Data Analysis for Vehicle Claims

Reena.D[1], Sathyan.S[2], Vinayagamoorthy.G[3], Sreeja.G.G[4]

UG Students[1,2,3] Assistant Professor[4]
Department of Computer Science and Engineering, KGiSL Institute of Technology, Coimbatore, TamilNadu, India

Abstract-

A system that helps in analyzes the motor insurance vehicle claims and visualizes the data in a form of charts and graphs. It will be useful for future prediction of the data by using those charts and graphs and comparison of that data with a past years vehicle insurance claim data activities. Claims analysis is a technique for examining the consequences of design features that are described in current or future scenarios of use. A "claim" is a statement of the consequences of a specific design feature or artifact on users and other stakeholders. This system is planned to build using the TABLEAU software to visualize the data. This software is a commerce ability tool for visually analyzing the data, and provides the information to improve the insurance claims.

Keywords: Data collection, Analyze and visualizing the data, Comparison of data, Prediction and Report creation.

I.   INTRODUCTION

Vehicle insurance (also known as car insurance, motor insurance or auto insurance) is insurance for cars, trucks, motorcycles, and other road vehicles. It will plays a key roll in financial industry and it is also importance for people economical life. Its primary use is to provide financial protection against physical damage or bodily injury resulting from traffic collisions and against liability that could also arise from incidents in a vehicle. Vehicle insurance may additionally offer financial protection against theft of the vehicle, and against damage to the vehicle sustained from events other than traffic collisions, such as keying, weather or natural disasters, and damage sustained by colliding with stationary objects. The specific terms of vehicle insurance vary with legal regulations in each region. Analyze and visualize the motor insurance vehicle claims automatically in order to make more understandable.

Claim analysis is a technique for examining the positive and negative consequences of design features that are described in current or future scenarios of use. A claim is a statement of the consequences of a specific design feature on user and stakeholder

The main objectives of the project are as follows:

• Database: Relevant information of vehicle insurance claim are collected and stored in the database

• Algorithm: For data cleaning process some sort of algorithms are used in order to retrieve the useful data.

In this our system, our main objectives are as follows:

• Working on large number of data

• Classify the insurance claim details

• Visualizing the data in the form of charts and graphs.

• Will predict the future claim rate of the insurance.

The study seeks to analyze the claims of motor insurance in public and private insurers and their effects. In this particular case, it was believed that the study would have the following specific significances.

• It briefly explains the importance of motor insurance and enhances knowledge of the customers.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

- It depicts the extent of customer satisfaction with existing motor insurance operational practices.

- It reveals the effect of centralization in prompt claim settlement and flexibility of premium discounts between the public and private companies.

- It will be used as a lesson and creates opportunity for detail and subsequent research on the same area.

The main goal of this project is to analyze and visualize the insurance claim by using some sample data .The analyzing process is done by some criteria.

II.   LITERATURE REVIEW

Nowadays, People choose various insurance policy to assist themselves and their properties. There are various insurance type which are followed by people such as property, health, medical and auto insurance. Car insurance is mostly imposed by the authority for those who own and drive vehicles and it was established in the early 1900s, and has been applied since then. It is now compulsory in most residences around the world.

Data analysis is a process of inspecting, cleaning, transforming, and modeling data with the objective of discovering useful information, informing conclusions, and supporting decision-making. In today's business, data analysis is playing a role in making decisions more scientific and helping the business achieve effective operation.[1] Data analysis is a process for obtaining raw data and converting it to useful information for decision-making by users. Data are collected and analyzed to answer questions, test hypotheses or disprove theories.[2]

The raw data may be incomplete, contain duplicate values, or contain errors. The need for data cleaning will arise from problems in the way that data are entered and stored. Data cleaning is the process of detecting and correcting these errors. Common tasks include record matching, identifying inaccuracy of data, overall quality of existing data,[3] reduplication, and column segmentation.[4] Such data problems can also be identified through a variety of analytical techniques.

Once the data are cleaned, it can be analyzed. Analysts may apply a variety of techniques and Data visualization may also be used to examine the data in graphical format, to obtain additional insight regarding the messages within the data. [5] Once the data are analyzed, it may be reported in many formats to the users of the analysis to support their requirements. The users may have feedback, which results in additional analysis.

Data comparison is the calculation and display of the differences and similarities between data objects, typically text files such as source code. Regression algorithms are used for predictive function. The linear regression model analyzes the relationship between the response or dependent variable and a set of independent or predictor variables.

III.   PROPOSED SYSTEM

Our proposed system is automated process which follows from collecting the data from the users, using the Google forms or an insurance company's own web or mobile application.

After collecting the necessary data from the subscriber the analyzing process is carried out using the table au software and analyzing the data means organizing or structuring the data.

The visualization will be shown accurately using the same software. The comparison will be carried out simultaneous and the future prediction is analyzed

Proposed System is to implement latest technology for making the system still more beneficial to cover the core areas. Apart from the previous mentioned techniques some of the other techniques involved in the paper are described below:-

A.   Collection

The data is required in order to process visualization. So we should collect data from the user and then process, analyze, visualize and so on. The only process we are carrying manually is the data collection. This is a manual process. This process is done by distributing forms to the subscribed users. The different type of vehicle insurance details of different customers will be collected by using the Google forms and it will be stored as a CSV file in a database.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

B.   *Data cleaning data storing*

Data cleaning will be performed in python by using Python's Pandas and NumPy libraries to use clean data

We will cover the following:

•   Dropping unnecessary columns in a Data Frame

•   Changing the index of a Data Frame

•   Using .str() methods to clean columns

•   Using the DataFrame.applymap() function to clean the entire dataset, element-wise

•   Renaming columns to a more recognizable set of labels

•   Skipping unnecessary rows and columns in a file

After the data cleaning will be completed that data will be stored in database with .CSV extension

C.   *Analyzing*

The cleaned data will be retrieved from the database and it will be analyzed based on the customer needs and problems. It will be performed by following steps

1 .Define a problem-The main objective will be defined

2 .Determine the problem-

3 .Design the problem-the problem and its solution will be designed

4 .Analyze the problem

5 .Sample

6 .Report- The report will be generated in a form of charts and graphs

Those graphs and charts will be used to perform a multiple operation like comparison and prediction

KMEANS, Linear Regression are the algorithms to use for the analysis process.

D.   *Visualization*

This is the graphical representation of our analyzed data. Visualization will be better if the analysis is better. This process is automated by jupyter followed by python. The more we see the more we understand. This is basic idea of visualization and this is what we carry out.

E.   *Comparison*

The analyzed data will be visualized in many forms of charts and graphs. These charts and graphs will be used to perform a comparison operation. Comparison is carried out for better understanding of data. We can compare it with our previous data and can predict our growth in which business call YOY called as year on year growth. Data must be compared for better understanding of ourselves. The data activity will be compared year by year and find the percentage rate of claim, deduction, income rate and   other details.

F.   *Prediction*

We should be able to predict the future responses. However the prediction may not be accurate. Data prediction is best when carried out with data comparison. Sometimes prediction can save us from huge disastrous for instance, when YOY starts decreasing, it can be a warning to change our strategy.
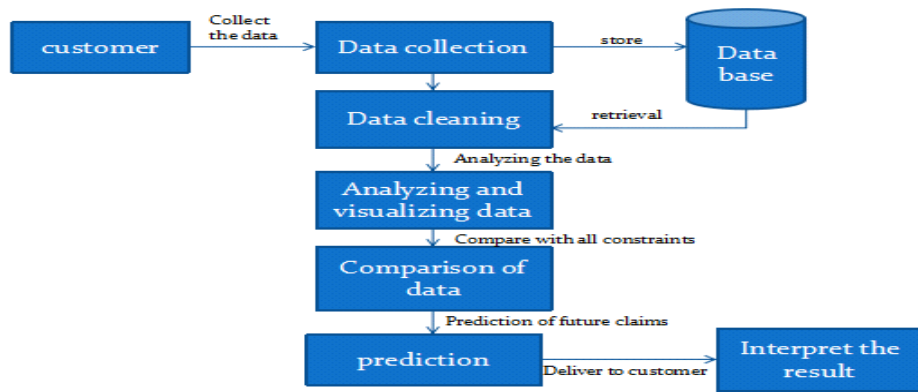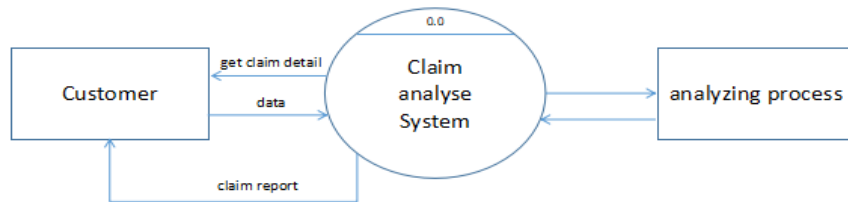
International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

Fig 1. Architecture Diagram

Fig 2. DFD level 0

Fig 3. DFD level 1

Fig 4. DFD level 2

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

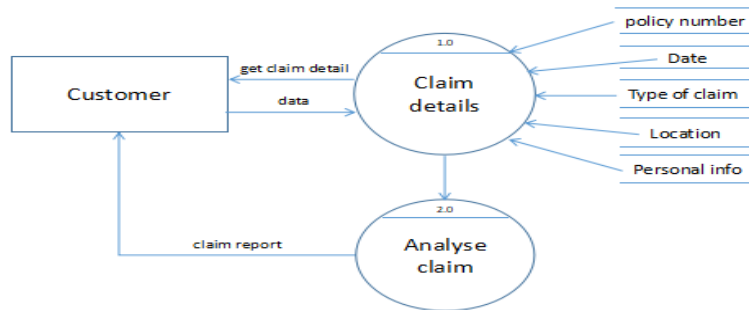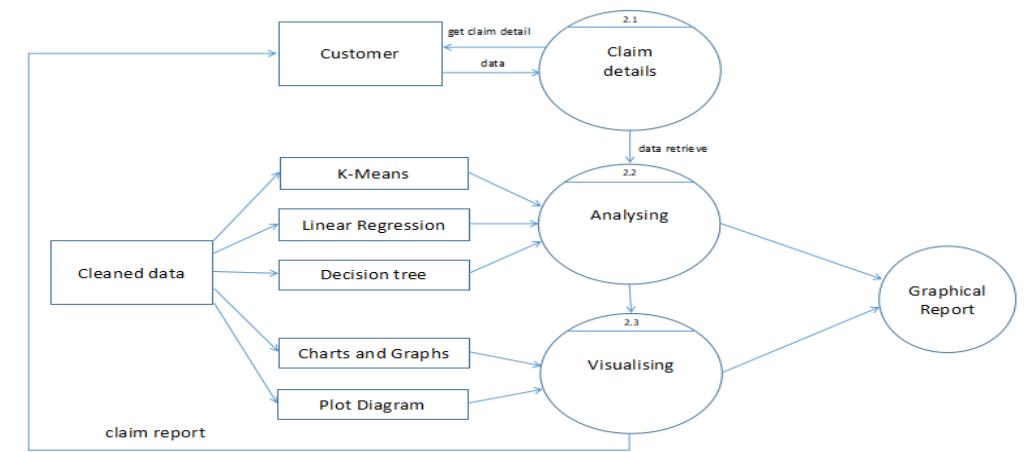## IV. CONCLUSION

The main aim of this paper is to provide the essential information to the insurance companies for increase their income and claim rate in a future. The key idea we applied here is the future prediction of data and automation. The automation we applied can save us from time, labor and much more. The simpler is now gone much simpler and the application can be managed by a few people.

The process we done above are collected data is transferred into visualized data and are compared and finally predicted. This will be very useful in vehicle insurance as the insurance customer can view the progress in detail such as subscription etc.

REFERENCES

[1]. Xia, B. S., & Gong, P. (2015). Review of business intelligence through data analysis. Benchmarking, 21(2), 300-311. doi:10.1108/BIJ-08-2012-0050

[2]. Judd, Charles and, McCleland, Gary (1989). Data Analysis. Harcourt Brace Jovanovich. ISBN 0-15-516765-0.

[3]. Clean Data in CRM: The Key to Generate Sales-Ready Leads and Boost Your Revenue Pool Retrieved 29th July, 2016

[4]. "Data Cleaning". Microsoft Research. Retrieved 26 October 2013.

[5]. Schutt, Rachel; O'Neil, Cathy (2013). Doing Data Science. O'Reilly Media. ISBN 978-1-449-35865-5.