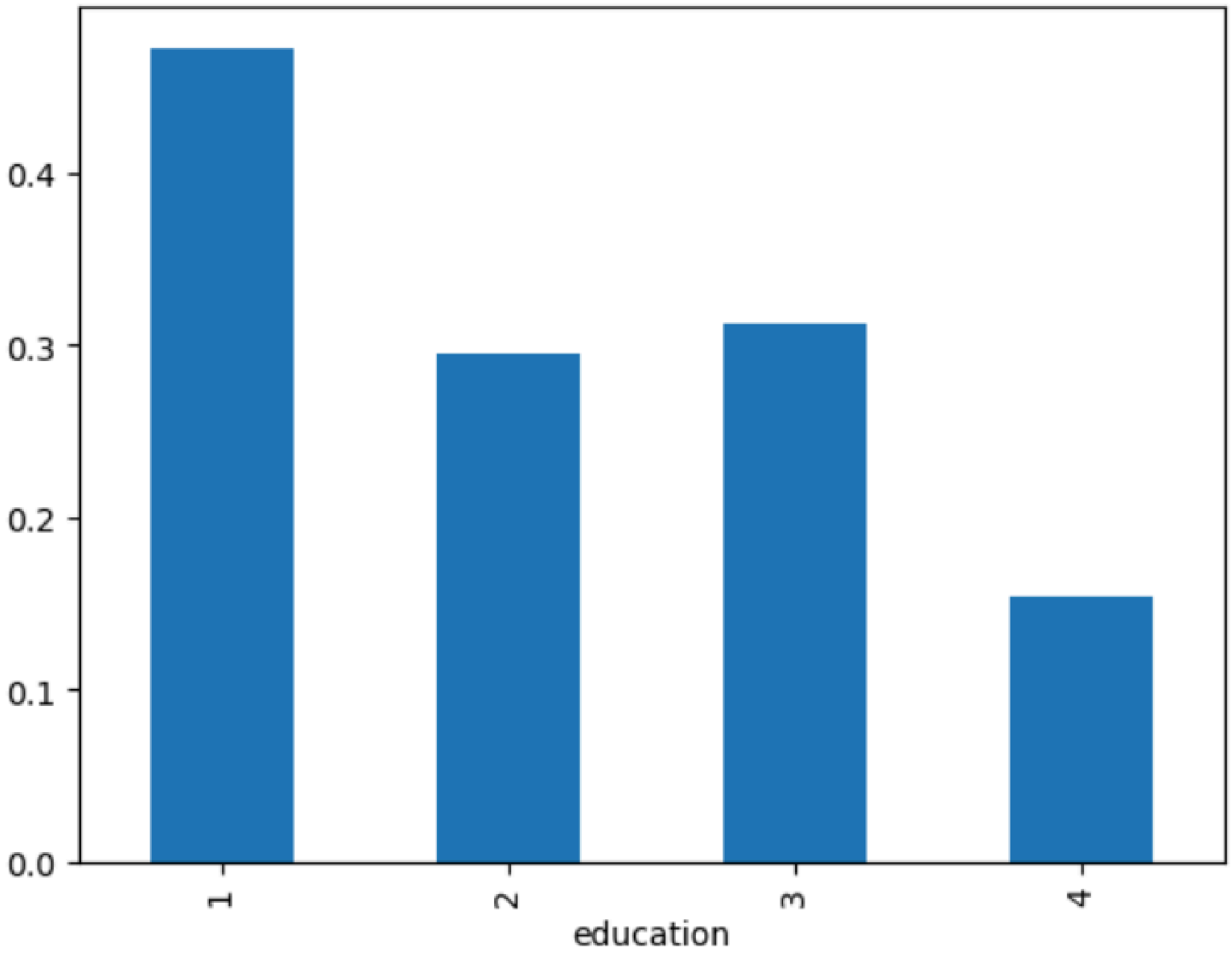
- 

Figure 3

Females default more as compared to males as per the graph (fig. 3) even females being less in number.

-





More educated people default less and have high pay to bill ratio.

- Other EDA graphs are provided in the codes file.
- Correlation was fetched using LGBM feature importance method which gave insight about important features to be selected for high metric's values, later the top 10 were selected.
- Got an idea about mean, median, mode, standard deviation, etc. of features and got to know which features are categorical and which are numerical and how many types of values are present in categorical features.

Financial analysis of features driving the next month default:

We derived a few features from the original ones which were thought to be important for the default such as

- credit_util_ratio = average bill amount / credit limit
- delinquency_count = number of months with overdue payments
- overpayment_count = bill less than zero
- longest_delinquency_streak
- percent_fully_paid_months

After all these the top 10 default predicting features were selected using the lgbm.feature_importances_ method of Light GBM Classifier, and they are namely:

['Bill_amt1', 'pay_amt1', 'credit_util_ratio', 'pay_amt3', 'pay_0', 'pay_amt2', 'PAY_TO_BILL_ratio', 'delinquency_count', 'pay_amt5', 'age']

Model comparison and justification for final selection:

The models which were trained and tested are logistic regression classifier, decision tree classifier, extreme gradient boost classifier, light gradient boosting machine classifier. Out of which the one which gave best F2 metric value was light gradient boosting machine classifier with F2 score of nearly 66%, solving major issue in credit risk management task.

Why F2 metric is most important:

In case of credit risk management, we needed to detect the defaulters to minimize the loss and for that we chose F2 score also called F Beta score.

F2 score gives more weightage or importance to recall as compared to precision.

What Recall means is the percentage or ratio of people who were predicted default and actually defaulted with respect to the actual total number of defaulters.

Recall = [1,1] / [[1,1] + [1,0]]

Precision = [1,1] / [[1,1] + [0,1]]

F2 score = (1+ beta**2)/beta**2[precision*recall/(precision + recall)]

And this metric is important because catching the defaulters is the most important thing in credit risk management.

Metric result of LGBM model on train dataset:

- F1: for 0 its 0.40 and for 1(default) its 0.37
- F2: 0.66
- Accuracy: 0.39
- Recall: 0.93 for default

Selection of classification cutoff or threshold:

The threshold is selected on the basis of best F2 Score using a loop of threshold values from 0.1 to 0.9 and it comes out to be 0.30 for best F2 score of 66%.

Business Implications:

- Loan officer of the credit card company can detect defaulters and avoid giving loan to them.
- Earlier detection of defaulters can help decrease churn rate because the company can suggest some offers so that customer may accept and continue to give outstanding money.
- Some customers may be unsatisfied due to wrongly detected as default since no model can be perfect.
- Classification cutoff can be tuned manually by the company as per their need.

Summary of findings and key learning:

This project focused on building a binary classification model using Light GBM to detect a rare but critical class, where missing positive cases could have serious business consequences. Due to the need of company the F2-score was used as the main evaluation metric to prioritize recall. Threshold tuning was performed across a range of values to identify the best cutoff that maximized the F2-score. The final model achieved high recall for the positive class, ensuring most important cases were captured, though with a trade-off in precision. This approach is valuable in credit risk where catching default cases is more crucial than non-defaulters. The project highlighted the importance of choosing the right metric and adjusting thresholds based on business needs. It also demonstrated the strengths of Light GBM in handling tabular and imbalanced data. Overall, the model serves as a strong decision-support tool, improving efficiency and reducing human oversight.

On the validation data it predicted 3222 as defaulters and 1794 as non-defaulters.