IST 718: CREDIT CARD FRAUD DETECTION

ABHIJIT GOKHALE
SHUBHAM SHARMA

Syracuse University
School of Information Studies

School of
Information Studies

Agenda

Why credit card fraud detection?

Understanding the Data

Project Objective and Lifecycle

Tackling Class Imbalance

Classification Models

Model Evaluation

Why Credit Card Fraud Detection?

- Most frequently used mode of transaction
- Second most commonly reported Fraud (theft) in 2020
- According to FTC, there were over 390 thousands reports of credit card fraud in 2020 and 149 million dollars were lost only in the United States.
- In 2019, the U.S. accounted for 33.57 percent of all gross card fraud losses worldwide.

Project Objective

- Solve the classification problem of predicting fraudulent transactions.
- Recommending suggestive solutions to reduce credit card frauds.

Project Lifecycle

1 Data Preprocessing

Remove redundant columns, Perform Visualizations and correlation

2 Feature Engineering

Historical Trends

4 Model Evaluation

Understand performance metrics

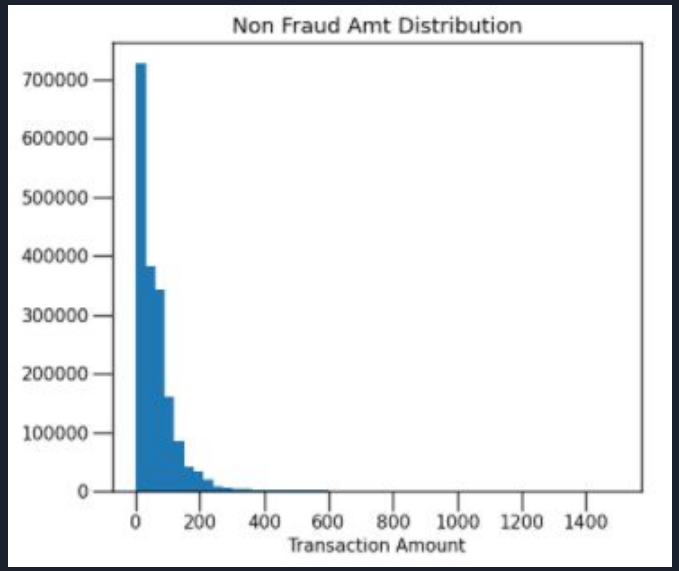
3 Model Creation

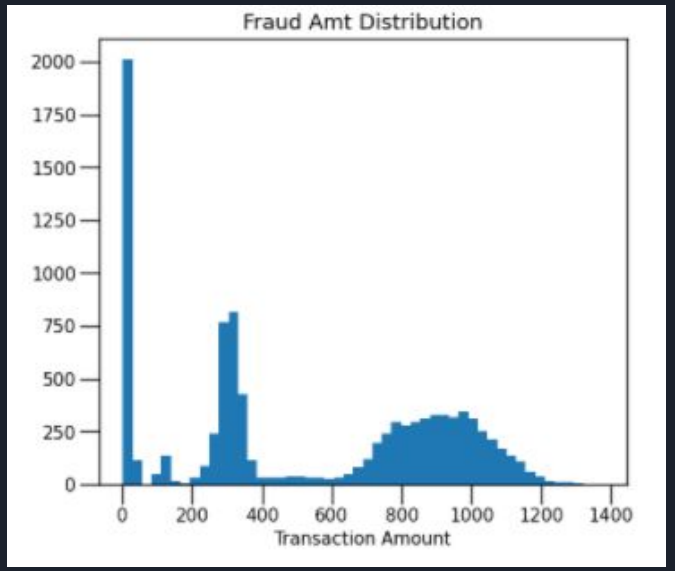
Classification Models

Understanding the Data (1)

- 23 features and 1.8 million transactions
- Target variable -> “is_fraud” contains “0: Legitimate Transaction” and 1: Fraudulent Transaction
- Transactions of 1000 credit cards from Jan 2019 to Dec 2020
- Geographical data with Merchants and Credit Card users details
- Transaction data related to amount and the time of transaction

Understanding the Data (2)





Mean of Non Fraud Transactions: 67.6

Mean of Fraud Transactions: 530.6

Tackling Class Imbalance

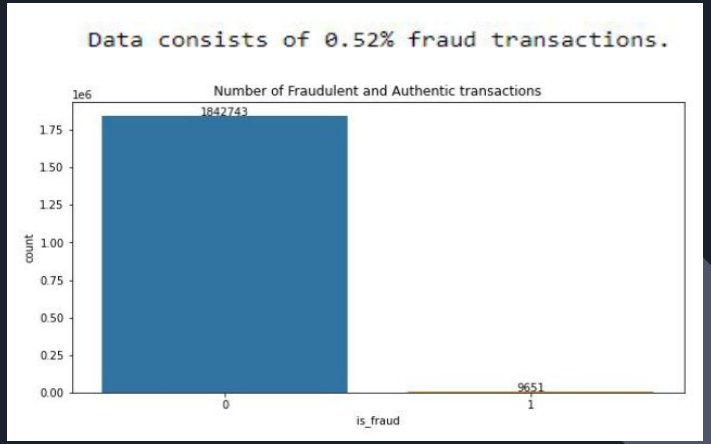
- We want to maximize recall while capping False Positive Rate
- A lot of True Negatives in the data which can lead to low FPR even when False Positives are high
- Need to find an optimum threshold probability for maximizing F1 Score

$$TPR = \frac{TP}{TP + FN}$$

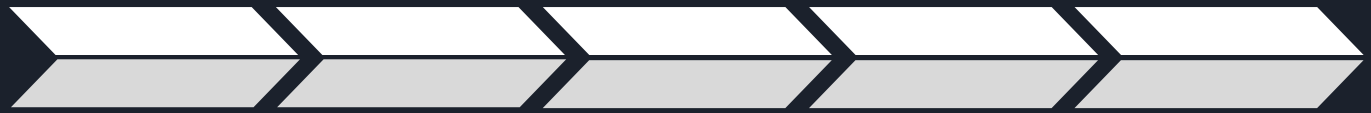
$$FPR = \frac{FP}{FP + TN}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

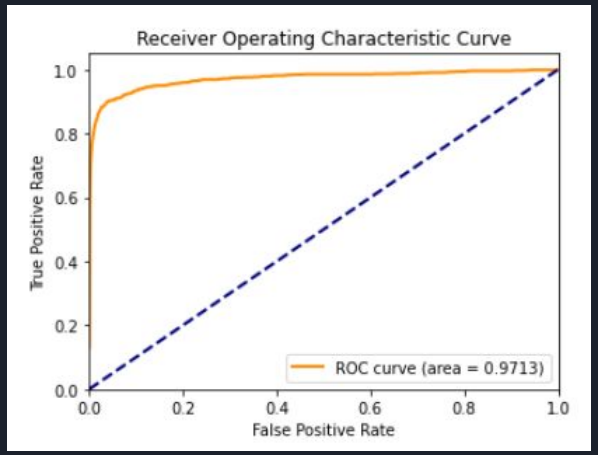


Classification Models

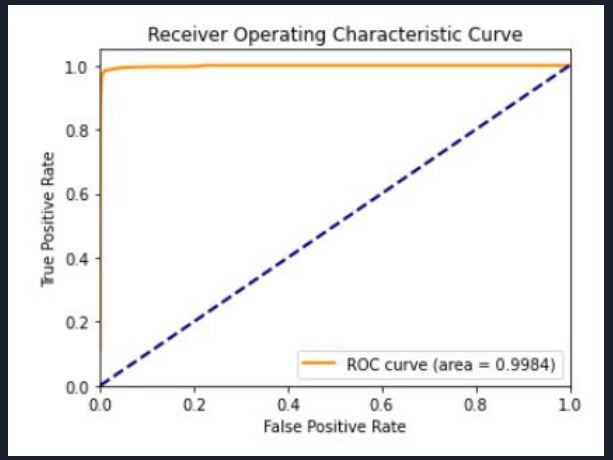


Logistic Regression	Logistic Regression with CV & Regularization	Random Forest	Random Forest with CV & Hyperparameter Tuning	Gradient Boosting
Default parameters	L1 and L2 penalty on the cost	Default Parameters	Tuning Impurity, No. of trees and max tree depth	Default parameters

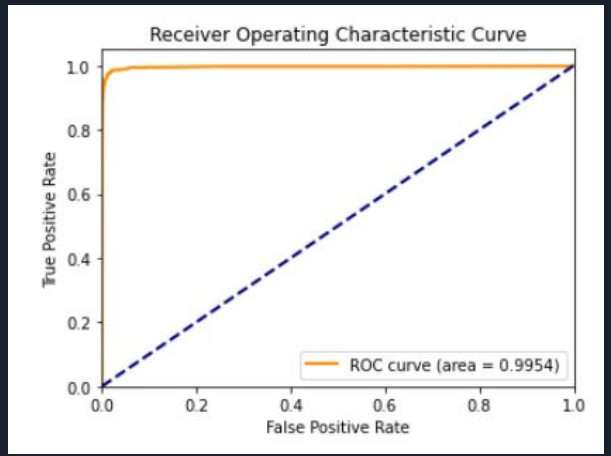
Area under ROC



Logistic Regression

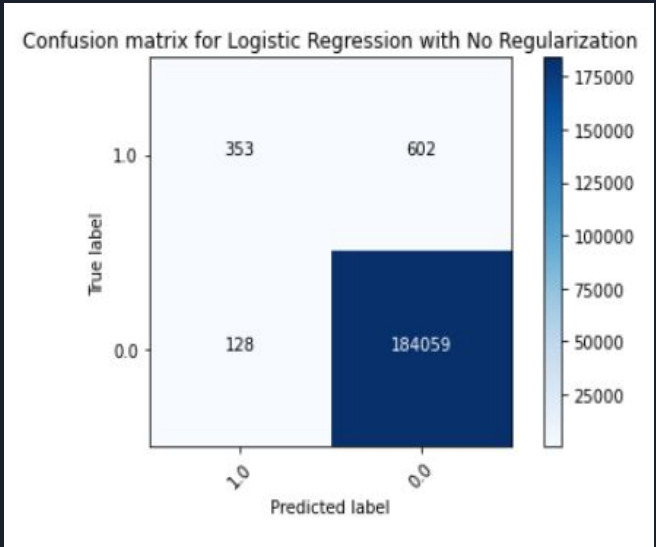


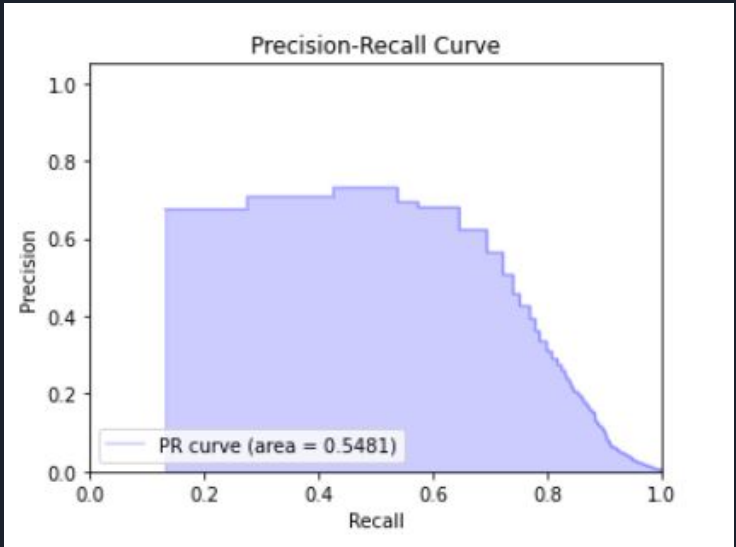
Random Forest



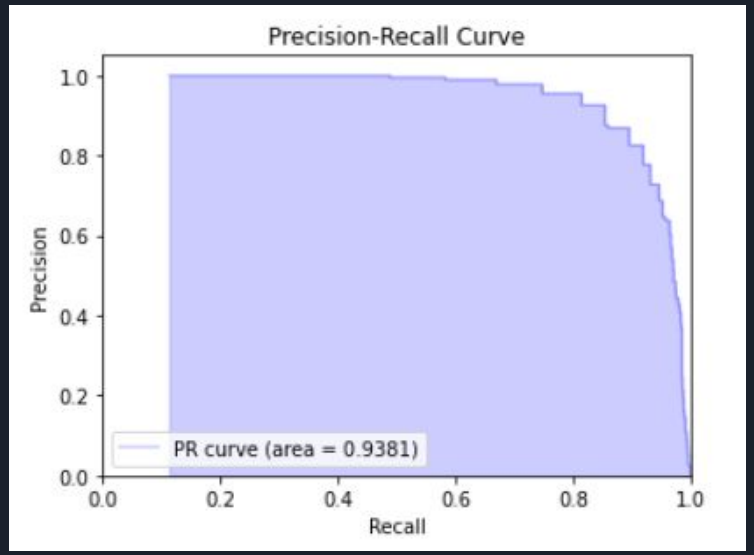
Gradient Boosting

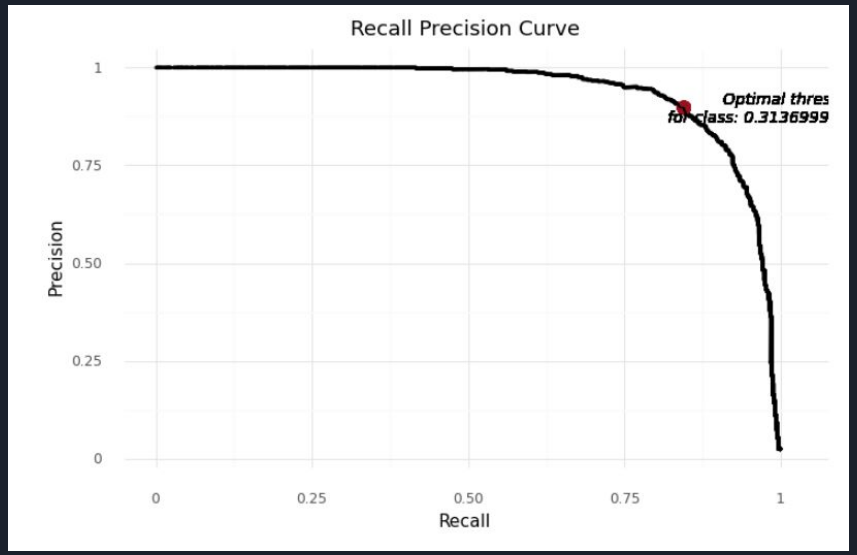
Area under Precision-Recall Curve (Logistic Regression)





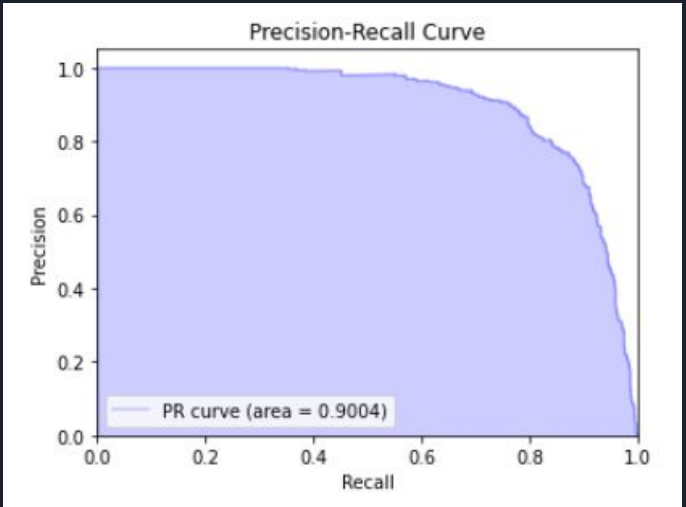
Area under Precision-Recall Curve (Random Forest)

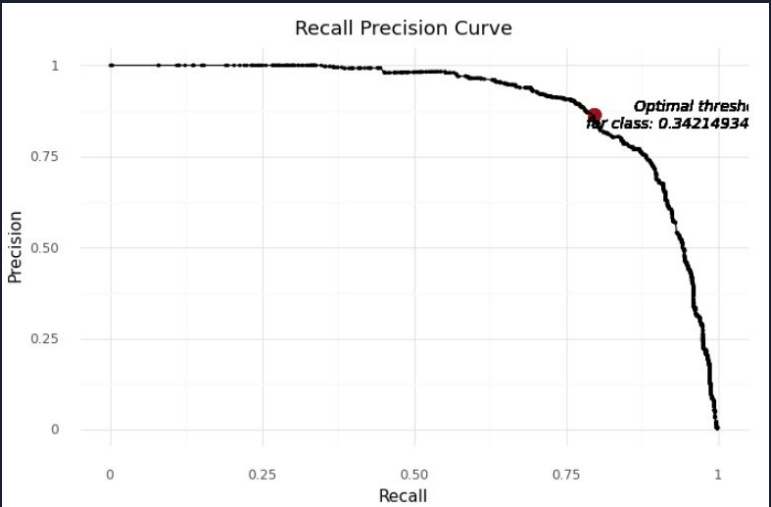





```
Best Threshold: 0.31369999051094055 with F-Score: 0.8704
Recall: 0.8452, Precision: 0.8971
```

Area under Precision-Recall Curve (Gradient Boosting)



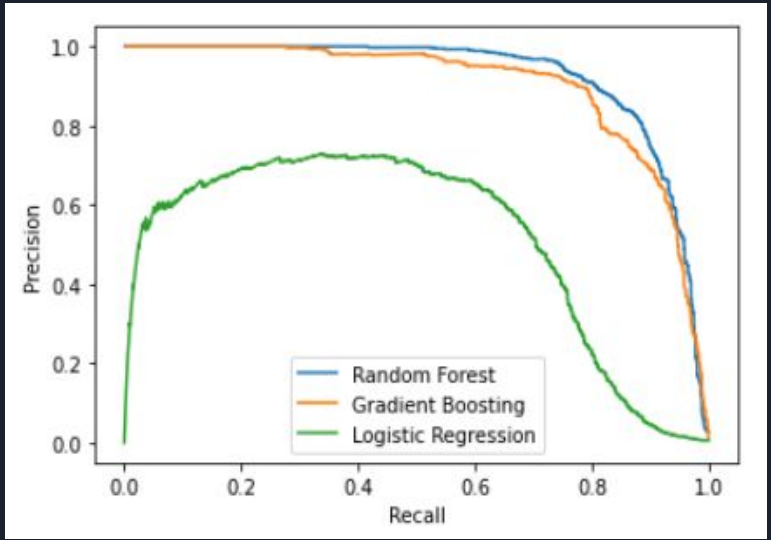


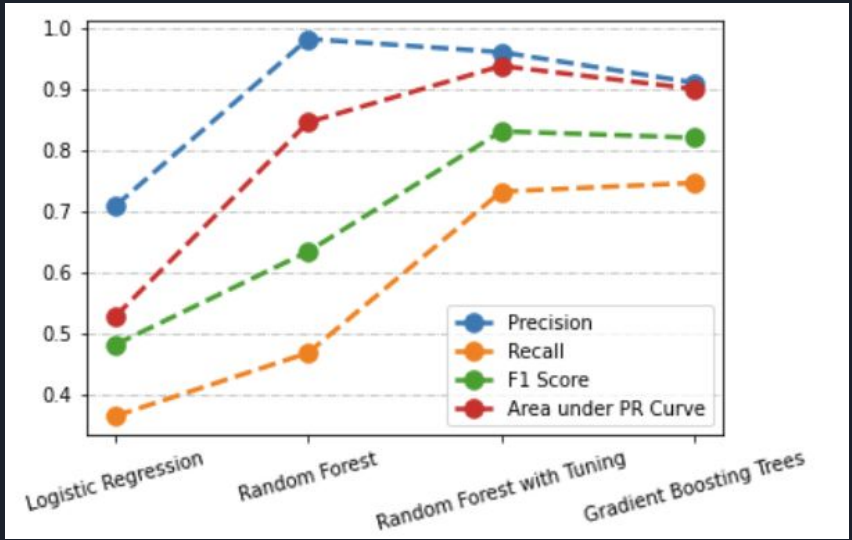

```
Best Threshold: 0.34209999442100525 with F-Score: 0.829
Recall: 0.7969, Precision: 0.8638
```

Consolidated Results

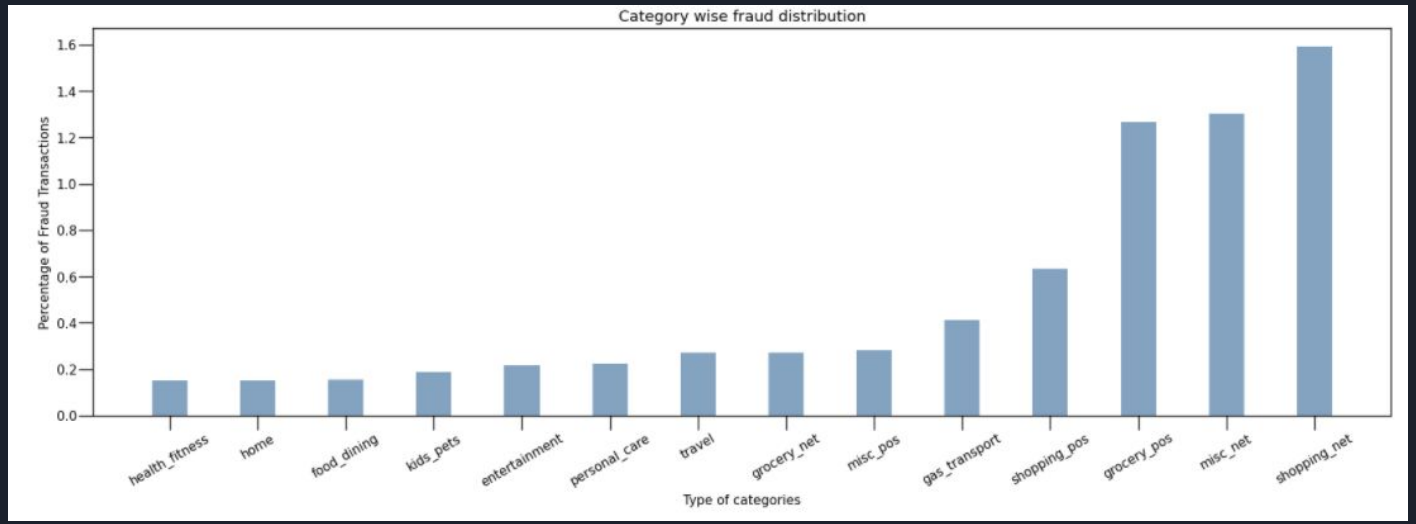
Performance Parameters	Logistic Regression	Random Forest	Random Forest with CV & Hyperparameter Tuning	Gradient Boosting
Precision	0.7082	0.9824	0.9606	0.9106
Recall	0.3643	0.4680	0.7322	0.7465
F1 Score	0.4812	0.6340	0.8310	0.8204
Area_ROC	0.9589	0.9949	0.9983	0.9953
Area_PR	0.5481	0.8457	0.9380	0.9003

Model Comparison





Recommendations: Reduce Credit Card Frauds (1)



day	is_fraud	count
Saturday	1	1227
Sunday	1	1216
Monday	1	1182
Friday	1	1079
Thursday	1	1008
Tuesday	1	935
Wednesday	1	859

hour	is_fraud	count
22	1	1931
23	1	1904
1	1	658
0	1	635
2	1	624
3	1	610
14	1	86
19	1	81

Recommendations: Reduce Credit Card Frauds(2)

- Don’t use unsecure websites and beware of phishing scams.
- Be on the lookout for skimmers and don’t post sensitive information on social media.
- Don’t save your credit card information online and never use debit cards for online purchases.
- Get a chip card with PIN capacity so one can make a habit to shop in stores that have chip readers.
- Don’t trust public Wi-Fi for financial transactions and set up a fraud alert or credit freeze if your card is lost or stolen.
- Audit your online financial accounts and credit card activity online weekly

Thank you!