Vehicle Insurance Fraud Analysis

Christine Gendron
Mitchel Diaz
Steven Dookhantie

Vehicle Insurance Fraud Analysis

Christine Gendron
Mitchel Diaz
Steven Dookhantie

Our Purpose:

Vehicle Insurance Fraud is an expensive, increasingly common, and avoidable problem for insurers.

*** find a stat **

This is not exclusive to auto insurers. Sophisticated fraud detection processes will be essential to the entire Financial Services Industry moving forward.

That's why this project seeks to use machine learning techniques to understand and predict fraud.

The Questions:

Can we accurately predict whether an insurance claim is fraudulent?

Which features are correlated with higher likelihood of fraud?

The Data:

Our dataset was generated by **Angoss KnowledgeSEEKER**, a provider of systems for predictive analytics, and found on **Kaggle**.

This set features over 15,000 claim samples with 33 columns, including information on:

- *Demographics of claimant*
- *Details on the vehicle in question*
- *Information about the claimants' policies.*
- *Whether fraud was detected*

Phase 1: Exploration

Slides will include:

- Which columns did we drop
- Which columns did we transform
- How did we make these determinations

Try to include images of exploratory plots/tables

Phase 2: Analysis

Slides will include:

- Rundown of tech, languages, tools, and algorithms used
- High-level overview of analysis/ml process (real details will be in readme/repo)
- Database stuff? (or can this just be reference in the readme/repo?)
- Visualizations and demo of interactive dashboard illustrating results
- Recommendations for future analysis
- What we would have done differently in retrospect

Our Purpose:

Vehicle Insurance Fraud is an expensive, increasingly common, and avoidable problem for insurers.

*** find a stat **

This is not exclusive to auto insurers. Sophisticated fraud detection processes will be essential to the entire Financial Services Industry moving forward.

That's why this project seeks to use machine learning techniques to understand and predict fraud.

The Questions:

Can we accurately predict whether an insurance claim is fraudulent?

Which features are correlated with higher likelihood of fraud?

The Data:

Our dataset was generated by **Angoss KnowledgeSEEKER**, a provider of systems for predictive analytics, and found on **Kaggle**.

This set features over 15,000 claim samples with 33 columns, including information on:

- *Demographics of claimant*
- *Details on the vehicle in question*
- *Information about the claimants' policies.*
- *Whether fraud was detected*

Phase 1: Exploration

Slides will include:

- Which columns did we drop
- Which columns did we transform
- How did we make these determinations

Try to include images of exploratory plots/tables

Phase 2: Analysis

Slides will include:

- Rundown of tech, languages, tools, and algorithms used
- High-level overview of analysis/ml process (real details will be in readme/repo)
- Database stuff? (or can this just be reference in the readme/repo?)
- Visualizations and demo of interactive dashboard illustrating results
- Recommendations for future analysis
- What we would have done differently in retrospect