Insurance Fraud Detection

Contents

[1. Hypothesis: 2](#_Toc98398572)

[2. Problem: 2](#_Toc98398573)

[3. Project objective: 3](#_Toc98398574)

[4. Aim of the project: 3](#_Toc98398575)

[5. Plan of the project: 3](#_Toc98398576)

[1. Extraction & Preparation of Data 3](#_Toc98398577)

[2. Pre-processing of data 4](#_Toc98398578)

[3. Work to be Done 4](#_Toc98398579)

[6.Project plan: 4](#_Toc98398580)

[7. References 4](#_Toc98398581)

[Bibliography 4](#_Toc98398582)

# 1. Hypothesis:

Given the abundance of broad knowledge linked to insurance, people nowadays are impacted by many payment types, including health insurance, automotive insurance, property insurance, and travel insurance. People are increasingly opting for such insurance, making it easier for con artists to defraud them (Roy, 2017).

Insurance fraud is a crime committed by either the customer or the insurance contract's vendor. Unrealistic claims & article policies, among other things, are examples of client-side insurance fraud. Insurance fraud occurs on the vendor side in the development of rules from non-existent firms and failure to submit premiums, among other things (Wang, 2018).

The insurance sector is losing millions and millions of dollars due to fraud. This research aimed to create a machine learning method to identify (Rukhsar, 2022). The study employed previous financial data for normal and fraudulent transactions to develop normal/fraud behavior traits machine learning, which were then used to determine whether a transaction was fraudulent. Comparative research was undertaken to determine which model is ideal for this project to train the behavior aspects of regular and unconventional transactions (Party, n.d.).

# 2. Problem:

The most serious problem that insurance providers confront is fraud, which results in massive losses for them that are often irreversible. Because combating fraud cases, particularly in insurance firms, is a difficult process, the key objective is to avoid fraudulent acts at all costs. According to reports, between 22% and 34% of auto insurance claims are believed of being fraudulent, although only 3% of situations are punished. The first step in minimizing fraudulent situations is to discover them, which is complex and just not "expense" because extensive and time-consuming examinations may irritate genuine clients (Gupta, 2022).

Increasing investigative expenses also make it more difficult to discover fraud. As a result, businesses fail to conduct necessary investigations, resulting in a slew of potential hazards. Because manual fraud detection is costly and inefficient, we now need to examine the fraud before approving the claim. Several artificial intelligence & ML approaches are effective in identifying fraud (Hanafy, 2022).

This project aims to develop a model that can predict insurance fraud. The difficulty with ML fraud detection would be that scams are significantly less prevalent than legitimate insurance payments.

Given the variety of fraud types and the low number of confirmed frauds in regular sampling, detecting financial fraud is difficult. When developing detection algorithms, the expense of false warnings must be weighed against the cost of loss avoidance. Machine learning approaches improve forecast accuracy, allowing loss controllers to cover more territory with fewer false positives.

Insurance fraud refers to various unethical behaviors that a person may engage in to obtain a favorable outcome from an insurance company. This might include arranging the event, exaggerating the circumstances, including the notable characters and the incident's cause, and lastly, exaggerating the magnitude of the harm.

# 3. Project objective:

The objective is to develop a model that can accurately predict which payments are likely to be fraudulent. The standard method of detecting fraud is to create algorithms based on fraud signs. A fraud judgment might be made in two ways based on these heuristics.

In other instances, guidelines would be defined to determine if the matter needed to be investigated.

A checklist with ratings for the different fraud indications might be created in other circumstances. The sum of these ratings and the claim's worth will decide if the case has to be investigated.

# 4. Aim of the project:

This project aims to construct a model for detecting insurance fraud using machine learning. The aim of forecasting a discrete random variable is different from analysis, which deals only data sets.

# 5. Plan of the project:

We use three main steps to implement in this research:

## 1. Extraction & Preparation of Data

It is necessary to provide the data to be evaluated for fraud prediction before discussing the various categorization algorithms. A car insurance fraud dataset was used to conduct this research.

## 2. Pre-processing of data

Data pre-processing greatly slows down data mining. Data is rare, free of impossible combinations, missing values, noise, inconsistencies, and other issues. Before implementing the method, the first and most important prerequisite is data quality. Data pre-processing may impact how the eventual results are understood.

Factual information made up of categorical variables or data is called categorical data. The dataset is first examined for categorical data.

## 3. Work to be Done

Several phases are involved in the fraud detection process.

1. Pre processing
2. Prediction
3. Evaluation

# 6.Project plan:

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Activity** | **Starting date** | **Target date** |
| 1 | Choose research area. |  |  |
| 2 | Literature search (find & confirm research gap). |  |  |
| 3 | Decide on research question and/or hypothesis. |  |  |
| 4 | Write the literature review draft. |  |  |
| 5 | Scope out research (initial thesis plan). |  |  |
| 6 | Write the proposal’s draft. |  |  |
| 7 | Submit the proposal. |  |  |
| 8 | Proposal approved. |  |  |
| 9 | Literature review. |  |  |
| 10 | Writing the report. |  |  |
| 11 | Submit the report. |  |  |
| 12 | Experiment the selected tools on identified dataset. |  |  |
| 13 | Develop the comparison criteria. |  |  |
| 14 | Develop the selection algorithms. |  |  |
| 15 | Develop the GUI (Graphical User Interface). |  |  |
| 16 | Feedback the entire thesis. |  |  |
| 17 | Submit results and findings. |  |  |

# 7. References

# Bibliography

Gupta, A. a. M. C. L., 2022. Comparative Analysis of Numerous Approaches in Machine Learning to Predict Financial Fraud in Big Data Framework.. *Soft Computing: Theories and Applications. Springer, Singapore, 2022. 107-123..*

Hanafy, M. a. R. M., 2022. Classification of the Insureds Using Integrated Machine Learning Algorithms: A Comparative Study.. *Applied Artificial Intelligence .*

Party, C., n.d. Machine Learning Working.. *Machine Learning in Insurance..*

Roy, R. a. K. T. G., 2017. Detecting insurance claims fraud using machine learning techniques. *In 2017 international conference on circuit, power and computing technologies (ICCPCT), pp. 1-6. IEEE, 2.*

Rukhsar, L. W. H. B. K. N. a. S. N., 2022. Prediction of insurance fraud detection using machine learning algorithms. *Mehran University Research Journal Of Engineering & Technology 41, no. 1 .*

Wang, Y. a. W. X., 2018. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud.. *Decision Support Systems 105 .*