

Use of Analytics in the field of Fraud Claims

Principal applications:

- Identification of new schemes of fraud claims and uncover hidden correlations.
- Earlier fraud detection in the insurance lifecycle.
- The system reveals connections between a great number of diverse factors, which can be imperceptible for a person.
- Huge amounts of data are processed in a relatively short period of time
- Fraudulent claims are detected with greater precision.

Purpose of Study

Unum Group is a Tennessee based Fortune 500 insurance company and pays over \$800m annually on Short Term Disability alone and facing fraudulent claim problems. Insurance fraud detection is a challenging problem, given the variety of fraud patterns and a relatively small ratio of known frauds in typical samples. Currently, most of the fraudulent activities are detected by looking at on-surface and evident signals which is time consuming and unable to find hidden and implicit correlations in data. With the help of analytics compared to rule-based, data processing will be faster and less manual work. For example, smart algorithms fit well with behavior analytics for helping reduce the number of verification steps.

Introduction

Globally it's estimated that 1 in 10 insurance claims is fraudulent. In the past, fraud detection was committed to claims agents who had to rely on few facts and a large amount of intuition. The detection of fraud claim is a very challenging problem due to different types of fraud patterns and very less known knowledge in the field of known frauds in typical sample. Moreover, the incidence of insurance fraud keeps growing every year and causing billions of money in loss for insurance companies. With the help of data mining and machine learning, few technique has introduced tools to make fraud review and detection possible as it allows improving predictive accuracy, allowing loss control units to gain higher reach with low false positive rates.

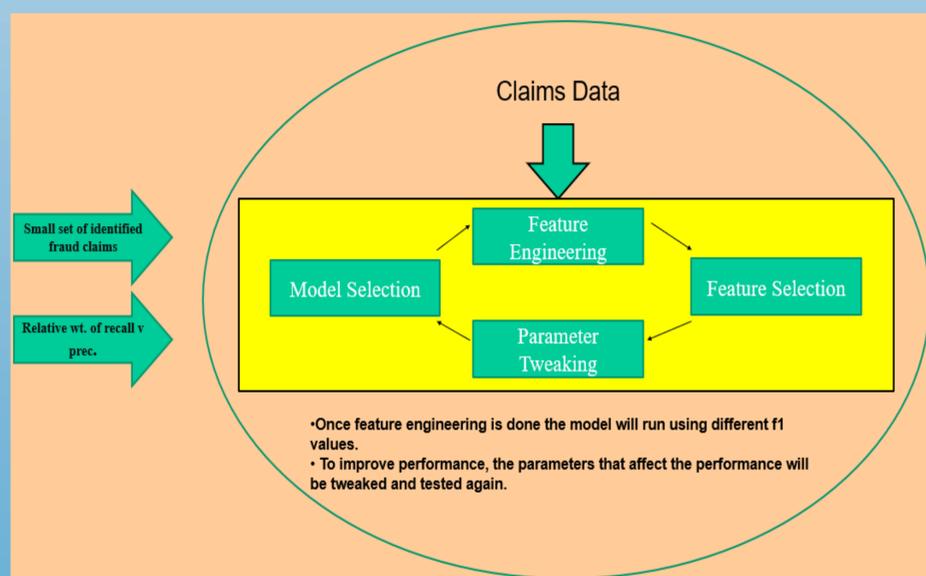


Figure 1: Model Architecture

References

1. AltexSoft. (2019). *Fraud Detection: How Machine Learning Systems Help Reveal Scams in Fintech, Healthcare, and eCommerce*. [online] Available at: <https://www.altexsoft.com/whitepapers/fraud-detection-how-machine-learning-systems-help-reveal-scams-in-fintech-healthcare-and-ecommerce/>.
2. Hargreaves, Carol & Singhanian, Vidyut. (2016). Analytics for Insurance Fraud Detection: An Empirical Study. *American Journal of Mobile Systems, Applications and Services*.
3. Maitra, S. (2019). *Data mining & ML to detect fraudulent Insurance claims — an analysis with complete Python code*. [online] Medium. Available at: <https://medium.com/fraud-investigation-using-machine-learning/fraud-detection-insurance-29d00b521525>.
4. www.wipro.com. (n.d.). *Predictive Analytics For Insurance Fraud Detection - Wipro*. [online] Available at: <https://www.wipro.com/analytics/comparative-analysis-of-machine-learning-techniques-for-detectin/>.

Arnab Roy
C00246500@itcarlow.ie
Supervisor: Jason Barron
M.Sc. in Data Science
Institute of Technology Carlow, Carlow, Ireland

Theory:

The research questions that this project conveys are as follows:

- To identify and investigate fraudulent insurance claims based on supervised and unsupervised machine learning algorithms?
- To investigate the normal/fraud behavior features?
- To build a model that produces fraud propensity and saves time spent on manual work.

The Dataset

- The data is provided by UNUM.
- This dataset contains 1,077,061 Short Term Disability claims with more than 50 features.

Technologies Used



The main motive of this research is to eliminate the heuristic approach around fraud indicators because of the inability to understand context-specific relationships between parameters manually. The bulk of the research is understanding and gathering the data because insurance data for every company is different from the other.

Discussion

This paper shows a devised implementation of data science in the field of detecting fraud in real-time.

- At first step large volume of previously processed claims is analyzed. By structuring primarily contrasting information and revealing connections between multiple factors with visualization, connection between different factors are revealed.
- Bulk of the work currently is done for data cleaning as insurance data has a lot of missing values and duplicate values.
- Further work will look into how the data is represented which will be done by Feature Engineering i.e. it is a process of transforming data into features that will improve the model performance.
- The model building activity involves multiple machine learning algorithm techniques. The algorithms are applied to previous claim data which will help making prediction on unseen data.
- Finally the model performance will be evaluated using different measures like Confusion matrix or Using ROC curves.

Conclusion

The earlier the fraud is detected in the insurance lifecycle, the lesser it costs to manage it. Analytics can play a very important role in identifying fraud because once the model is ready, we can feed more data into the training phase of the model and more training data in training phase can make a better model and we can iterate constantly. Preferably this will save a lot of manual investigation work which is the rule based approach i.e. by looking at on-surface and evident signals.