Credit Card Fraud Detection Mini Project

George Messih, Matt Zambetti, Siddharth Akalwadi, Judd Babbitt

Meet our Team

Matt Zambetti
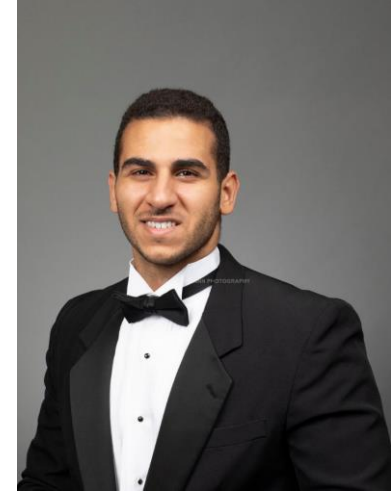BS/Meng Computer Science
Blacksburg Campus

Siddharth Akalwadi
BS/MEng Computer Science
Blacksburg Campus

Judd Babbitt
BS/Meng Computer Science
Blacksburg Campus

George Messih
M.S. Industrial Systems Engineering
Virtual Campus

Table of Contents

Background

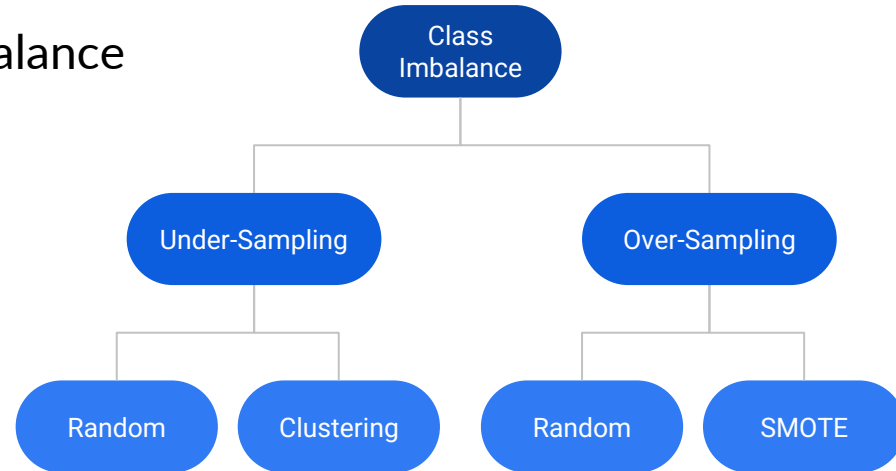**What is Credit Card Fraud?**

How can Credit Card Fraud be detected?
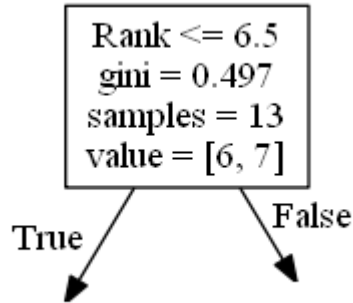
Why Credit Card Fraud?

Data Pre-Processing

- Data ingestion
    - 1990004x11
- Split into train-test sets
    - 75%, 25%
- Overcome class imbalance

True	Fraud
1987514	2490
~99.9987%	~0.0012%

Decision Trees

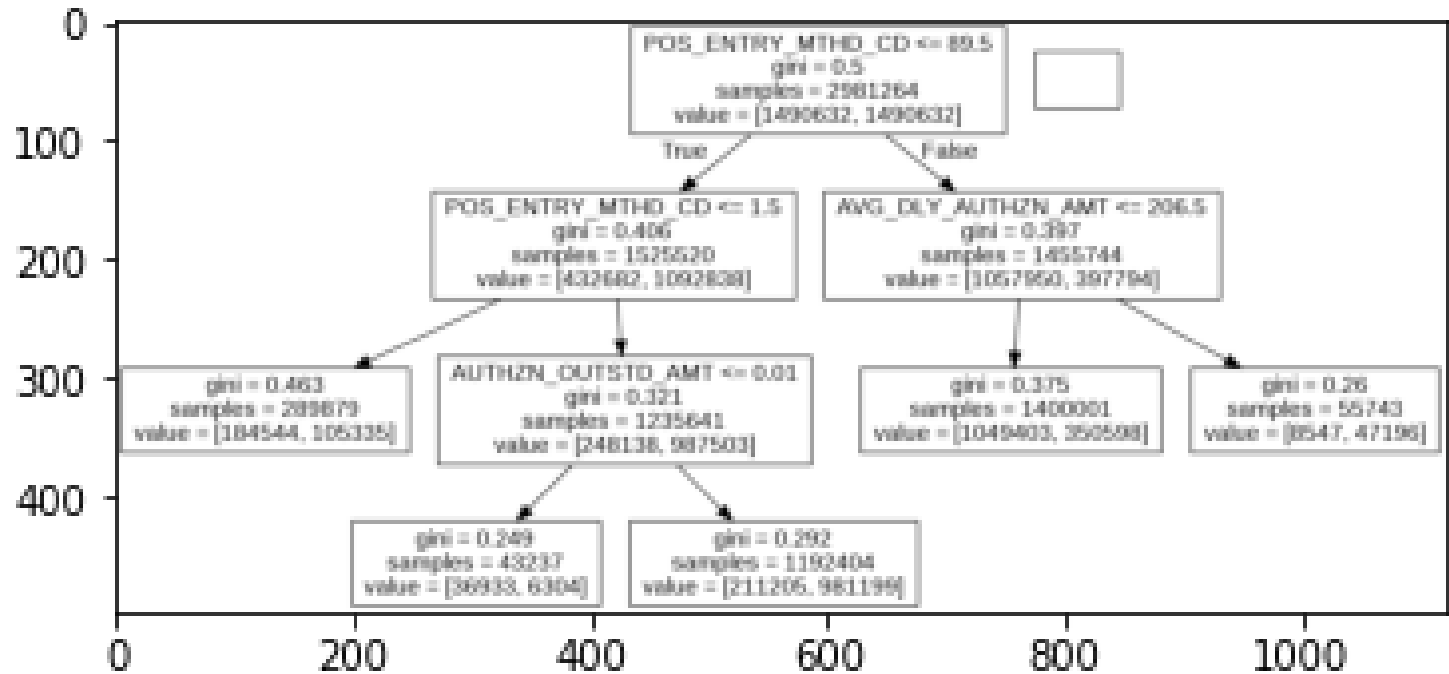Rank <= 6.5
gini = 0.497
samples = 13
value = [6, 7]

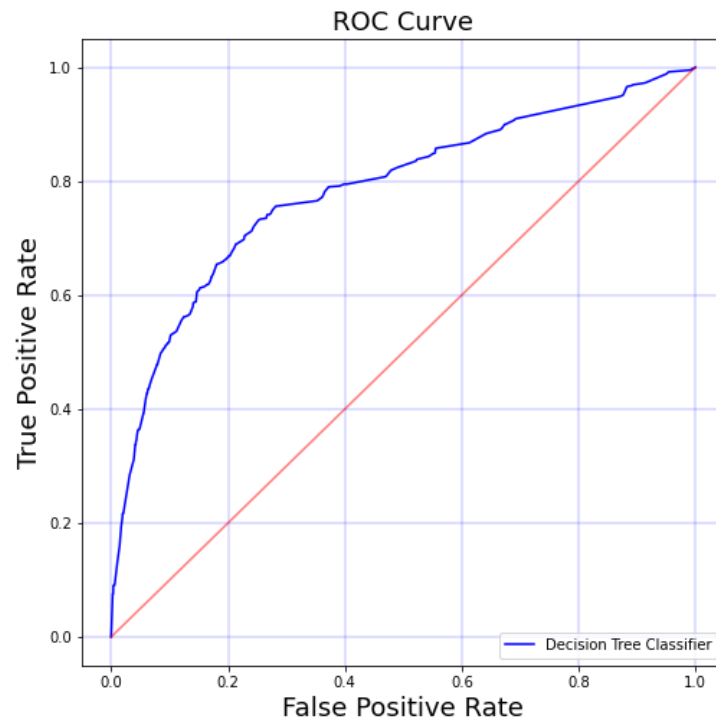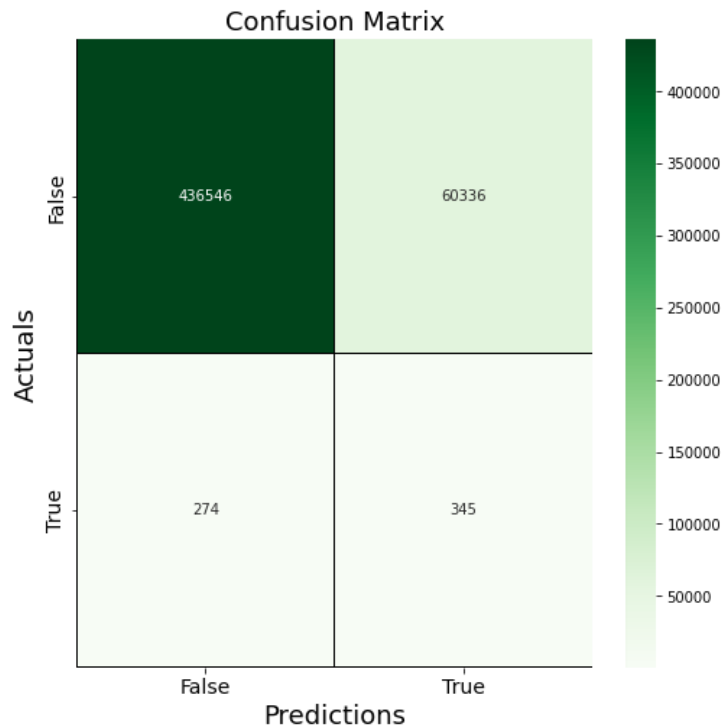True   False

Our Decision Tree Model

- Binary tree that controls the flow of

- Rank: A value that splits the data

- Gini: The quality of the split

- Samples: How many samples are left at this split

- Value: Shows how many were True and False
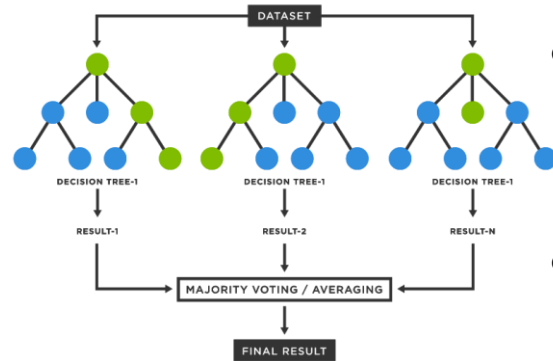
Example Decision Tree

Results from Well Fitted Tree

Confusion Matrix

	False	True
False	436546	60336
True	274	345

Actuals / Predictions

ROC Curve

Decision Tree Classifier

True Positive Rate vs False Positive Rate

Random Forests

What are Random Forests?
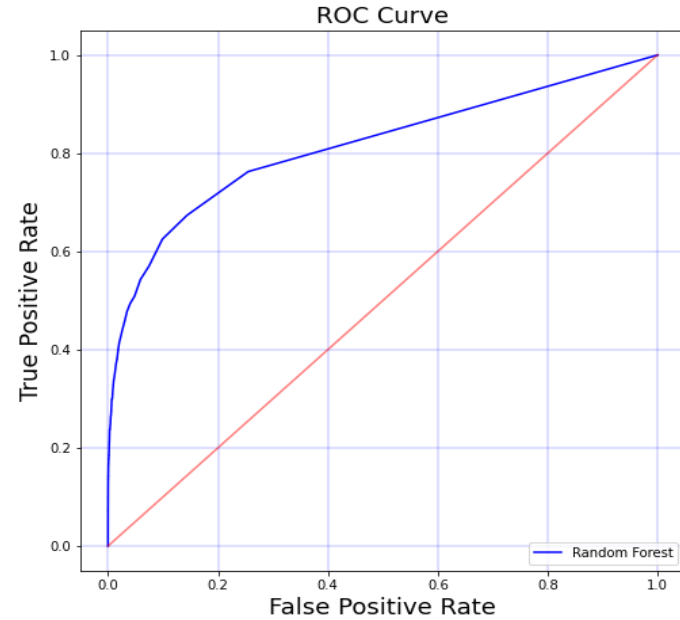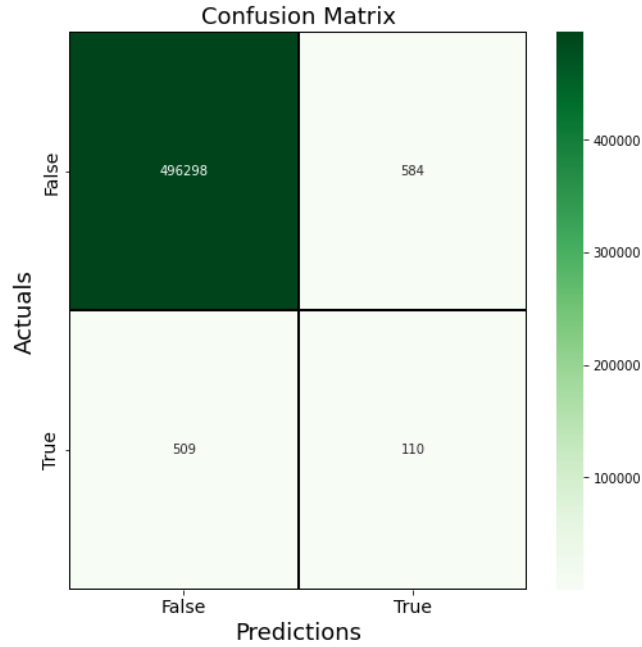

- Algorithm consists of a collection of multiple decision trees

- Used for classification and regression

How did we use this for our project?

- Compare Random Forest results with it's simpler counterpart Decision Trees.

Random Forest Results

Confusion Matrix

	False	True
False	496298	584
True	509	110

ROC Curve — Random Forest
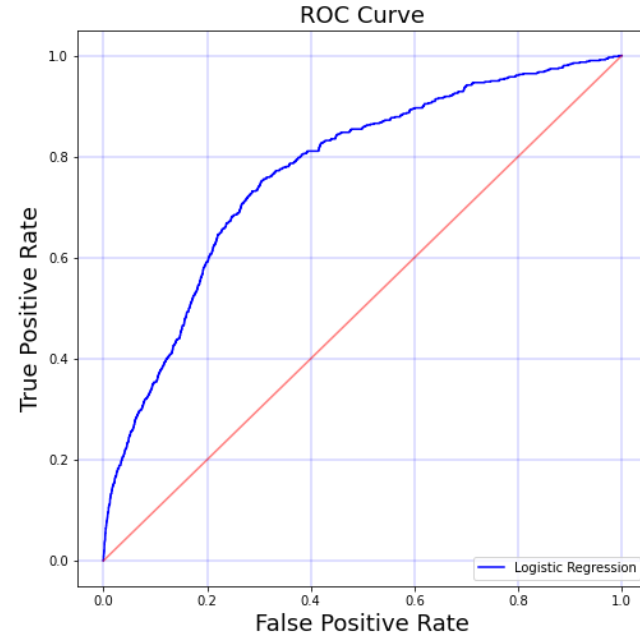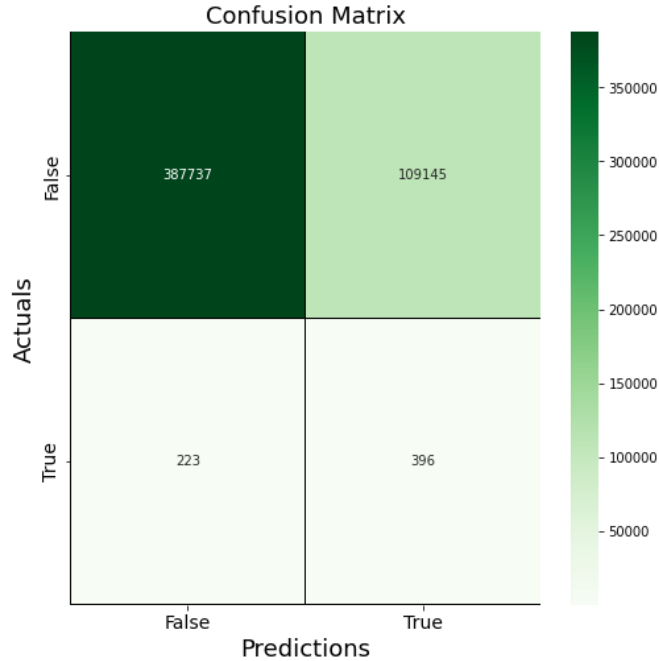
Logistic Regression

What is Logistic Regression?

- Shows relationships between two variables by estimating probabilities on a logistic regression curve

- Used for predicting the class of a categorical variable

How did we use this for our project?

- It is almost logical to use this method for our project since the class is if the transaction was fraud or not.

Logistic Regression Results

Results

- Random Forest performed better overall than the Decision Tree
- Logistic Regression and Decision Tree were able to classify more frauds, but the false positive rate was also much higher

Final Comments

1. Next steps?

2. What we would've done differently

3. Our experience