



# Credit Card Default

By: Anthony Schams & Ngoc Tran  
August 16, 2019

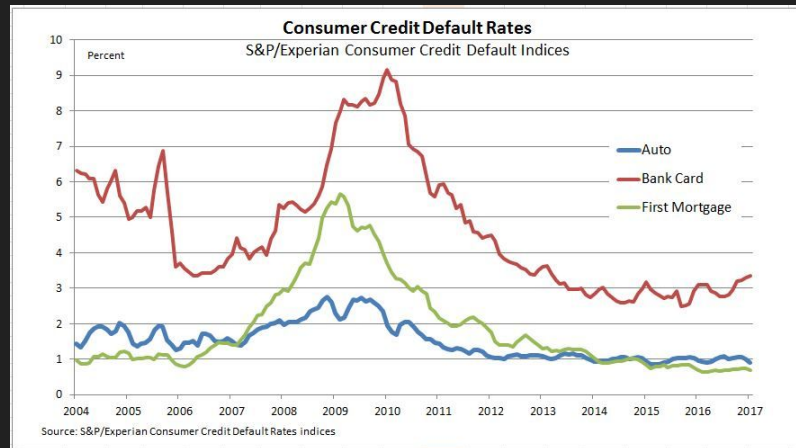
# Introduction

## Goal:

- Predicting who is about to default on his/her credit card

## Data:

- UCI Machine Learning Repository
- Number of Instances: 30,000
- Number of Attributes: 24



- Predictors:
  - credit limit
  - basic personal info
  - past payments and bill statements info
- Target Variable:
  - default payment next month

# Processed Data

Convert all categorical variables into dummy variables.

Transform all numerical features using Yeo-Johnson transformation.

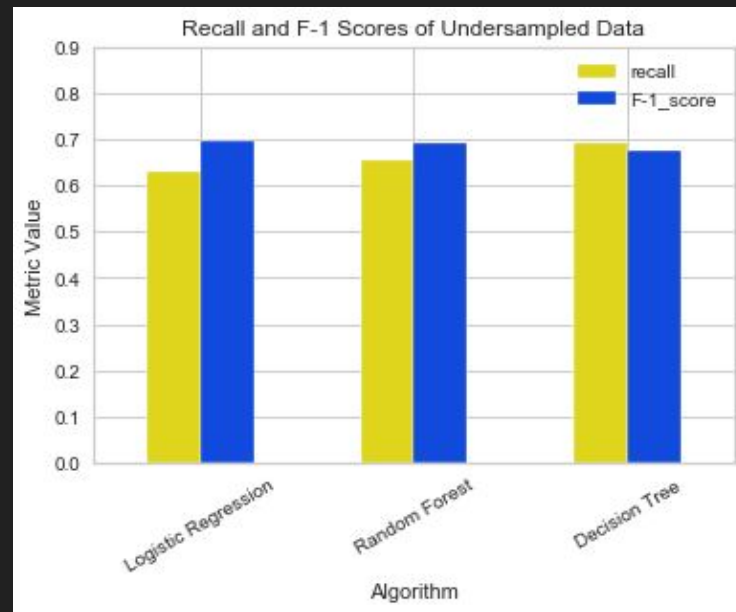
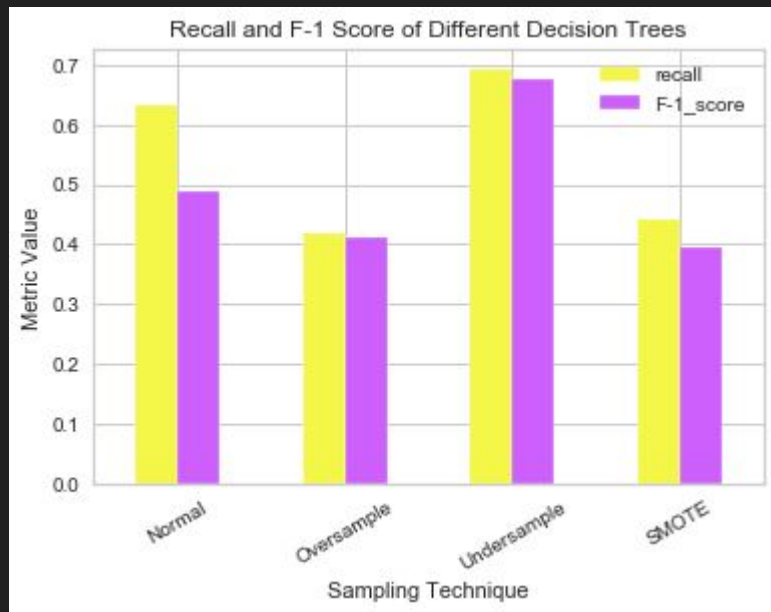
Scale all numerical features using Min-Max scaler.

Final datasets:

- As-is
- Over-sampled
- Under-sampled
- SMOTE



# Models Comparison



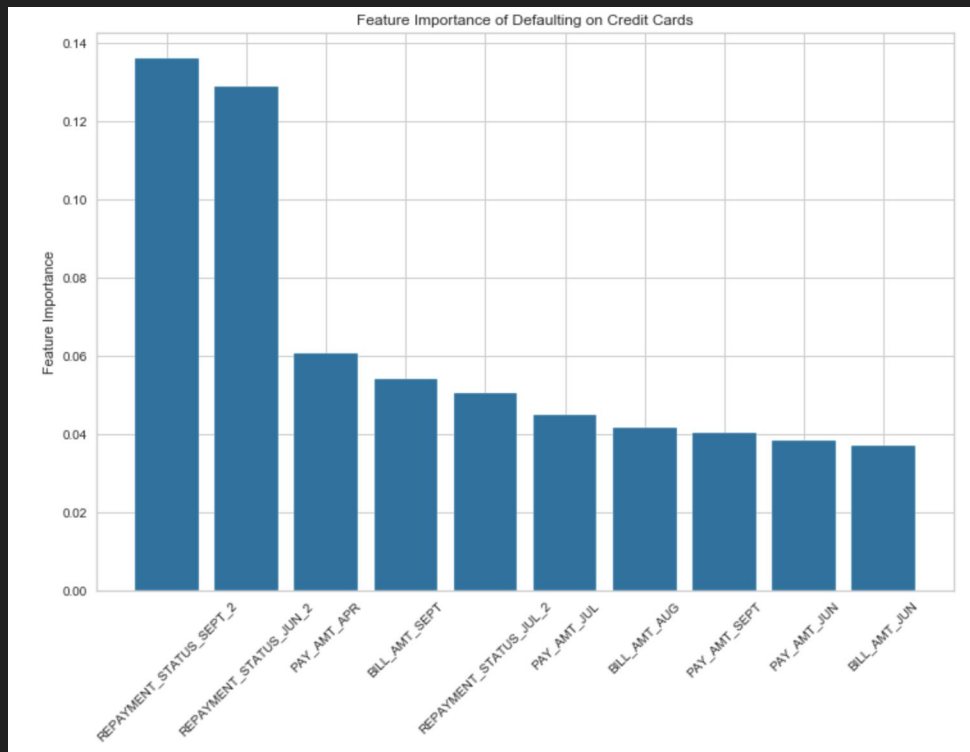
# Best Model

Model:

- Decision Tree

Data:

- Under-sampled



# Conclusions

- Decision Tree on undersampled data yielded the best model, according to recall
- Being 2 months late on payments is a strong signal that a customer will default
- Bill amount also associated with default prediction

# Next Steps for Future Improvements

- Collect more (and relevant) features (IS\_EMPLOYED, SALARY, etc.)
- Collect more data
- Try more Machine Learning algorithms
- Try Anomaly Detection algorithms

Questions?



THANK YOU!

# Sources

- <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>
- <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PowerTransformer.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
- <https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>
- [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)
- <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>