

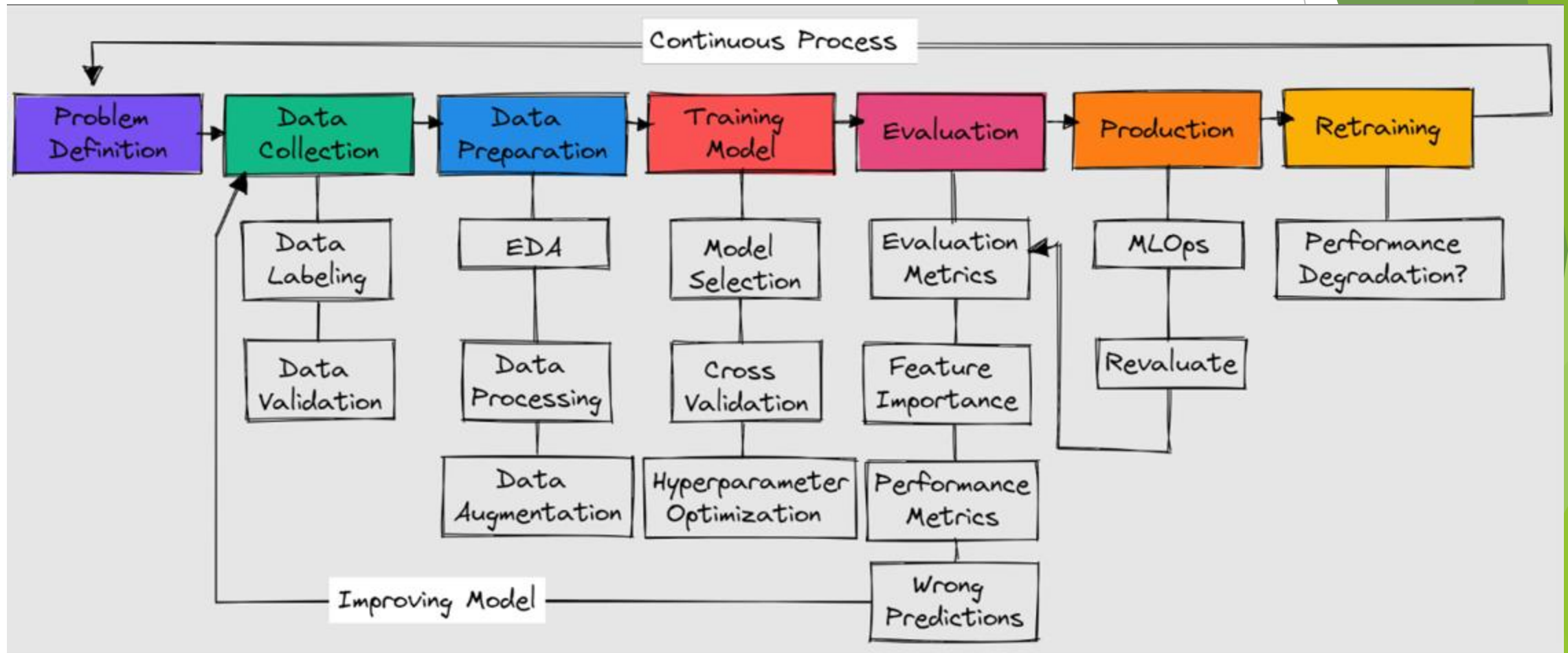
CREDIT CARD DEFAULT PREDICTION

Submitted by: Pooja Rathore

OVERVIEW

- Banking/Financial Institutes play a significant role in providing financial services.
- To maintain integrity, banks/institutes must be careful when investing in customers to avoid financial loss.
- Before giving credit to borrowers, the bank must come to about the potential of customers.
- The term credit scoring, determines the relation between defaulters and loan characteristics.

ARCHITECTURE



DATA PREPROCESSING

- Data set is divided in 80:30 ratio for train and test respectively. ID column was dropped as it was unnecessary for our modelling.
- The attribute name 'PAY_0' was converted to 'PAY_1' and 'default.payment.next.month' was converted to 'default' and later "DEFAULTER" for naming convenience.
- Pay_0: No consumption of credit card=-2, Pay duly (paid on time)=-1, payment delay for one month=1, payment delay for two months=2, payment delay for nine months and above=9.
- No Null values in the dataset

DATA ANALYSIS

- There are more women than men in our dataset and men have a slightly higher chance of default.
- The probability of default was higher for men.
- Most people in our dataset are between 25 and 40 years old. There is also an impression that around that age the chance of default is a little lower.
- Most customers have a 200k or less credit limit. And it seems that we will find a higher concentration of customers in default on that range.
- Those who have a negative bill statement have a lower chance of default than the rest. What stands out is that there is a slightly higher chance of default for those who didn't have a bill in the previous months.
- There is a higher default rate among those who paid nothing in previous months and lower rates among those who paid over 25k NT dollars.

MODEL SELECTION & PREDICTION

Here we train and select the best machine-learning model for predicting credit card defaults based on the provided data. We tried and tested multiple models such as LogisticRegression, Support Vector Classifier, KNeighborsClassifier, RandomForestClassifier, GaussianNB, AdaBoostClassifier, and GradientBoostingClassifier for the model and came up with the model with the best performance.

The Accuracy of GradientBoostingClassifier was 82.22 and Accuracy_score was 69(around).

CONCLUSION

- We investigated the data, checking for data unbalancing, visualizing the features and understanding the relationship between different features.
- We used train-test split to evaluate the model's effectiveness to predict the target value i.e. detecting if a credit card will default next month.
- We started with LogisticRegression, Support Vector Classifier, KNeighborsClassifier, RandomForestClassifier, GaussianNB, AdaBoostClassifier, and GradientBoostingClassifier, the accuracy all are different.
- We choose the GradientBoostingClassifier model based on the Accuracy score which was higher than other models.
- This would also inform the issuer's decisions on whom to give a credit card to and what credit limit to provide.

Thank You