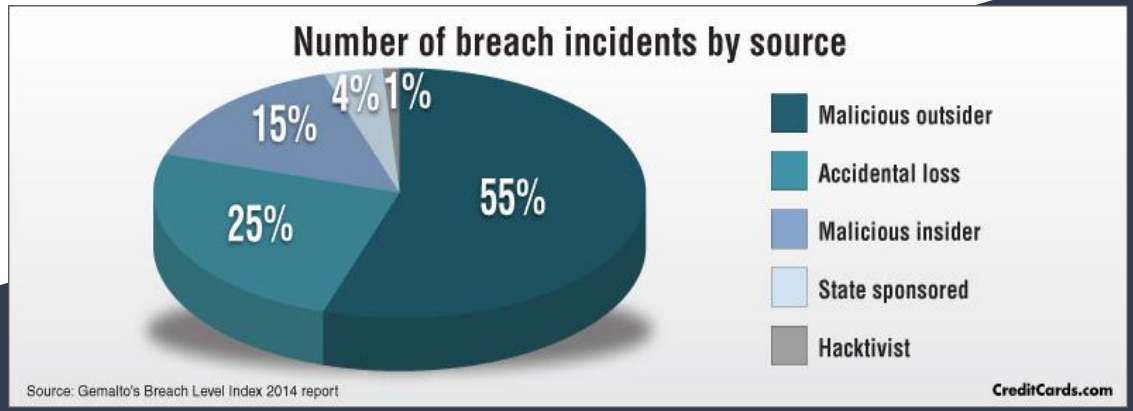
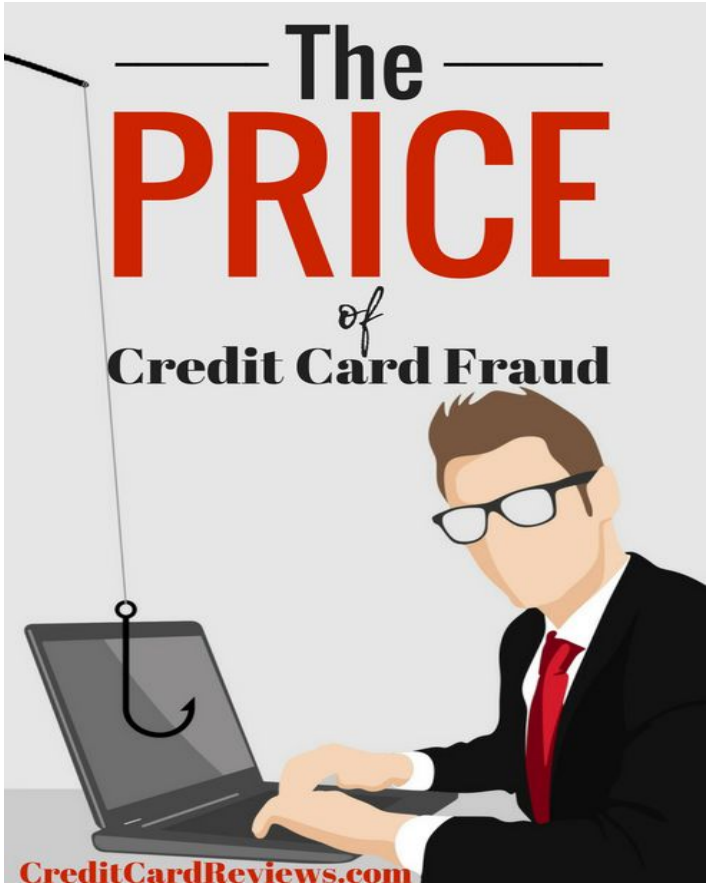
Welcome!
Credit Card Fraud detection



By: Nick Oseland - Data Scientist
Machine Learning

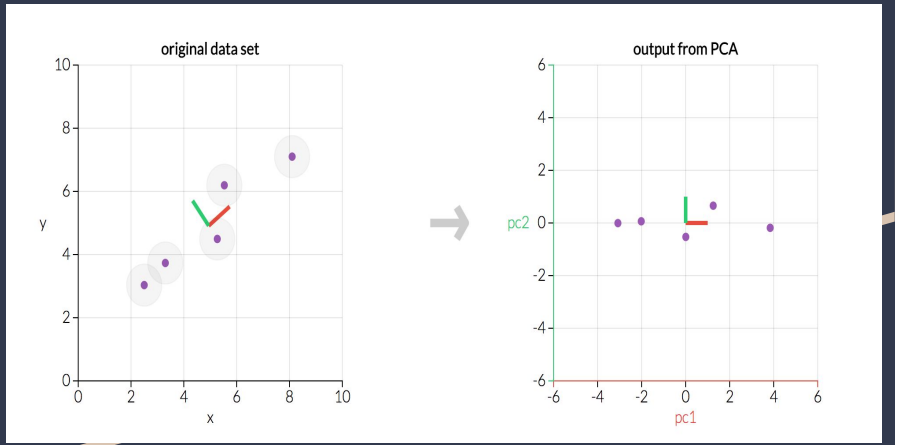
Problem



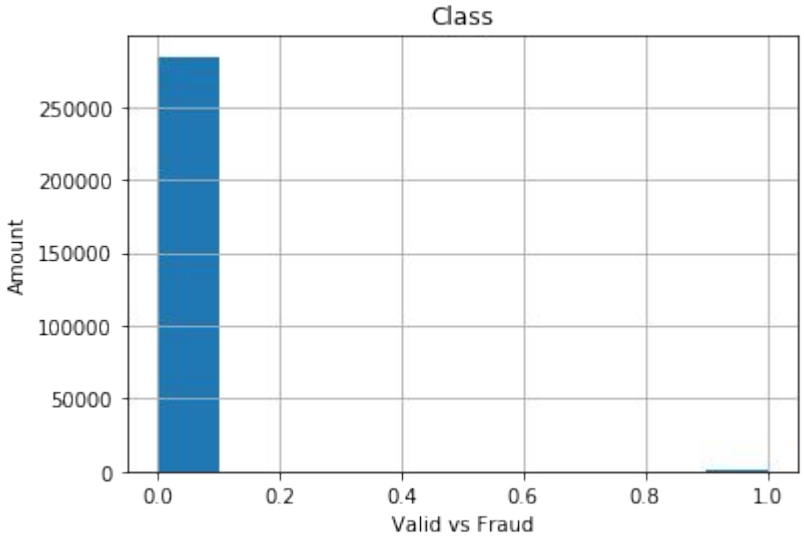
Viewing the Data

- 284,807 Transactions over a 2 day period
- Numerical Data
- 30 features
- Class = Fraudulent(1) or Valid(0)
- 492 frauds

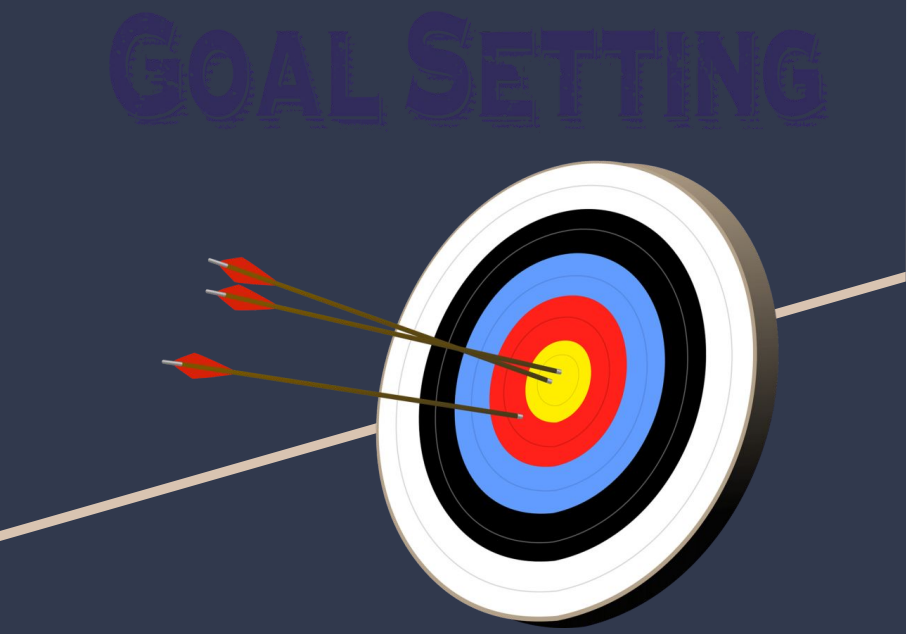
Viewing the Data cont.



- 2 labeled features
- “Time”
- “Amount”
- 28 unlabeled features
- Highly unbalanced



The Goal?



Create a model that can accurately predict fraud moving forward.

Modeling Decision Tree

Valid transactions = 100%(rounded)

Got 93,786 correct out of 93,838

Fraudulent transactions = 70%

Got 119 correct out of 149

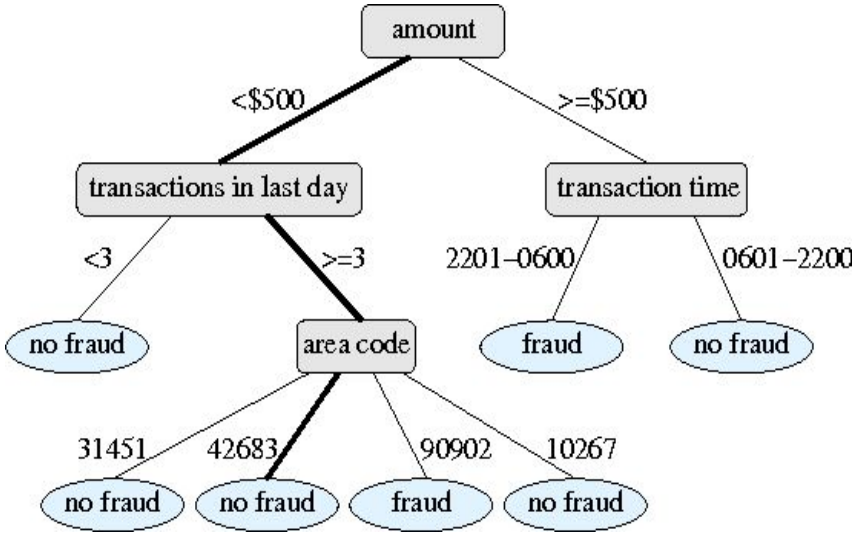
Confusion Matrix and ROC Curve

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

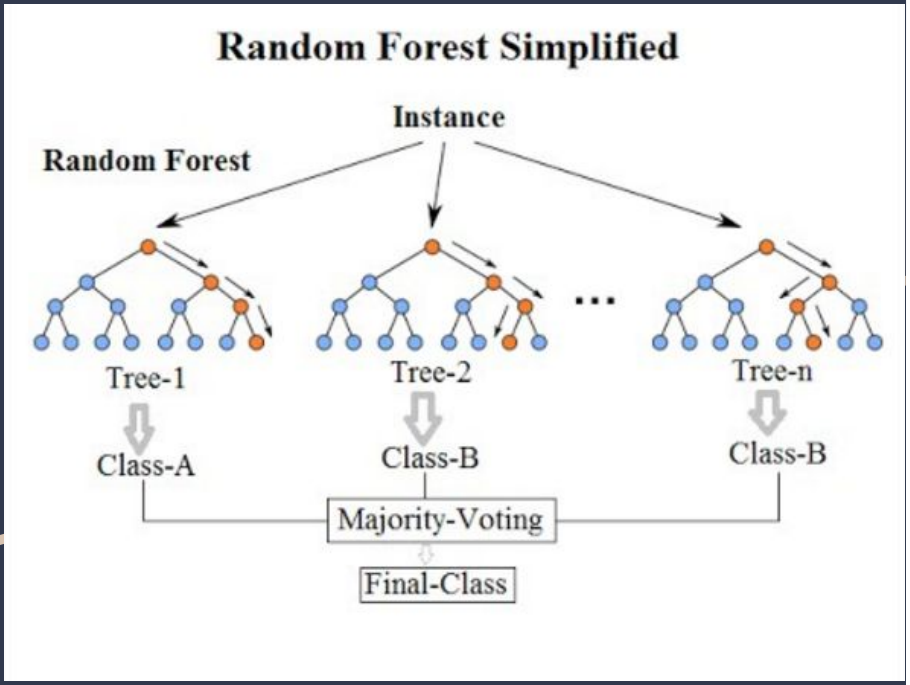
TN	True Negative
FP	False Positive
FN	False Negative
TP	True Positive

Model Performance

Accuracy	= (TN+TP)/(TN+FP+FN+TP)
Precision	= TP/(FP+TP)
Sensitivity	= TP/(TP+FN)
Specificity	= TN/(TN+FP)



Modeling
Random Forest



- Much more versatile
- Longer computational time

- Precision score for Valid transactions still 100% (rounded) but only missed 7 out of 93,838(missed 52 with Decision Tree)

- Precision score of 95% for fraudulent transactions(+25%), getting 121 correctly identified fraudulent points out of 149

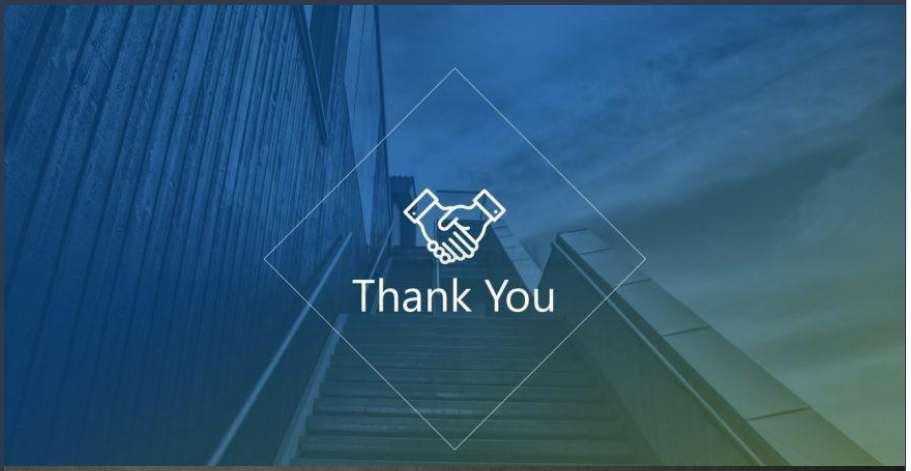
Conclusion

- Decision trees worked well

- Random Forest worked better

Future work:

- Gather more data
- Experiment with different classification algorithms.
- Discuss best metrics for model success


Thank You

ANY
QUESTIONS
?

YEAH !!!
WE FINISHED