AUTO INSURANCE FRAUD DETECTION THROUGH MACHINE LEARNING

84% CONFIDENCE IN IDENTIFYING FRAUDULENT INSURANCE CLAIMS

Lost premiums not only impact companies but also affect customers. Up to 14% of what you pay for car insurance premiums goes toward covering your insurance company's lost premiums.

AUTHOR

Caroline Floss Bresolin
Data Analytics for Business
CCT College

ACKNOWLEDGEMENTS

I am thankful to the lecturers of CCT for their invaluable assistance throughout the course, as well as to my friends and family, whose unwavering support made this study possible.

01. INTRODUCTION

Auto insurance companies lose billions of dollars per year due to insurance fraud. Applicants, policyholders, mechanics and accident victims all participate in car insurance fraud. Car insurance companies lose $29 billion per year because of this, according to a 2017 study by Verisk. American families pay an additional $400 to $700 per year in insurance premiums to help cover the cost of insurance fraud, according to the FBI.

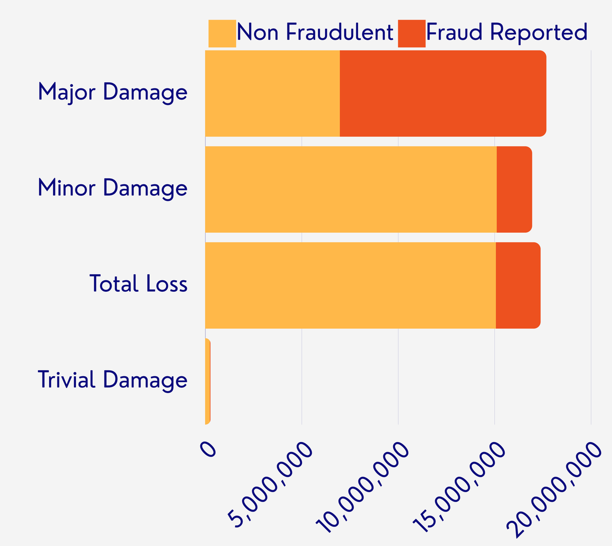
02. OBJECTIVE

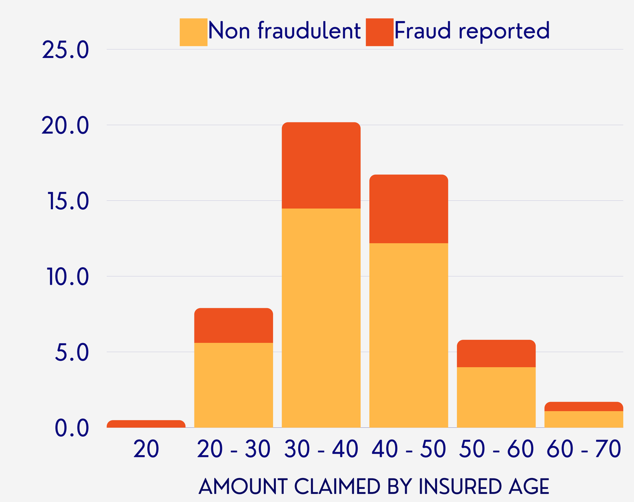
The insurance industry faces significant challenges in detecting and preventing fraudulent claims, which can lead to substantial financial losses. In response to this challenge, my study aims to leverage machine learning techniques to develop predictive models capable of identifying potential instances of insurance fraud based on a dataset that include details from insured individuals as well as information regarding claims within three states of the US: OH, IL and IN.

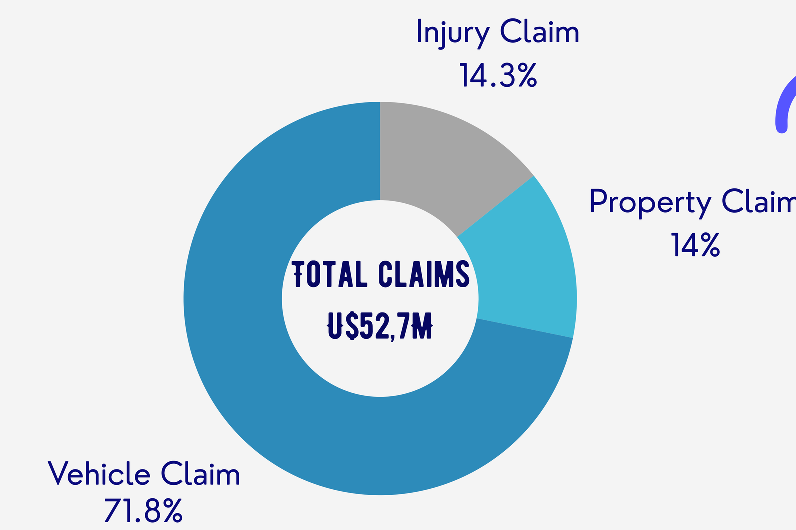
MOST COMMON TYPES OF AUTO INSURANCE FRAUD

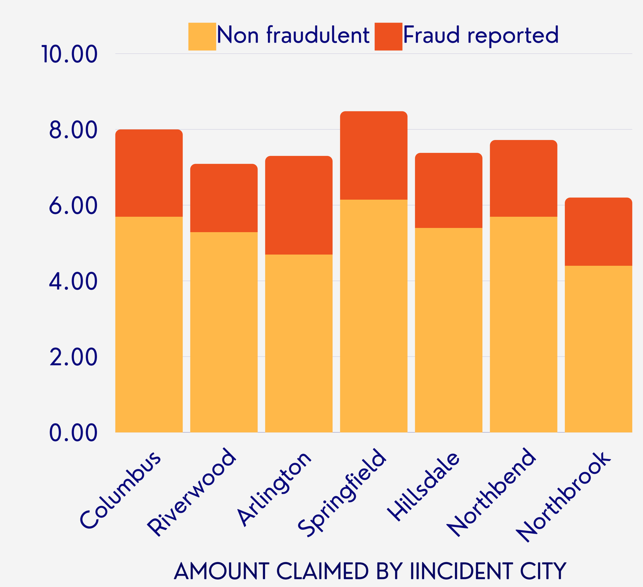
- **Lying on your insurance application to get a lower rate**
- **Faking or exaggerating an injury**
- **Filing multiple claims for the same accident**
- **Filing a claim for an accident that never happened**
- **Causing an accident on purpose to file a claim with another driver's insurance**

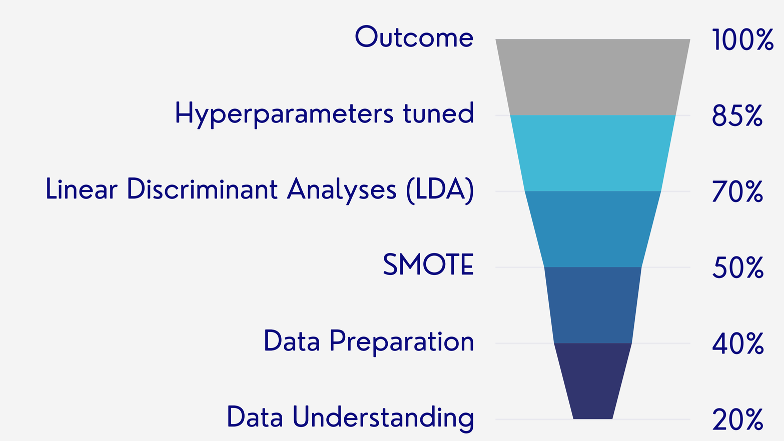
05. ANALYSIS



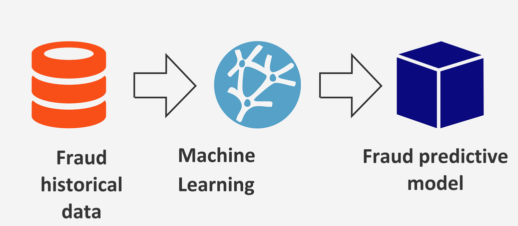


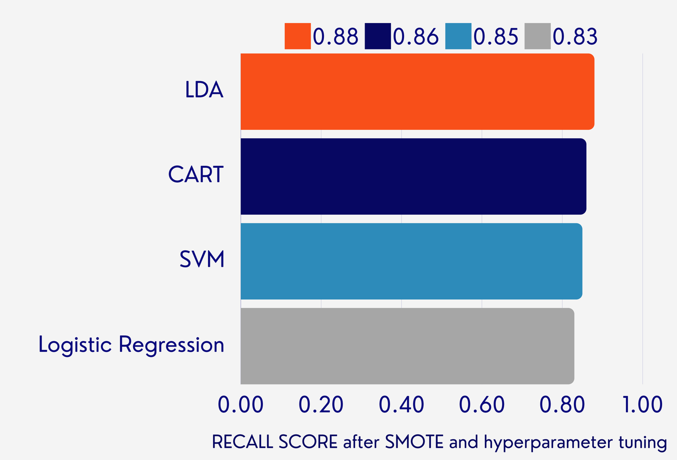


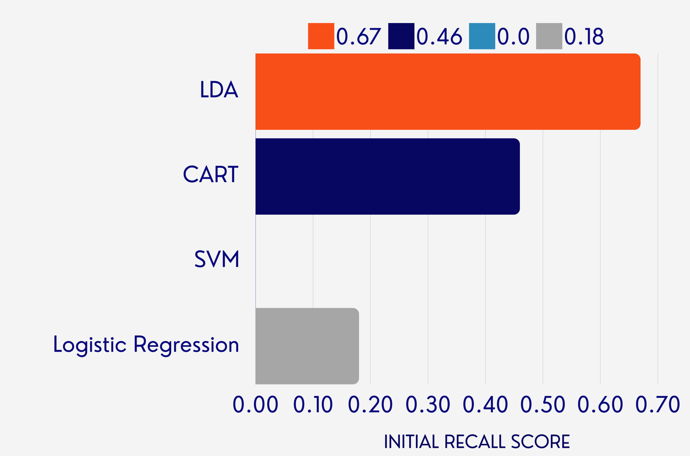


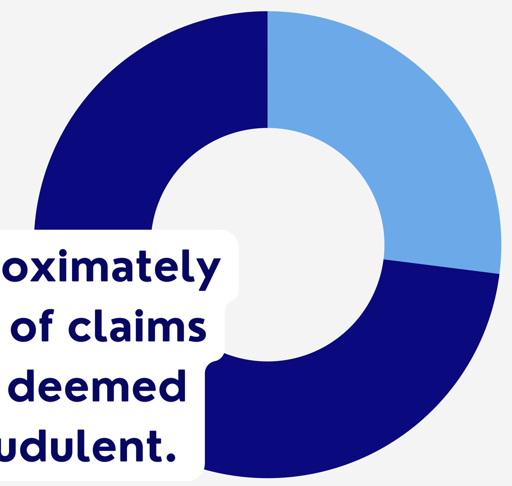


The evolution of metrics in the model is significant when tackling challenges like imbalanced targets and hyperparameter tuning, highlighting the importance of comprehending both the dataset and the employed model. The following graphs outline the disparity in scores









Approximately 30% of claims are deemed fraudulent.

MACHINE LEARNING MODEL CAN DETECT FRAUDULENT CLAIMS WITH UP TO 85% OF ACCURACY

03. METHODOLOGY

This project was conducted in Python, with a dataset sourced by Kaggle, adhering to the CRISP-DM methodology

- **Data Preparation;**
- **Machine Learning Modeling: Models were trained and tuned;**
- **Performance Evaluation;**
- **Cross-Validation;**
- **Class Imbalance Handling;**
- **Model Interpretability with Lime and SHAP**

Research was conducted to gain a deeper understanding of the subject.

04. RESULTS/FINDINGS

- Interpretability techniques like SHAP and Lime reveal the importance of features on model outcomes and indicate that 'incident severity' exerts a substantial influence.
- The city where the incident occurred may influence the importance of certain features.
- Linear Discriminant Analysis (LDA) emerges as the optimal method for addressing insurance fraud, boasting an accuracy rate of 85% and a recall rate of 88%

06. CONCLUSION

Overall, this study contributes to enhancing fraud detection capabilities in the insurance industry by employing accurate and interpretable predictive models. The interpretability tools offered insights into model predictions, showcasing influential features and their significance in fraud detection.

With a well-trained model properly implemented, the detection of fraudulent claims has the potential to save around 12M and detect up to 85% of fraudulent claims based on this dataset.

REFERENCES

valuepenguin. (2024). Insurance Fraud Statistics. [online] Available at: https://www.valuepenguin.com/auto-home-insurance-fraud.
www.forbes.com. (n.d.). Insurance Fraud Statistics 2024 – Forbes Advisor. [online] Available at: https://www.forbes.com/advisor/insurance/fraud-statistics/.
InsuranceFraud.org. (n.d.). Fraud Stats. [online] Available at: https://insurancefraud.org/fraud-stats/.