IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 7, July 2020

Auto Insurance Fraud Detection

Kavya Priya M L[1], Anusha Y G[2], Amrutha T[3], Harsha R[4], Harshitha M R[5]

Assistant Professor, Department of Computer Science and Engineering,

Maharaja Institute of Technology, Mysore, Karnataka[1]

Student, Department of Computer Science and Engineering, Maharaja Institute of Technology, Mysore, Karnataka[2,3,4,5]

Abstract: Fraud is the activity which will cause distress to corporations. This Financial fraud has been a huge worry for many organizations across the industries, this insurance industry comprises of over and above thousands companies all across the board, which gathers greater than a trillion dollars premium every year and billions of dollars are being lost every year because of this fraud, thus detection of Insurance Fraud is a burdensome task for the insurance companies. The conventional outlook for detecting insurance fraud was completely dependent on evolving heuristics around the fraudulent indicators. The auto insurance fraud is being considered to be one of the leading categories of fraud, which will be carried out by faking accident claim. In this study, we are concentrating towards tracking down the auto insurance fraud by making use of Machine Learning Techniques (Naïve Bayes Classifier), by using this techniques time complexity will be declined and also depicts the results accurately.

Keywords: Machine Learning techniques, Auto Insurance, Fraud detection, Naïve Bayes Classifier.

I. INTRODUCTION

Insurance Fraud: It is defined as an improper activity which may be committed by individuals in order to obtain a beneficial outcome from the insurance company. There is different type of insurance provided for example we have Health Insurance, Agricultural Insurance, Business Insurance, Life Insurance, Auto Insurance and many more. In this project we are considering only Auto Insurance Fraud Detection

This Insurance Fraud is classified into two types

1. Hard Insurance Fraud: This is a type of fraud in which accident has not taken place at all but even though the individuals are going to claim for an insurance.
2. Soft Insurance Fraud: This type of fraud is also referred to as opportunistic fraud. If accident has taken place or if any automotive collision has occurred an insured person might claim for an insurance by increasing the severity of damage to the vehicle This soft insurance fraud is more common when compared to that of hard insurance fraud.

Some of the potential satiation in which fraud can takes place are:

1. Some of the situation such as Drink and driving, performing some risky activities, using mobile phones while driving etc. if accident takes place during these circumstances then insurance companies are not responsible for providing the insurance, yet if accidents takes place during these situations' individual is going to claim for an insurance.
2. Accident has not taken place at all, but what an individual does is that he is going to create a beautiful story about the accident and narrate it to the insurance companies.
3. Accident has taken place, but the amount of damage to the vehicle is very less but what an individual does is that insured person is going to claim for an insurance saying that huge amount of loss has occurred.

India is one of the Biggest markets for insurance industries all over the world, yet it is not free from risk the reason why is it not free from risk is that , Indian Insurance Industry Loses to around $6 billion every year to this insurance fraud in India. FBI which is also an insurance company in USA, under which there are 7000 companies and according to it the overall value lost to this fraud is potentially greater than $40 billion per annum. Hence the Insurance Industry have a urgent need to develop a capability which can help them identify weather the given insurance claim is fraud or genuine with high degree of accuracy and with less amount of time. This will also help in maintaining the customers satisfaction and also the trust towards that insurance company.

II. EXISTING SYSTEM

Traditional approach consists of two ways:

1. Based on Certain rules: Here there would be a committee and that committee would create certain rules based on that rules, they would define weather the given insurance claim is fraud or genuine, if it is fraud then the it would be sent to investigation.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 7, July 2020

2. A checklist would be prepared based on the aggregation method: Here they would prepare a checklist which includes certain indicators along with the scores associated with that and what they would do is that , an aggregation of all the scores along with the value of the claim would be calculated and then this calculated value will be correlated with the boundary value, which would have been calculated earlier, if this calculated value is larger than the boundary value then the case would be sent to investigation else insurance would be provided.

These traditional approaches has certain limitation

1. Minimum number of parameters
2. Maximum human intervention

So, to overcome these drawbacks we are developing a machine learning model which predicts whether the given claim is fraud or genuine.

III. LITERATURE SURVEY

A paper titled **"Robust Fuzzy rule-based techniques to detect frauds in vehicle insurance"** [1] uses Fuzzy logic by framing the fuzzy rules to improve the Fraud detection. The fuzzy rules-based techniques will be enforced on to the training dataset and based on instance the level of fraudulent and genuine is detected, this procedure will be used for high capacity and huge dataset, but the major drawback of this rule is that it involves lot of human interventions.

A paper titled **"Medicare Fraud detection using machine learning methods"** [2]. This paper does an investigational survey by studying multiple supervised as well as unsupervised classification methods to discover the fraud cases. They have considered 3 groups or methods, 1. The Supervised learner which includes Random forest, Deep Neural Networks, Naïve Bayes Etc., 2. The Unsupervised learner includes KNN. Autoencoder etc., 3. The Hybrid learner includes NN Model. Their results show that Supervised learners are significantly better when compared to that of other type of learner.

A paper titled **"Nearest Neighbors and Statistics Method based for detecting fraud in auto insurance"** [3]. This study uses Nearest Neighbors method and statistics method to detect the occurrence of fraud and these methods are explained in detail in this paper. They made the comparison between SVM Method and the method that were used in this study, and found that when compared to SVM and interquartile range (statistic method), Nearest Neighbors method provides best performance.

A paper titled **"Detecting insurance fraud by using Data mining Techniques"** [4]. Makes use of 3 algorithms they are Bayesian Network, C4.5, Decision tree-based algorithm for predicting and analyzing fraud pattern from data. This model prediction will be aided by Bayesian naïve visualization, Decision tree visualization and rule-based classification. They looked at model performance matrices like recall, accuracy, and precision. This will be stronger when compared to that of class skew, by making it reliable performance matric in numerous prime insurance fraud detection functional areas.

A paper titled **"Application of data mining techniques in Health Fraud Detection"** [5]. In this study they have proposed a health fraud detection using separate procedures by using ID3 (iterative dichotomiser 3), J48 and Naïve Bayes. Based on the survey they could clearly say that the highest accuracy is ID3 with cent percent and least accuracy will be J48, and finally concluded that Decision tree is best when compared to other 2 algorithms.

The outcome of this survey is that, we came to know that supervised learners are better when compared to that of other learners and decision tree classification algorithm is best when compared to other algorithms, but the major disadvantage of using this decision Tree is that, it works fine for smaller datasets, but as the datasets increases performance decreases, hence in this proposed system Naïve Bayes algorithm is being used which works fine for both smaller as well as large datasets and works fine for any type of datasets.

IV. METHODOLOGY

This proposed system includes 3 modules / 3 actors, they are

- **Admin:** The one who maintains the complete application.
- **Branch in-charge:** The one who receives the insurance claim.
- **Public:** The one who claims for the insurance.

A. ARCHITECTURAL DESIGN.

When public claims for an insurance, branch in-charge receives the claim and performs some investigation like, whether the public has claimed for an insurance before, if so what type of claim was it whether it's a fraudulent or genuine case, whether the premium has been paid properly then sends 8 parameters as input to our Insurance Fraud System. Our system

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 7, July 2020

performs some analysis and outputs as either fraud or genuine. If genuine then payment will be provided else case will be sent to investigation.

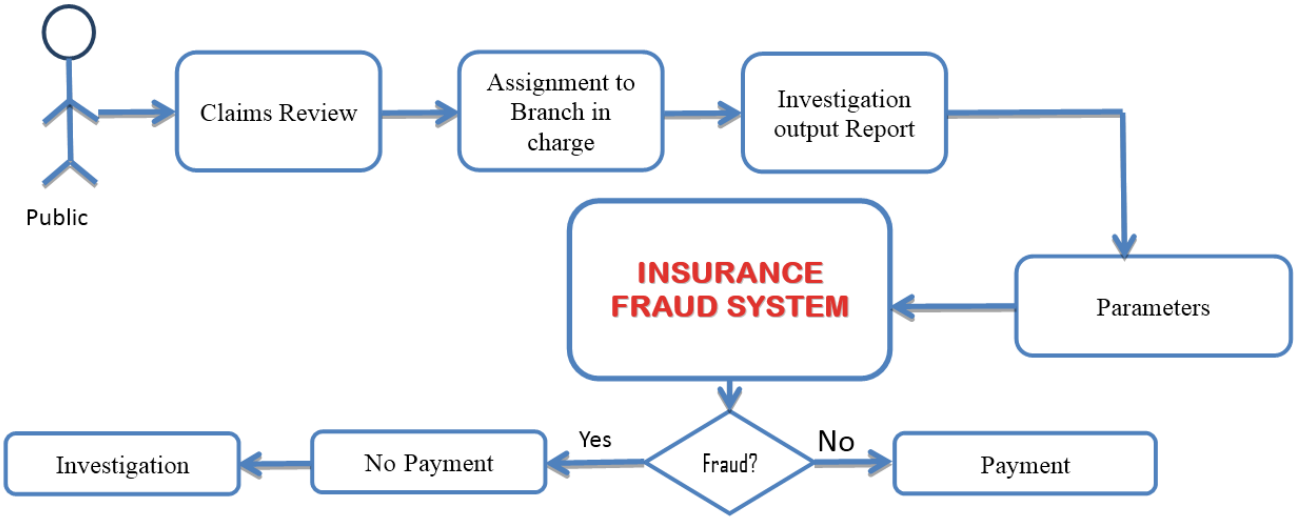

Figure 1: Architectural design of the proposed system

B. FLOW CHART

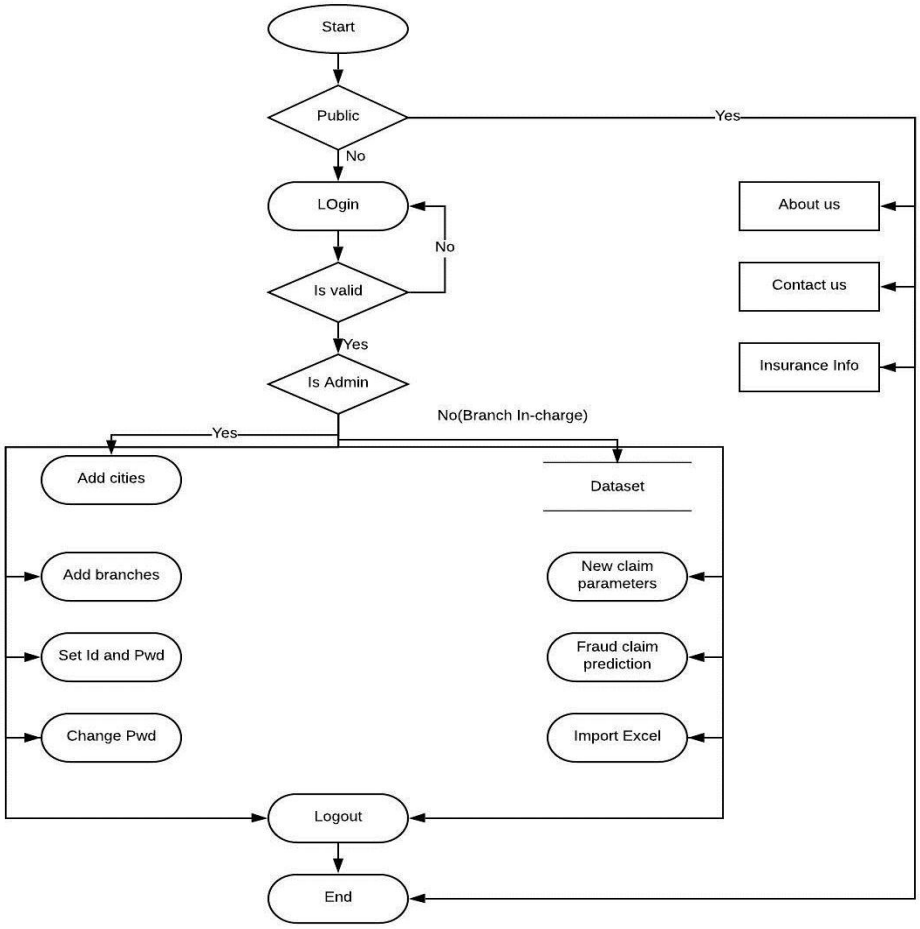

Figure 2: Flow chart of the model

Suppose public visits the application then, there is no need for public to login into the system, public will be directly directed to the home page where he can view different information related to insurance like about us, contact us, insurance info etc, if not public then that have to login into the system. On successful login they will be directed to their home page

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 7, July 2020

if not they will be redirected back to the login page, suppose if the logged in person is Admin then admin can perform various operations like he/she can add different cities, within a city there could be multiple insurance company so adds branches, these n number of cities and n number of branches can not be managed by a single person i.e, Admin, hence he/she creates a branch in-charge by providing them the Id and password , then admin can change his/her password. If the logged in person is Branch in-Charge then various function which he/she could perform are can add/delete a item to/from the Datasets, inputs 8 parameters into the system, receives the output and based on it he/she will make decision, suppose branch in-charge is using this application for the first time then he/she needs to import the dataset from excel sheet then from the next visits no need to import it again and again. Finally they can logout from the system, they will be directed to the application home page.

Table 1 shows the list of parameters that is being consider along with their description

Sl.n o	*Parameters*	*Description*
1.	DCOD_CR D	Variation between claim event date and claim turn date (datatype - numerical)
2.	DCRD_CO PD	Variation between claim turn up date and claim open date (datatype - numerical)
3.	DPE_COD	Difference between policy effective and claim event date (datatype - numerical)
4.	CDS	Are claim document submitted (1-yes, 0-no)
5.	PCD	Part cost difference (datatype- numerical)
6	CR	Credit rating (1-yes, 0-no)
7.	PP	Policy premium (datatype -numerical)
8.	CCC	Count of customer communication (datatype -numerical)

Table1: Parameters used

C. ALGORITHM USED

Algorithm that is being used is Naïve Bayes Algorithm

Step 1: Browse the dataset.
Step 2: Probability of each parameters value will be calculated.
Step 3: Probability will be calculated using this formula

$$P(\text{attrvalue}(ai)/\text{subjectvalue}(vj)=(n\_o+ n*p)/(n\_v+n)$$

Where,

- n_v = the number record in training dataset where v=vj
- n_o = number of records in training dataset where v = vj and a = ai
- p = probability outcome
- n = number of parameters

Step 4: calculated probabilities will be multiplied with each other.
Step 5: Probability of case being fraud and genuine will be compared and the one with highest value is sent as output.

V. RESULT ANALYSIS AND DISCUSSION

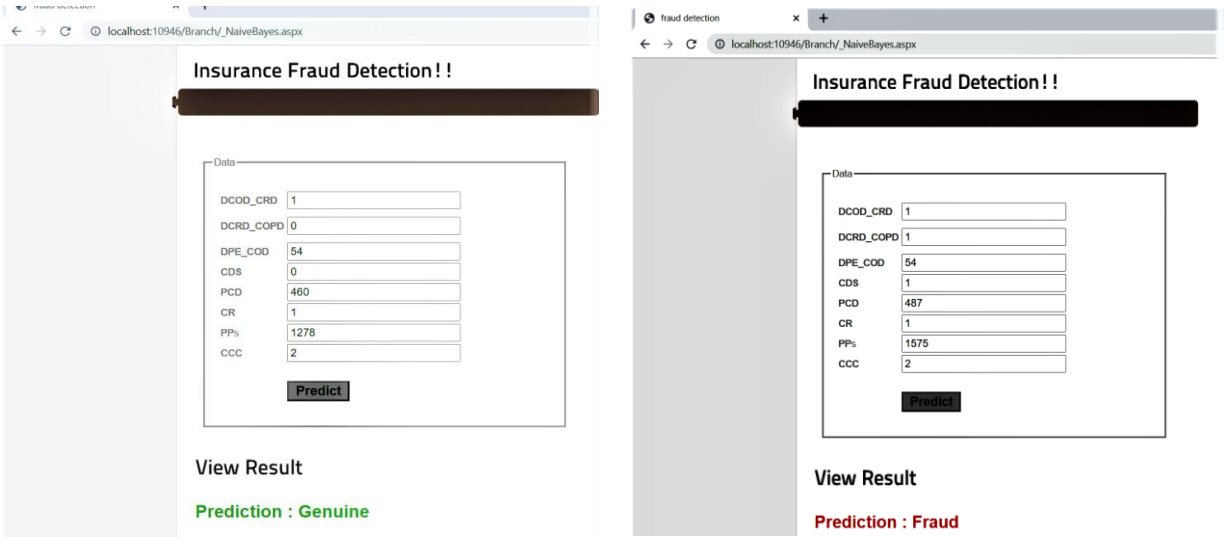

Figure 3: Insurance Fraud Prediction

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 7, July 2020

When public claims for the insurance, branch in-charge receives it and inputs all 8 parameters into the system and when he/she clicks on the predict button, internally naïve bayes algorithm will be executed and the system provides the output as either fraud or genuine.

From Fig 4 we can see that by using Naïve bayes classifier we have obtained the accuracy rate 95% and the time taken to execute in milliseconds is 751 and also we can observe that 95% of the test cases are being correctly classified while 5% of them are incorrectly classified.

Naive Bayes	Constraint
Accuracy	95%
Time (milli secs)	751
Correctly Classified	95%
InCorrectly Classified	5%

Figure:4 Result Analysis

VI. CONCLUSION

Insurance fraud detection is a rough task, this industry has grappled with challenges of insurance claim fraud from the very beginning. Proposed system aims at developing a system that can help to recognize possible frauds with peak magnitude of accuracy. Proposed system predicts whether the claimed insurance is “FRAUD” or “GENUINE”. Thus helping the insurance companies to spot frauds with fewer amount of time and with good accuracy rate

VII. FUTURE ENHANCEMENT

System can be enhanced to predict the insurance fraud in bulk. System can be enhanced by adding feedback module, where users can posts feedbacks to clarify their doubts and also it can be upgraded by adding a report module, so whenever the output is Fraud, if branch in- charge clicks on the report button complete report must be sent to the nearby police station.

REFERENCES

[1] “Robust Fuzzy rule based techniques to detect frauds in vehicle insurance”, K. Supraja , S.J. Saritha , 2017.
[2] “Medicare Fraud Detection using Machine Learning Methods”, Richard A. Bauder , Taghi M. Khoshgoftaar, 2017 .
[3] “Nearest Neighbour and Statistics Method based for Detecting Fraud in Auto Insurance”, Iwan Syarif, Tessy Badriyah, Lailul Rahmaniah.,2018.
[4] “Detecting Auto Insurance Fraud by Data Mining Techniques”, Bhowmik, R., Journal of Emerging Trends in Computing and Information Sciences, Volume 2 No.4, april 2011.
[5] “Application of Data Mining Techniques in Health Fraud Detection”, Rekha Pal and Saurabh Pal, October 2015