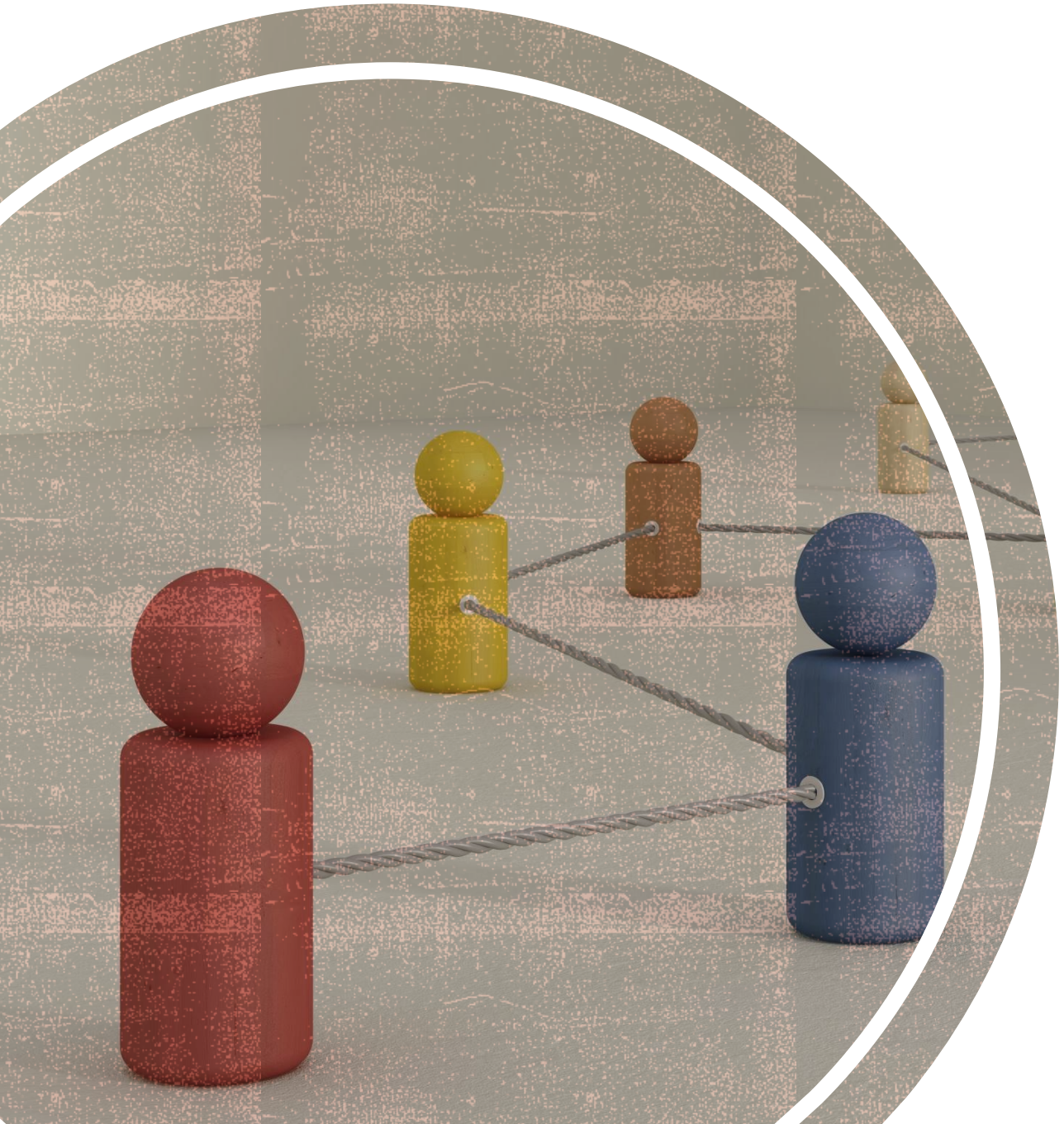
PREDICTING CREDIT CARD FRAUD

Shashi Bhushan
Spring 2024
DSC 680



AGENDA

Introduction

Data Selection

Modeling and Methods Used

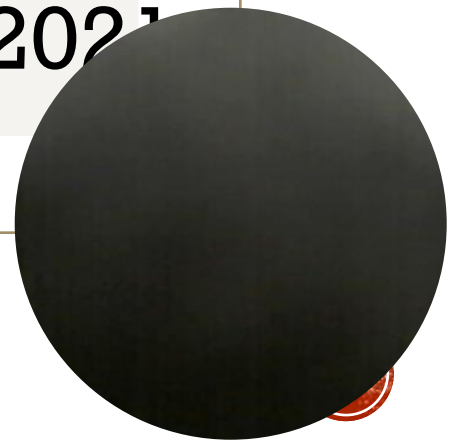
Interpretation of Analysis / Model Results

Conclusion

INTRODUCTION

Federal Trade Commission reports that credit card frauds are a big issue .

390,000 credit card identity theft cases being reported to the agency in 202 alone

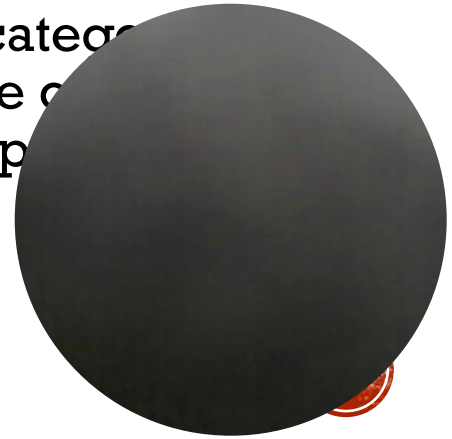


DATA SELECTION

Kaggle

DATA SELECTION

- Resoure Link: Credit Card Fraud Prediction (kaggle.com)
- Has 22 variables
- Fraud noted as 1 and 0
- Collected in 2020
- Has categorical and numerical data
- 555,719 rows
- Features -transaction date, credit card number, merchant name, merchant categ amount, names, gender, street address, city, zip code, location, job, and date o of the card holder, transaction identification number, transaction time stamp merchant location (latitude, longitude)



MODELING AND METHODS USED

VISUALIZATIONS

- Bar chart showing count of males and females who were subjected to fraud .
- Bar chart showing positive counts of fraud by state.
- Bar chart showing positive counts of fraud by merchant category.
- Age distribution (histogram for positive fraud cases) .

DATA PREPARATION

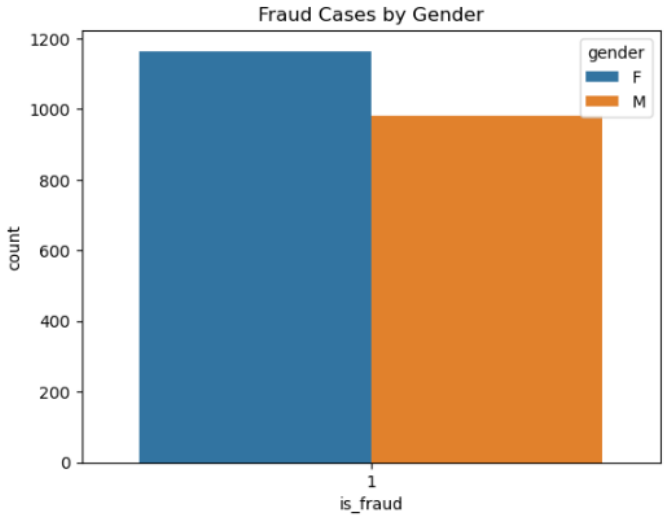
- Categorical variables were encoded numerically
- Features observed to be not relevant removed

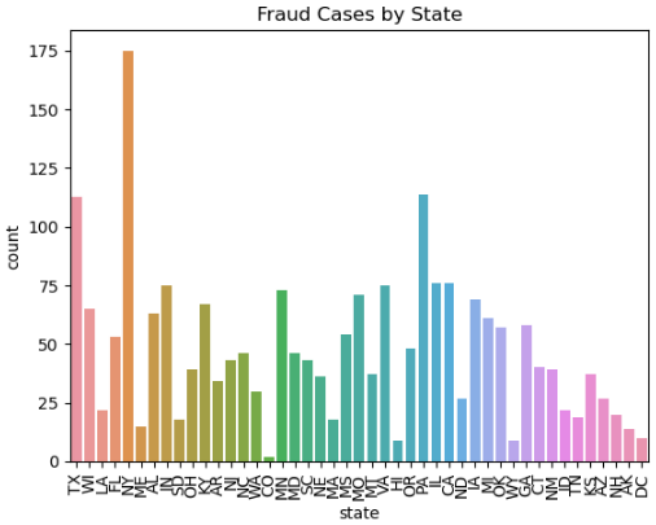
MODELING

- Target outcome is 1 or 0 / Binary
- Created two models:
  - Logistic Regression
  - Nearest Neighbor Algorithm

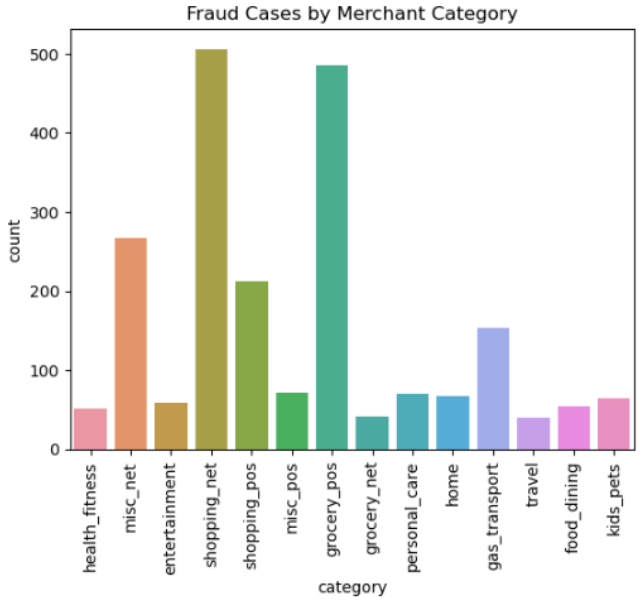
INTERPRETATION OF ANALYSIS / MODEL RESULTS

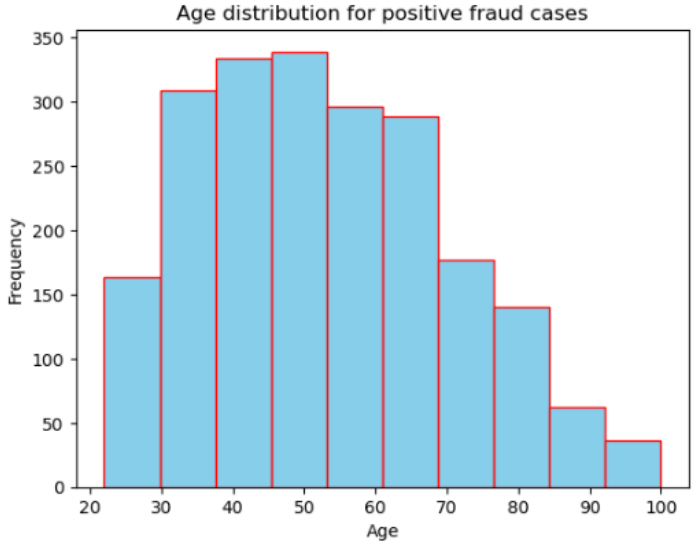
VISUALIZATIONS





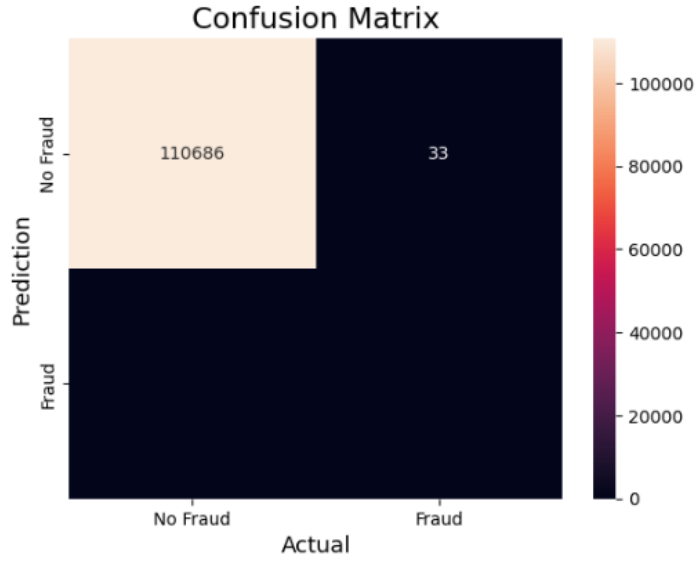
VISUALIZATIONS CONTINUED





MODEL RESULT INTERPRETATION

- Logistic Regression

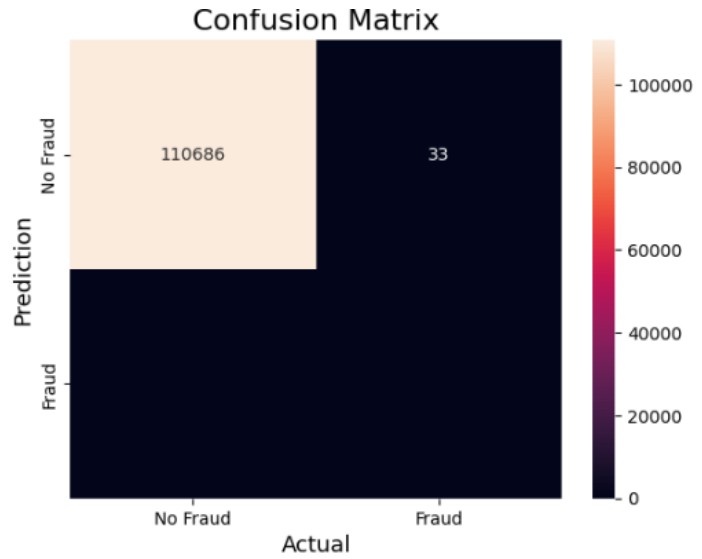



```
print(classification_report(y_test, y_test_pred, labels=[0,1]))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	110719
1	0.00	0.00	0.00	425
accuracy			1.00	111144
macro avg	0.50	0.50	0.50	111144
weighted avg	0.99	1.00	0.99	111144

MODEL RESULT INTERPRETATION - CONTINUED

- Nearest Neighbor




```
print(classification_report(y_test, y_pred, labels=[0,1]))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	110719
1	0.59	0.45	0.51	425
accuracy			1.00	111144
macro avg	0.79	0.72	0.75	111144
weighted avg	1.00	1.00	1.00	111144

CONCLUSION

- Recommend KNN model- Higher accuracy together with high precision scores for predicting "credit card fraud" between the two models created.
- Only slight overfitting observed.

THANK YOU

Shashi Bhushan