

Predictive Analysis on Credit Card Attrition

Danhe Tang
Purdue University, Daniels School of Business
tang631@purdue.edu

ABSTRACT

This project evaluated classification models to determine the most accurate model for banks to better predict the credit card attrition. Credit card attrition can have negative impacts on banks, causing revenue loss. It's a imperative capability for banks to predict credit card customer churn as it allows them to address customer concerns and retain business to boost the overall profit. I built and evaluated K-Nearest Neighbors (KNN), Logistic Regression and Decision Tree models of 10,127 observations and select the most accurate model to forecast if the customer will churn based on misclassification rate. The model provides a way for bank practitioners to strategically identify the potential customer churn.

BUSINESS PROBLEM

In recent years, credit card attrition has emerged as an issue of significant concern for the banking sector. As shown in Figure 1, credit card is one of the biggest revenue sources of corporate banks. However, each corporate bank faces the challenge of customer attrition, especially giving that the cost of acquiring new customers is higher than that of retaining existing customers. In respect with the high proportion the credit card revenue making up the total revenue, we can tell credit card attrition has a significant impact on profitability. In Figure 2, BCG research shows that corporate banks annually lose 10% to 15% of gross revenues to attrition. Thus, predicting credit card attrition (churn) rate is an imperative task to banks. The main stakeholders of this study are corporate bank practitioners, who want predict credit card churn to boost gross revenue. The goal of this analysis is to help the bank practitioner to have a good prediction on credit card attrition and increase the revenue by using this information.

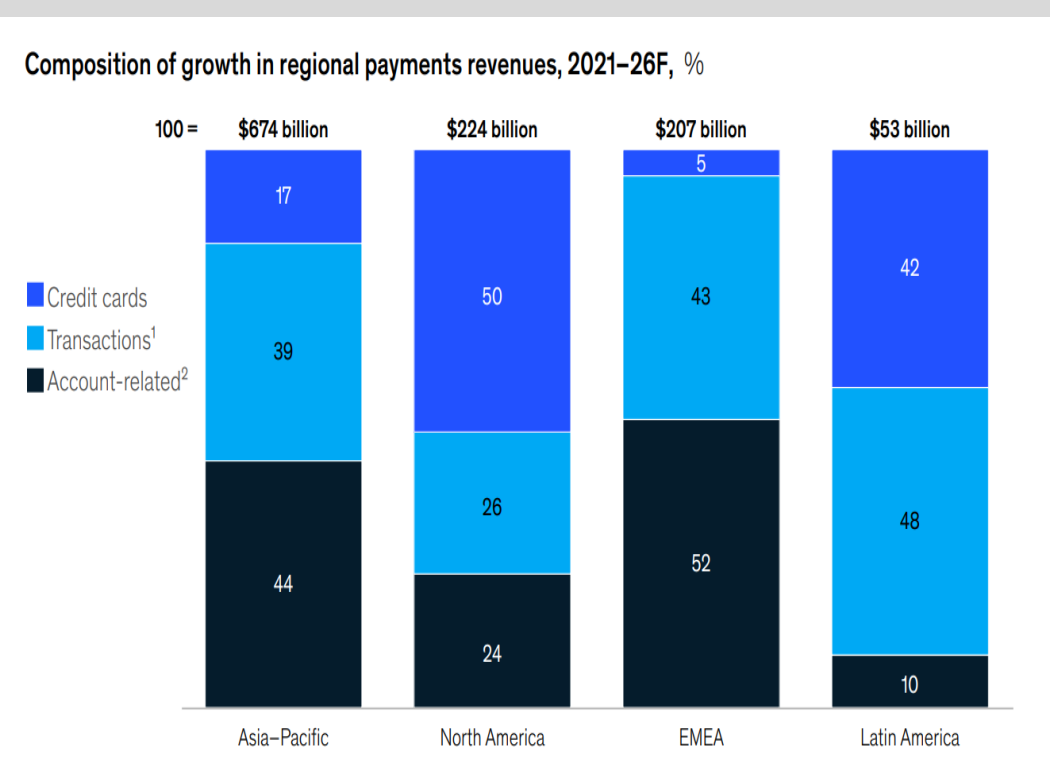


Fig1. Credit Card payment revenue growth percentage in bank revenue

Source: <https://www.mckinsey.com/industries/financial-services/our-insights/global-banking-annual-review>



Fig2. Banks gross revenue loss caused by attrition

Source: <https://www.bcg.com/publications/financial-institutions-marketing-sales-how-banks-close-back-door-attrition>



ANALYTICS PROBLEM FRAMING

To realize the business goal, the goal for this analysis task is to build a predictive model of bank customer churn and determine whether customers will churn or not. This analysis problem can be identified as a classification problem, given that the credit card customers can be divided into attrited customer and existing customer. The task is to predict the target feature, Attrition.



RESEARCH QUESTIONS

1. What predictive model would be the best in assisting banks in predicting if the credit card customer will churn.
2. How accurately can we predict credit card customer attrition (churn).

DATA

This dataset sourced from Kaggle contains 10,127 samples and 20 features. By looking at the distribution, the age of customers ranges from 26 to 73, with 46.3 on average. The ratio of customers churning vs not churning is which is 8500:1627.



Fig3. Distribution of features (part)

Steps below are conducted in data preprocessing in this case.

1. Feature engineering: Delete the CLIENTNUM, which is customer number that will not help in further classification.
2. Handle outlier: Identify that Months_on_book has outliers. Delete the value smaller than 16 or larger than 52 in this feature.
3. Handle missing value: No missing value identified, so skip imputation.
4. Standardize numerical value.
5. Before building classification model, choose the validation set approach for this case.

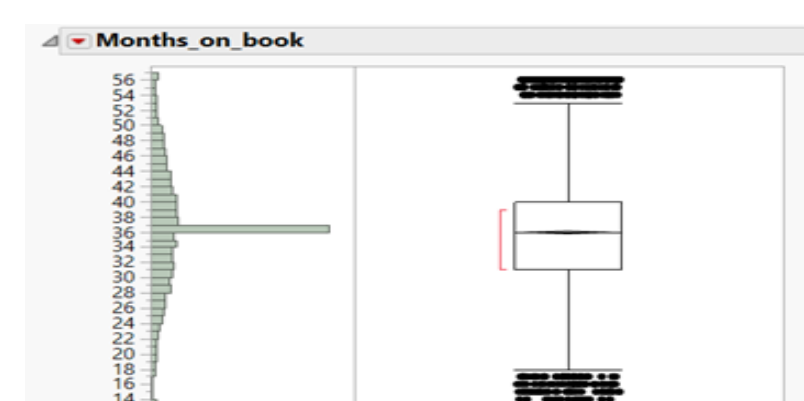
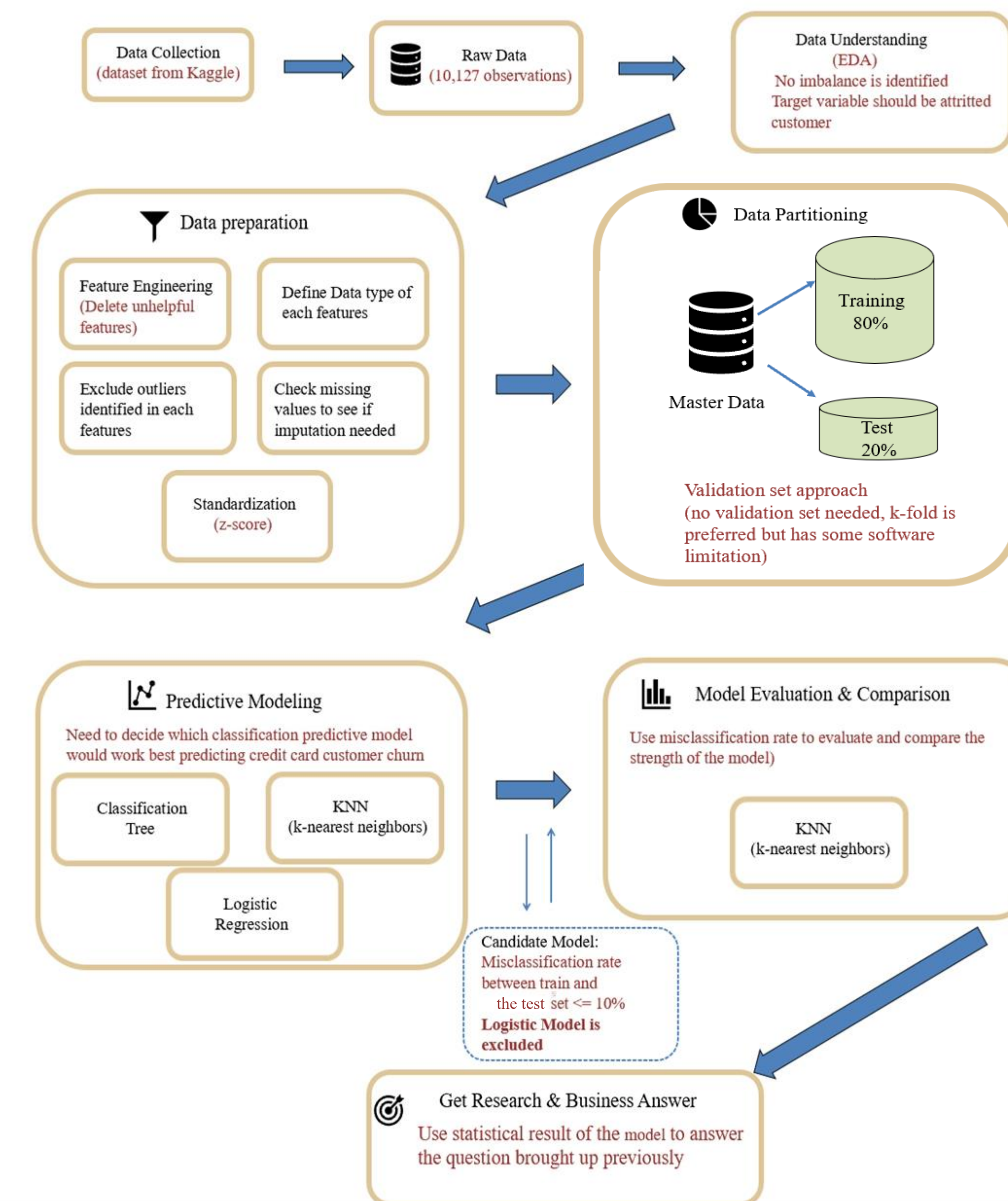


Fig4. Outliers in data



Mitchell E. Daniels, Jr.
School of Business

METHODOLOGY

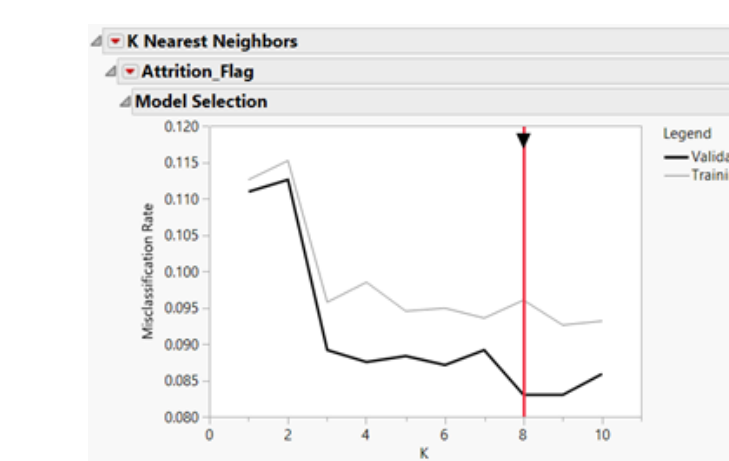


MODEL BUILDING AND EVALUATION-STATISTICAL PERFORMANCE

The statistical measurement used in this analysis will be misclassification rate. The misclassification rate shows how often your confusion matrix is incorrect in predicting the actual positive and negative outputs. I set the measurement difference between training and test set greater than 10% will be considered as over-fitting. Logistic model was excluded from this analysis. The candidate model would be decision tree and KNN. After looking at misclassification rates, the KNN model is decided to be the best model because it has the lowest misclassification rate. (When k = 8)

Validation	Creator	Entropy	Generalized	Mean	Misclassification
Training	Fit Nominal Logistic	0.4550	0.5636	0.2368	0.2670
Training	Partition	0.4419	0.5494	0.2429	0.2682
Training	K Nearest Neighbors			0.1441	0.0996
Test	Fit Nominal Logistic	0.2547	0.3695	0.1535	0.1777
Test	Partition	0.3109	0.4501	0.1492	0.1393
Test	K Nearest Neighbors			0.3302	0.0915
					0.0884

Fig5. Model Comparison



Since this task is to predict whether a customer churns or not, I used the classification accuracy to evaluate the effectiveness of the model, which defined as:

$$ACC = \frac{Samples_true}{N}$$

(Samples_true is the number of correctly classified samples and N is the total number of samples.)

92.2% of the churn can be identified by the KNN model accurately. Beside, through the decision tree model, I also obtained the key factors that lead to customer churn, like total transaction amount, total revolving balance, total count change etc.

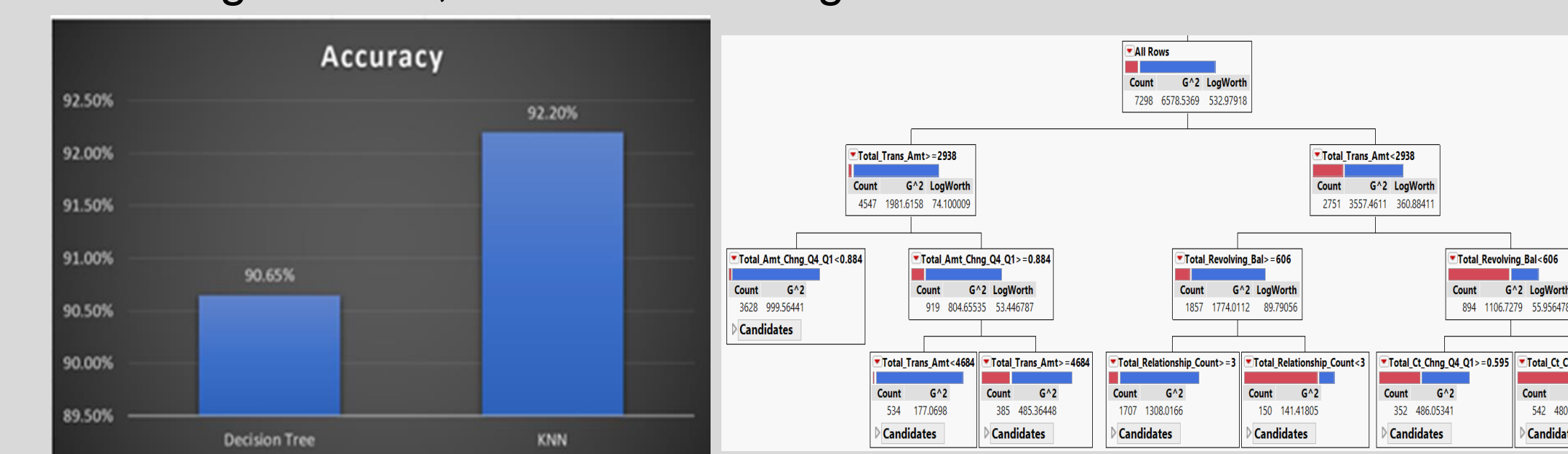


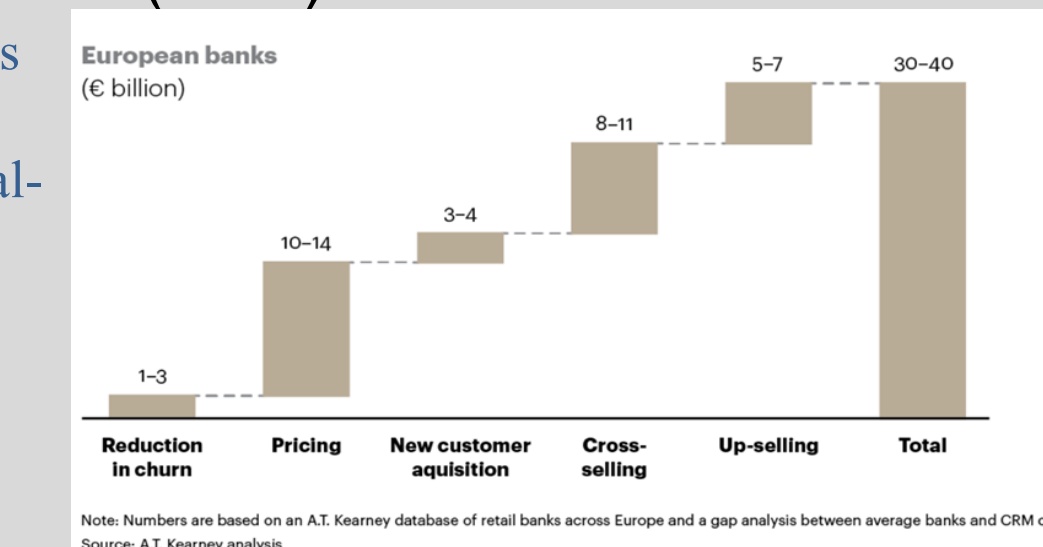
Fig6. Model Accuracy comparison

Fig7. Possible factor leading churn presented by decision tree

MODEL EVALUATION – BUSINESS IMPLICATIONS

The model can help corporate banks with credit card service to predict if the customers owning their credit card will churn. The biggest benefit of having the accurate prediction of potential credit card attrition is that, banks practitioners can take in-time action for further customer relationship management like retention strategies to retain the customers, reducing the churn rate and cost of acquiring new customers. In European banks, reduction in churn will make a 1-3 billion euros approximately growth in the gross revenue. Thus, predictive model applied will make approximately 1.1-3.2 billion dollars increase in revenue to banks in US. To every bank, it is possible for a retention by 5% leaded by accurate predict churn to make a 25% to 95% increase in profits. (Bain)

Fig8 Revenue potential for European banks
Source: <https://www. Kearney.com/industry/financial-services/article/-/insights/cultivating-the-customer-relationship-in-banking>



CONCLUSIONS

It's crucial for a bank to have an accurate prediction on credit card attrition, as it helps to identify the potential churn and take action to boost the total bank's profitability.

- KNN was able to predict the attrited customer in credit card service with a 0.0780 classification rate, which is about 2% better than other candidate model.
- The accuracy of the KNN model to predict the attrition is about 92.20%.

Limitations:

1. Machine learning has numerous algorithms in classification but only few of them are conducted in this analysis.
 2. There is only one dataset collected and processed from a specific bank, which only represents a small part of the industry.
- The model can be further improved by collecting more real-time data on credit card attrition in the industry, and more models can be built to be compared with the current one to see if there are better one to apply.