

PREDICT BANK CHURNING

USING A KAGGLE DATA SET

Business Problem

A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate if one could predict for them who is gonna get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction.

Variables Used

Attrition Flag: Whether the customer was churned or not

Customer Age: Age of Customer

Gender: The customer's gender, 'male' or 'female'

Dependent count: The number of dependents

Education Level: Can be uneducated up to Doctorate

Marital Status: single, married, or divorced

Income Category: Broken every 40k starting at 40 to 120

Card Category: The type of credit card issued by bank

Months on book: How long they have used the bank

Total Relationship Count: number of products held

Months Inactive 12 mon: No. of months inactive in the last 12 months

Contacts Count 12 mon: No. of Contacts in the last 12 months

Credit Limit: Credit Limit on the Credit Card

Total Revolving Bal: Total Revolving Balance on the Credit Card

Avg Open To Buy: Open to buy credit line

Total Amt Chng Q4 Q1: Change in transaction amount

Total Trans Amt: Change in transaction amount

Total Trans Ct: Total Transaction Count

Total Ct Chng Q4 Q1: Change in Transaction Count

Avg Utilization Ratio: Average Card Utilization Ratio

Data

Cleaning/Splitting

Cleaning the data was pretty

easy, you can see the

screenshots below, but all we

had to do was change the

data types for the object

variables to integers so that

we can use them for

analytics. We did a 60-40 split

because we had roughly

10,000 data points and

others used the same split. I

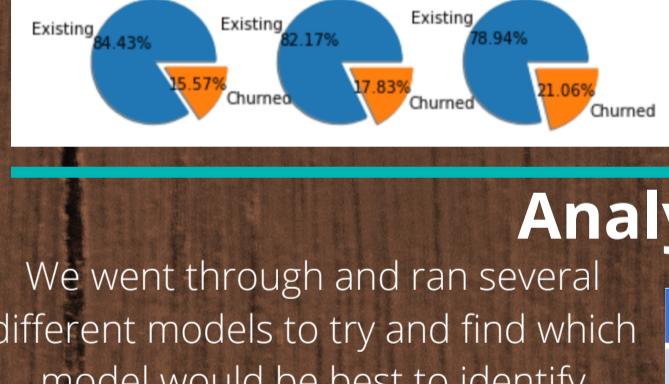
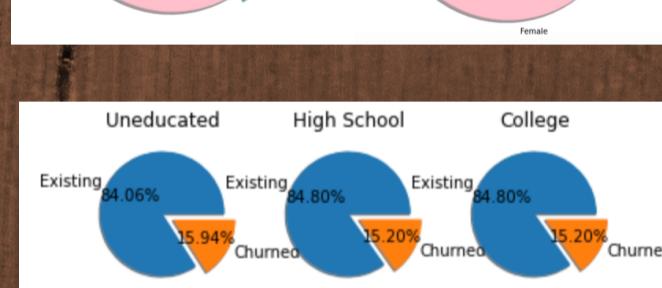
felt that since it was the

established practice for this

data set, it was appropriate

```
bc['Gender'].value_counts()
F 6368
M 4769
Name: Gender, dtype: int64
#Cleaning up the gender variable
#Male is 1 and Female is 0
bc.loc[bc["Gender"] == "M", "Gender"] = 1
bc.loc[bc["Gender"] == "F", "Gender"] = 0
bc['Gender'] = bc['Gender'].astype(int)
bc['Gender'].value_counts()
0 5358
1 4769
Name: Gender, dtype: int64
```

Exploratory Data



One of the first things we looked at was the demographics of existing customers and churning customers. As you can tell, this was not much of an indicator being that they all had similar churn rates.

Analysis

We went through and ran several different models to try and find which model would be best to identify customers that are churning. After running several models we discovered the best model to be the Random Forrest Classifier because it has the highest AUC Score.

Model Type	AUC Score
Random Forrest Classifier	0.930
AdaBoost Classifier	0.924
Gradient Boosting Classifier	0.923
GaussianNB	0.786
K Neighbors Classifier	0.762

Conclusion

The best model to predict which customers will churn is the Random Forrest Classifier because its AUC Score was 0.93 and it also had a high F1 score. The credit card company should use this model to predict which customers are going to churn.