International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Case Study on Analytical Tools for Insurance Fraud

Dr. Kashmira Mathur

Management Development Institute, Gurugram, India

Abstract: ***Corruption in the auto insurance industry is a worldwide concern. For insurance firms, manually addressing fraud is always expensive. Data science may be quite beneficial in the fraud detection process and can aid insurance companies to detect fraud. For the fraud analysis, typically, there are probably over forty variables. The purpose of this study is to identify the factors that are crucial for detecting fraud and to offer a framework for doing so. This paper also uses empirical research to demonstrate the commercial use of data analytics for detecting insurance fraud. It shows how the insurance firm can accurately identify fraudulent claims by adopting a few business guidelines, which will probably lead to cost reduction and higher profitability for the business.***

Keywords Fraud Detection, Analytics, Insurance, Significant Variables, Business Guidelines, Framework

1.Introduction

An anticipated 10% of claims made with the Indian insurance industry are fake, while only a single figures percent of all claims are typically stopped or reclaimed by claimant management financial crime units. [8]

What is fraud?

Fraud occurs if someone intentionally utilizes trickery to acquire something illegally or wrongfully. The conduct of fraud can be categorized either as a criminal or civil offense in the majority of countries. While obtaining rewards of value is the core reason fraud is done, fraudulent activity can also happen for the sole purpose of manipulating another individual or organization. For example, based on the scenario, making incorrect declarations may be considered fraud.

Statistical hypothesis testing techniques will be used in this paper. In this paper, it is recommended that by deriving business rules to facilitate fraud detection, it will become feasible to determine the variables that are crucial for detecting and preventing and establish a structure for insurance fraud detection.

Further, this paper uses empirical auto insurance data and is structured into 6 sections.

Section 1	The introduction focuses on the importance of detecting fraud and establishing frameworks that can warn the company when something is out of the norm and demands further inquiry.
Section 2	Provides a concise literature review.
Section 3	Provides the purpose of the study
Section 4	Outlines the possible methods for detecting fraud.
Section 5	Exposes the results of the data analysis, including the statistical significance of relevant factors and solutions for the adoption of obtained business requirements to improve the fraud detection phase.
Section 6	Displays the conclusion

2.Literature Review

Automobile insurance scam is a global issue, not simply an Indian one. Our concentration is on automobile insurance, although fraud occurs in other types of insurance as well. Insurers understand the value of data analytics in fraud detection and are ready to select premium fraud solutions that do not match their gaps or strengths. In fact, Spathis claims that fake accounts have become more widespread in recent years. Furthermore, there is an emerging trend for improved access, clarity, and more content in financial statements. Spathis developed a model for detecting false income reports. He used a numerical tool with two financial ratio inputs. The claimed performance rate was higher than 84%. This study justifies and promotes our use of analytics to detect fraud. Coston highlights how business rules and anomaly detection are generally the first steps in fraud detection, analysing each claim against tools helps to track common malpractices by recognizing specific pattern types. Constant transaction analysis empowers an enterprise to discover fraudulent activity on a daily, weekly, or recurring basis. Fraud detection with a focus on evaluating false claims using faulty data. The study revealed great results, with performance ranging from as 87%. Therefore, businesses should concentrate on active surveillance operations on specific trade segments or risk-prone geographic regions.

As a result, we focused on statistical hypothesis techniques to detect fraud. Long-term usage of data analytics techniques should lead to a reduction in the rate of fraud, a decrease in the average amount of time spent on fraud research for each claim handled, a decrease in the total amount paid out, and a decrease in the unallocated loss adjustment costs involved to investigate fraud. Additionally, this research was started with this idea.

3.Objective of Study

Three fundamental goals constitute this paper.

Licensed Under Creative Commons Attribution CC BY

1) The purpose of this study is to show how analysis of the data can be utilized to find fraud and the most important factors in detecting fraud.
2) The purpose of this study is to generate business rules from important variable information. The marking of risky material requires the business rules that were derived. After being informed, the company conducts a further analysis of these reported occurrences to determine whether they are actually fraudulent or not.
3) This research will propose a methodology for detecting fraud based on the experimental data used, which may be applied to similar detecting fraud studies.

4.Methodology

4.1. Framework for Fraud Detection

In order to help business users, recognize fraud, we have created a framework. To determine which characteristics were significant and could help identify fraud, we conducted statistical experiments on 31 variables using the data set given from Angoss Knowledge Seeker software. We analyze fraud and non-fraudulent claims using the significant variables that have been established. The main factors and the fraud profile also aid in the development of the business rules used to spot future fraudulent claims. The steps taken to recognize and detect fraud are shown in figure 1 below.

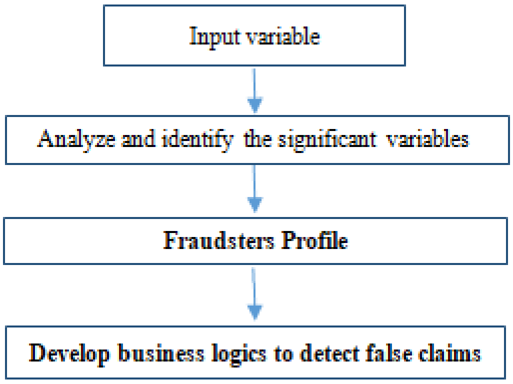

Figure 1: Fraud detection framework

According to the framework, our first step is to analyze the importance of several elements in identifying fraud. Once we've identified the significant traits, we may use them to identify criminals and build business criteria for detecting fraudulent claims.

Identification & Usage of Significant Variables

Given that the data set contains a lot of input variables (31), it makes sense to assume that some of them may not have any real effect on detecting fraud. The data were subjected to the Chi-Square test and independent sample t-test, two distinctive significant tests, using SPSS Statistics 22 to help identify which variables were actually important and might help catch fraud. Any relationship between any two of the category variables under consideration was checked using Pearson's Chi-square Test for Dependence. If there is less than a.05. significance level (in the case of a 95 percentage error range), the null hypothesis that there is no correlation between the two variables is rejected. This demonstrates how the group and forgery are related to one another. A further testing requirement is that less than 20 percent of the total cells have a score of less than 5. The Pearson Chi-Square criterion is used in this instance to determine the p-value. If this presumption is incorrect, we examine the likelihood ratio metric's p-value. The independent sample was used to compare the means of the data samples for the two distinct groups, fraud and non-fraud. The p-value is less than.05 if there is a sizable difference in means between fraudulent and legitimate claims, denoting that the two inputs help identify fraud. Important aspects will be used to distinguish between fraudsters and non-fraudsters and to create business regulations.

5.Analysis of Results

5.1. Significant Variables

The Chi-Square test was used to examine association and the Independent Samples T-Test was utilized to test differences between means for variable significance testing. For detecting fraudulent claims, 31 variables (30 categorized and 1 continuous) are taken into consideration.

Table 1, below provides a list of the 31 variables used.

Test of Variable Significance			
Month*	**Week of month claimed**	**Age ***	**Rep Number**
Weekday	Make Fin *	Fault *	Deductible *
Week of the month *	Accident Zone*	Policy kind*	Driver rating
Month claimed	Gender *	Vehicle class*	Policy accident
Days of week claimed	Marital status	Vehicle price*	Policy claim
Previous number of claims *	Age of vehicle *	Age PHF in *	Police report filed*
Eyewitness present	Agent kind*	Amount of supplements*	Address variation claim*
Numbers of cars	Year*	Base policy*	

The 0.05 significance threshold was used to determine whether or not a variable was useful for detecting fraud, i. e., if the p-value was 0.05, we concluded that the data was meaningful. In the previous paragraph, significant variables were denoted with a '*' in table 1. Twenty of the 31 factors were found to be statistically significant. These aspects were crucial in defining business requirements for detecting fraudulent claims and profiling false claims.

Fraud Team Portfolio

We use the analysis findings from the determination of Significant Factors in segment (5.1) above to develop profiles of the fraudsters. Our primary goal was to take a closer look at the significant variables to understand why

Licensed Under Creative Commons Attribution CC BY

there was such a large difference between fraudulent and authentic claims for the 20 key parameters.

The following is an overview of the major conclusions from the analysis of data and important factors:

Demographic Category of the Fraud Group:

1. Fraudulent claims typically occur in urban areas.
2. On average, men commit scam much more often than women do. Year: It is falsely asserted that the ten events took place within the first two years, rather than later.
3. Drivers under the age of 36 are more likely than older drivers to commit fraud. Updated address policyholders appear to be more likely to commit fraud.
4. In comparison to other parties, insurers are more likely to be negligent and dishonest.
5. False claims are much more likely to exist than none at all.
6. Usually, there is no need to file a police report in the case of fraudulent claims.

Vehicle Category of the Fraud Group:

1. Honda and Toyota vehicles are typically the subject's majority of false claims.
2. Owners of vehicles five years old or older are more likely to file fictitious claims.
3. Claims pertaining to sedans are frequently more likely to be faked.
4. Low-value vehicles are more likely to be the subject of fraudulent claims (under RS.30, 000).

Policy Category of the Fraud Group:

1. Liabilities are less likely to be false claims than Accident or All Peril kinds.
2. It is more likely that a third party will handle the false claim.

Claim Features of the Fraudulent Group:

1. Typically, fraudulent claims have a history of two to four prior claims.
2. The middle of the month is when the majority of fraudulent claims are submitted.
3. The months with the highest accident rates are frequently January, March, June, July, October, or December.
4. The months with the highest likelihood of fraudulent claims are January, May, October, and November.

The twenty traits listed above aid in giving us a picture of the fraudulent group. We derive the 20 business rules for spotting fraudulent claims in section 5.3 below.

5.3. Derived Business Rules for Detecting Fraud

We can summarise the findings of our analysis and the insights found in the twenty rules derived below after carefully analyzing data and variables for fraud identification. We suggest that these 20 rules be applied as follows to all claims in the future:

5.3.1. Developed Business Rules to Determine if the Claimant Meets the Demographic Requirements for a Fraud Profile

The following four generated rules are very likely to be used to categorize the fraudster in terms of their demographic traits:

1. Is the claimant a "Man"? Give the claimant a score of 1 if the answer is "yes," or else, a score of 0.
2. The driver must be "less than or equivalent to 36" years old. Give the claimant a score of 1 if the response is "yes," otherwise, a score of 0.
3. Has the policyholder's "Address" ever changed? Give the applicant a score of 1 if the response is "yes," otherwise, a score of 0.
4. Was the accident actually the policyholders' "fault"? Give the claimant a score of 1 if the response is "yes," otherwise, a score of 0.

At this point, a claimant may receive either the lowest score of 0 or the highest score of 4. A claimant who receives the highest score of 4 is advised to have their claim processed immediately using the business rules that were inferred from their "Claim Features" score.

5.3.2. Tests Based on Derived Business Rules to Determine if the Claimant Meets the "Claim Attributes" of a Fraud Portfolio

The following ten rules were created to assess whether a claimant is likely to be a fraudster based on the claim criteria:
In which of the following "months"-January, March, June, July, October, or December-was the accident said to have occurred? In that case, give the applicant a 1, otherwise a 0.

Whether the "month claimed" occurred in January, May, October, or November? Give the applicant a 1 if this is the case; otherwise, a 0.

Was it the middle of the month during the "week of month claimed"? Give the applicant a 1 if this is the case; otherwise, a 0.
Have between two and four "past claims" been submitted by the claimant? Give the applicant a 1 if this is the case; otherwise, a 0.

Is the claim made within the policy's first two "years"? If yes, give the claimant a '1'; otherwise, give them a '0'.

Are the first two "years" of the policy covered by the claim? If so, give the applicant a score of 1, otherwise a score of 0.

Was the "accident area" situated in a city? If so, give the applicant a score of 1, otherwise a score of 0.

Did it say "no supplements" in the claim? If so, give the applicant a score of 1, otherwise a score of 0.

Licensed Under Creative Commons Attribution CC BY

The claim may have stated, ”No police report filed." If so, give the applicant a score of 1, otherwise a score of 0.

A third-party "Agent Category" handled the claim, if so. If so, give the claimant a rating of 1, otherwise a score of 0.

At this point, we can compute the overall score for each claimant. The minimum and maximum overall scores are 0 and 14, respectively. The company can select the optimal score for the claimant's 'Vehicle Class' characteristics to be processed immediately.

5.3.3. Derived Procedures used to determine whether a claimant has the 'Vehicle' Features of a Fraud Credentials.

In terms of vehicle characteristics, the four main guidelines were developed to aid in determining if the claimant is inclined to be a fraudster:

1. Is your car a Toyota or a Honda? If yes, award a '1'; otherwise, award a '0'.
2. Is your vehicle five years old or older? If yes, award a '1'; otherwise, award a '0'.
3. Is the sedan in your 'Vehicle Category'? If yes, award a '1'; otherwise, award a '0'.
4. Is your car's "vehicle cost" less than RS.30, 000? If yes, award a '1'; otherwise, award a '0'.

The claimant's score at this point can range from 0 to 18, with 18 being the maximum. The business may select the threshold score at which they want the claimant's 'Policy Class' features processed right away.

5.3.4 Derived Business Rules to Determine if the Claimant Meets the Policy Form

The four major rules were created to help determine whether the applicant is likely to commit fraud based on the features of the claimant's policy type:

1. Is your 'policy type' sedan-all perils or sedan-collision? If so, give the applicant a score of 1, otherwise a score of 0.
2. Which perils or collisions are covered by your "base policy"? If so, give the applicant a score of 1, otherwise a score of 0.

At this point, the claimant's rating can range from 0 to 20. The company may decide that all claimants with a final overall score of 16 or higher should have their claims processed right away, with all pertinent data provided, and that solid evidence should be sought to ascertain whether or not the claim is fraudulent. If time and resources permit, claimants with rankings greater than ten but less than 16 ought to be highlighted and given more attention.

An overview of the resulting regulatory requirements is shown in Figure 2 below. It should be observed that step 4 of figure 1 is represented in figure 2, and information on the varying data analysis scores is provided in sections 5.3.1 to 5.3.5. Businesses are strongly advised to use their derived norms in a systematic way to ensure the stability of their fraud identification and tracking process. It should be noted that the generated codes can be modified and improved as new data becomes available over time.

6.Conclusion

Insurance companies have started to recognize the value of business intelligence in the domain of fraud detection and have hurriedly chosen pricey fraud remedies that are not a good fit for their company's strengths and weaknesses. Insurance companies should use straightforward data analytic techniques, like statistical significance testing, to identify fraudulent claims. From there, business rules can be derived, allowing for the development of a framework similar to the one in this paper. This paper illustrated how key variables, such as the claimant's demographics, the claim characteristics, the policy, and the vehicle type, may be used to quickly identify fraudulent claims. By focusing on fewer variables (in this case, twenty instead of thirty-one, a reduction of 35 percent inside the number of variables), the business can become more efficient and save time and funds. This greatly aids in the investigation of fraud as more time and attention can be given to the important variables by using the derived business requirements. By cutting down on both the time and expense of conducting fraud investigations, the framework for fraud detection outlined in this paper is expected to increase the effectiveness of fraud investigations.

Licensed Under Creative Commons Attribution CC BY

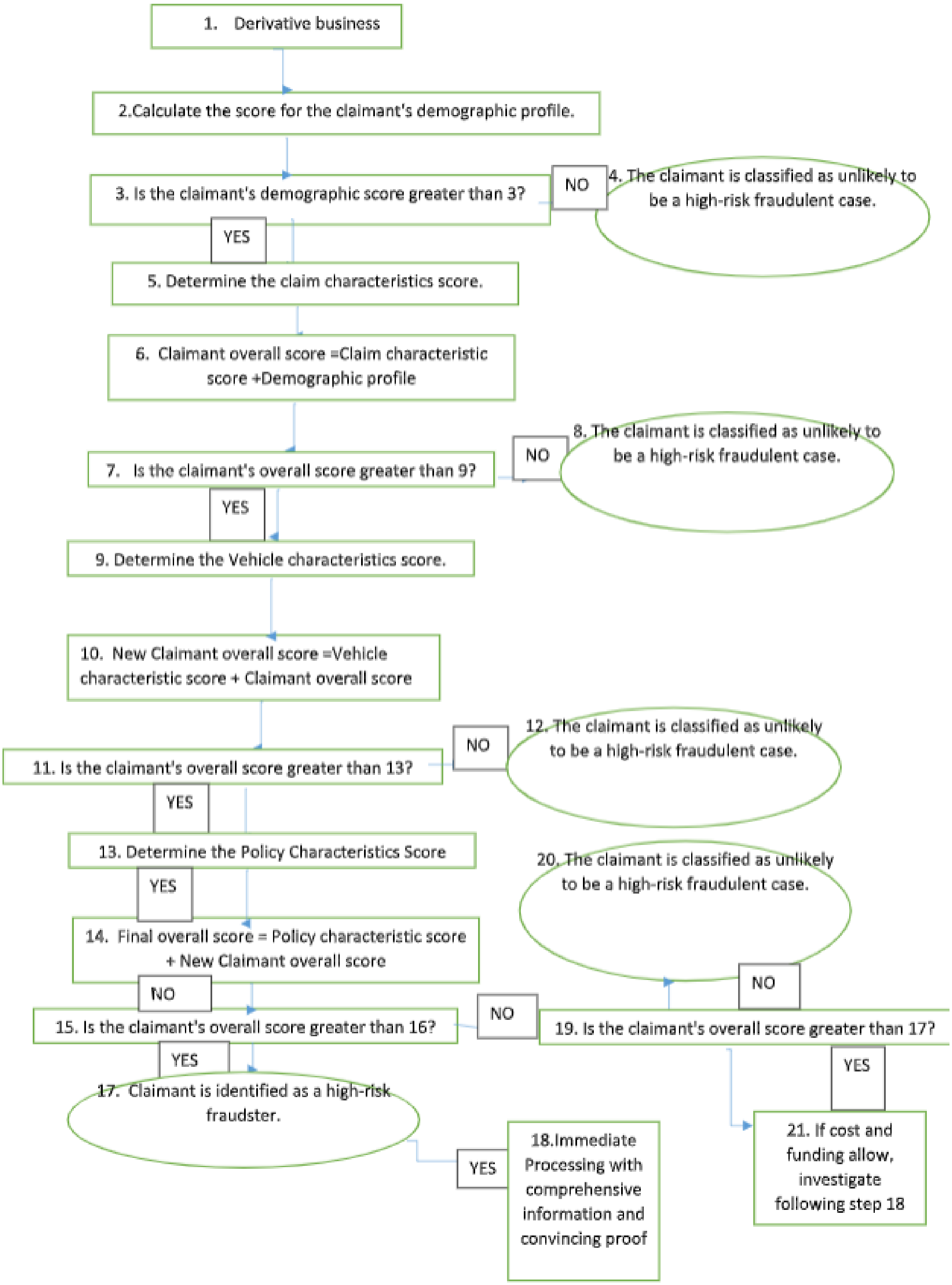

Figure 2: Obtained Procedures for Detecting Fraud

References

[1] Bolton, R. J. & Hand, D. J. (2002). "Statistical Fraud Detection: A Review", Statistical Science, Vol.17. No.3, 235-255.

[2] Artis, M., & Mercedes, A., & Montserrant, G. (2002). Detection of Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims.

[3] Verma, R. & Sathyan, R. M. "Using Analytics for Insurance Fraud Detection: 3 innovative methods and a 10-step approach to kick start your initiative". Digital Transformation. Pages 1-10.

[4] Costons, M (2010). "Analytics and Claim Fraud: Assembling the proper toolbox to prevent and detect scams". Claims Magazine. Page 43 - 45.

[5] Spathis, C. (2002). "Detecting falsified financial statements using published data: some evidence from Greece ". Managerial Auditing Journal, 17 (40, 179 - 191

[6] Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: a

Licensed Under Creative Commons Attribution CC BY

comparative study using multicriteria analysis and multivariate statistical techniques.
[7] The European Accounting Review, 11 (3), 509-535. [11] Phua, C., Lee, V., Smith, K. & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research, Artificial Intelligence Review (2005) 1-14. [12] www.angoss.com/

Licensed Under Creative Commons Attribution CC BY