

Vehicle Insurance Fraud Detection

Morgan Didjurgis

03 Feb 2022



Outline

- Business Problem
- Data
- Model
- Results
- Conclusions



Business Problem

1

Provide automatic
fraud detection to flag
claims for further
investigation

2

More important to
find fraud than 'guess
correctly' on each
claim

3

Balance detection of
actual fraud with not
flagging too many
non-fraud claims for
investigation

Data

- List of over 15,000 vehicle insurance claims
- 30 features, 19 used in model
- Labeled Fraud / Not Fraud
- 6% of claims were Fraud
- Kaggle Dataset, originated with Oracle
- Claims occurred in 1990s

Features included in model:

| | |
|-------------------------|-----------------------|
| Fault | Day of week |
| Base Policy | Marital Status |
| Vehicle Category | Month |
| Policy Type | Driver Rating |
| Address Change on Claim | Vehicle Price |
| Past # Claims | Age of Policy Holder |
| Deductible | Age, |
| Make | Week of Month |
| Month claimed | Week of Month claimed |
| Number of Supplements | |



Model

- Gradient Boosting Algorithm for Classification
 - Using CatBoost Library
- Model Performance:
 - Can find 90% of fraud while flagging only 36% of non-fraud claims
 - Capable of flagging more fraud, but false positives would increase
 - Most important features are:
 - Fault (claimant or third party)
 - Base Policy
 - Vehicle Category
 - Policy Type
 - Address changes on claim
- Limitations of the model:
 - In order to detect fraud, which occurs less frequently, overall accuracy is sacrificed
 - Positive results should be seen as ‘potential fraud’

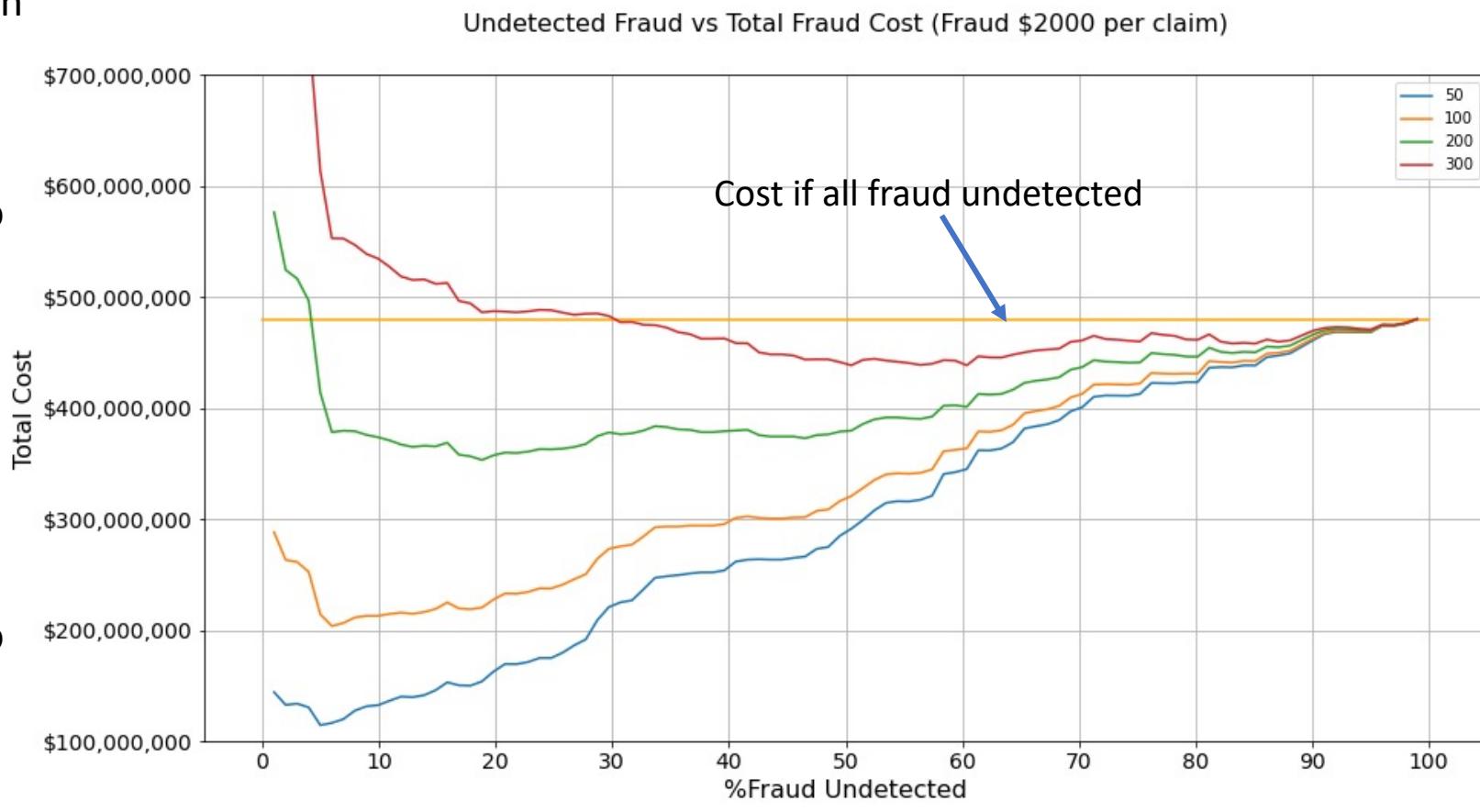
Results: Example cost analysis

Assume that a company processes 4 million claims each year and 6% are fraudulent (based on data set)

Evaluate acceptable level of fraud to let go undetected based on overall cost. This includes:

- Cost of investigating flagged claims (actual fraud and non-fraud)
- Cost of undetected fraud

If we assume that a fraudulent claim costs the company \$2000 on average, cost of no fraud detection is \$480 million



Plots show the total cost for two different 'costs to investigate fraud'

Conclusions

- Model flags potential fraud for investigation
- Aims to reduce costs by identifying fraudulent claims. Must balance the cost of investigation with the cost of fraud not found.

Next Steps

- Cost analysis of fraud and investigation process
- Potential to build model to detect certain types of fraud, especially ones with higher average cost

