IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, May 2023

Data Driven Analysis of Insurance Claims Using Machine Learning Algorithm

Dr. Velmurugan. K[1], Mr. K. Pazhanivel[2], Divyasree. R[3], Gowtham. E[4], Guruharan. S[5]
Professor, Department of Computer Science and Engineering[1,2]
Students, Department of Computer Science and Engineering[3,4,5]
Anjalai Ammal Mahalingam Engineering College, Thiruvarur, Tamil Nadu, India

Abstract: *The insurance industry is undergoing major changes with the integration of big data and artificial intelligence technologies. The design and research of an insurance survey claims system based on big data analysis aims to improve the efficiency and accuracy of insurance claims processing. The system uses image recognition, computer vision systems, language recognition, and other AI technologies to analyze case information and accelerate the speed of insurance claims settlement. The system also includes an intelligent customer service feature that uses AI algorithms such as language processing and big data statistical analysis to provide processing suggestions to policyholders. The system implements an individualized insurance service that collects, stores, and analyzes data on policyholders to create personalized insurance products and perform precise marketing. Big data analysis in the insurance personalized service primarily uses association rule analysis, classification and clustering analysis, and change and deviation analysis to improve the service. The one-click reporting function simplifies the reporting process for policyholders, allowing them to report a case from anywhere at any time. The intelligent claims processing feature separates liability in claims cases and deals with non-controversial cases through the use of AI, shortening the processing time and reducing manpower costs. The insurance survey and claim system has undergone five iterations under an agile development model and has achieved the goals of personalized insurance services, one-click reporting, intelligent claims processing, and intelligent customer service. The practical application results demonstrate that the system can improve the efficiency and accuracy of insurance claims processing while also providing policyholders with a more convenient and personalized experience. In conclusion, the design and research of an insurance survey claims system based on big data analysis has the potential to revolutionize the insurance industry and greatly benefit both insurance companies and policyholders.*

Keywords: Insurance Industry

I. INTRODUCTION

This project aims to predict insurance claim amounts using machine learning algorithms. The goal is to build a model that can accurately predict the amount of a claim based on various factors such as the age and gender of the policyholder, the type of insurance policy, the amount of coverage, and any past claims. The project involves gathering and cleaning data, selecting an appropriate algorithm, training the model, and evaluating its performance. The model can then be used to make predictions on new data, helping insurers better understand and manage their risk exposure. The success of the project will depend on the quality of the data, the chosen algorithm, and the ability of the model to accurately predict claim amounts

1.1 Insurance Claim Analysis

The process of analyzing insurance claims to learn more about how well an insurance company's business activities are performing is known as insurance claim analysis. Every day, policyholders submit a large number of claims to insurance firms. By examining these claims, insurers can spot patterns and trends that can guide their business decisions.Large amounts of data, including details about the policyholder, the type of claim, the amount sought, and the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, May 2023

claim's outcome, are usually collected and analyzed in the course of insurance claim analysis. In order to optimize their operations and make better choices, insurers can use this data by applying various analytical techniques, such as machine learning algorithms. Improvement of claims management procedures, the identification of high-risk clients, the detection of fake claims, and the prediction of future claims are all objectives of insurance claim analysis. In the end, insurance claim analysis can assist insurers in increasing profitability, decreasing expenses, and enhancing customer satisfaction. The insurance industry is crucial to modern society because it provides financial protection against unforeseen tragedies like accidents, natural calamities, and health issues. Yet another challenge faced by insurance firms is managing a sizable volume of claims, which can consume a lot of time and resources. To address this issue, insurance firms are using machine learning techniques to manage and examine insurance claims.

1.2 Machine Learning

A kind of artificial intelligence called machine learning enables computers to learn from data and enhance their performance over time. Insurers can spot patterns and trends in insurance claim data that may not be obvious to human analysts by utilizing machine learning algorithms. This can facilitate the claims process, improve the accuracy of fraud detection, and enable insurers to make more educated choices on the pricing and underwriting of policies.

This research aims to investigate the use of machine learning methods for insurance claim analysis. We will specifically look at how machine learning may be used to analyze various insurance claims, including auto, health, and property insurance. We will also go over the difficulties and possibilities of using machine learning to analyze insurance claims, including issues with data privacy, model interpretability, and the requirement for sizable and varied datasets.

This paper's overall goal is to present a thorough overview of machine learning's current state in insurance claim analysis and to highlight how it has the potential to revolutionize the sector.

II. RELATED WORKS

A Model for the Detection of Insurance Fraud, Geneva Papers on Risk and Insurance Theory

This article's goal is to create a model that will help insurance companies make decisions and ensure that they are better prepared to combat fraud. Based on the methodical exploitation of fraud signs, this tool.We first propose a procedure to isolate the indicators which are most significant in predicting the probability that a claim may be fraudulent. The Dionne-Belhadji study's data were subjected to the procedure (1996). We were able to see from the model that 23 of the 54 indicators used had a significant impact on predicting the likelihood of fraud. The accuracy and detection power of the model are also covered in our study. The reference point for this discussion is the detection rates attained by the adjusters who took part in the study.

Insurance Fraud and Optimal Claims Settlement Strategies

Because claimants can permanently misrepresent their damages by participating in expensive claims falsification, we look at the best claims settlement method for liability insurance. In this scenario, claims auditing is not a feasible deterrent to fraud, and the settlement strategy consists of an indemnification profile that ties the insurance payout to the alleged amount of loss. It is demonstrated that the ideal indemnity profile involves systematic underpayment of claims at the margin as a way to prevent exaggeration of losses, with the level of underpayment being constrained by anticipated litigation costs and probable bad-faith claims. The key testable implication of the theory is that the extent of underpayment should be greater for classes of claims for which loss exaggeration is easier.

Detecting insurance claims fraud using machine learning techniques

There are thousands of companies in the insurance industry globally. and collect premiums totaling more than $1 trillion each year. When a person or business submits a false insurance claim in an effort to get funds or benefits to which they are not legally entitled, this is known as insurance fraud. The established method for detecting fraud is focused on creating heuristics around fraud indicators. The most prevalent form of insurance fraud is auto fraud, which is accomplished by filing false accident claims. This research focuses on the use of machine learning to identify auto-vehicle fraud. Additionally, the performance will be compared using a confusion matrix computation. Calculating accuracy, precision, and recall can be made easier using this.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, May 2023

III. SYSTEM ARCHITECTURE

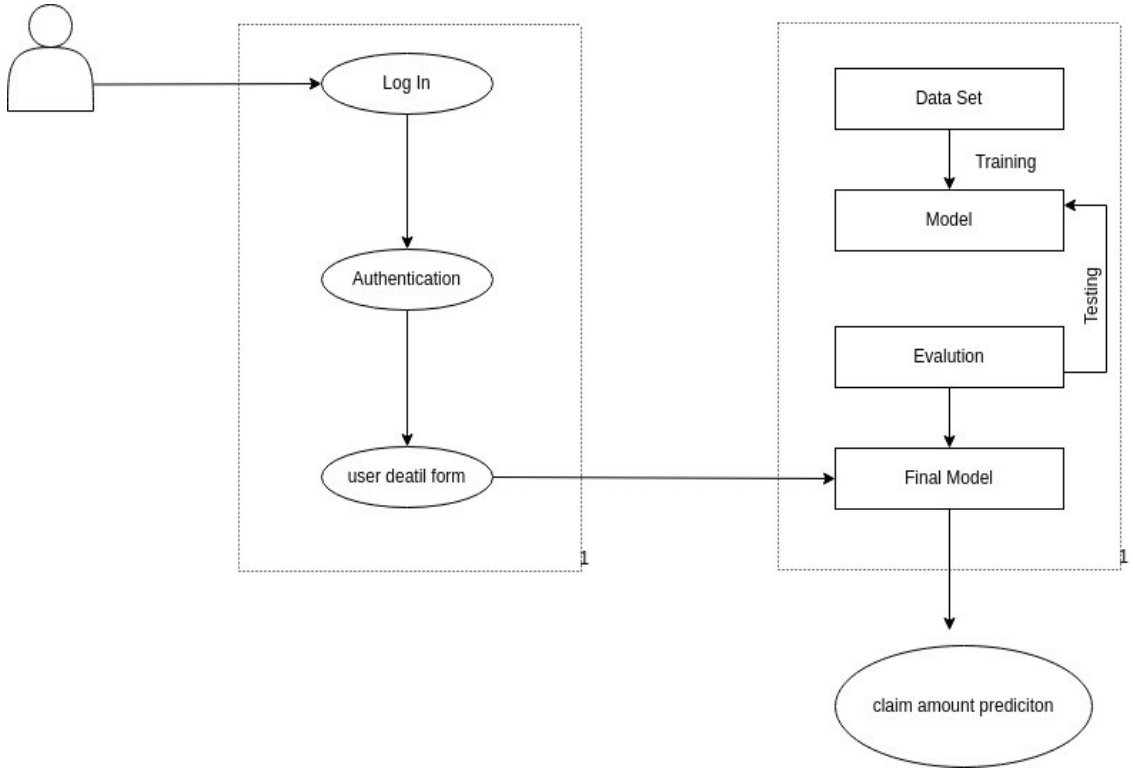

Fig. 1. System Architecture

- **Data Collection:** The first step in the study of insurance claims is to gather the pertinent data. Data from insurance policies, claim forms, damage reports, photographs, and other sources may be included in this. The information is then kept in a database for future use.
- **Data Preprocessing:** The preprocessed data is subsequently put into a format that is appropriate for machine learning techniques. To get the data ready for analysis, this may involve procedures like feature engineering, normalization, and data cleaning.
- **Machine Learning Models:** The machine learning models that are trained on the preprocessed data are the brains of the system. These models might use unsupervised learning techniques like clustering and anomaly detection in addition to supervised learning techniques like decision trees, neural networks, and support vector machines.
- **Prediction:** The machine learning models can be used to anticipate insurance claims once they have been trained. For instance, based on the reported damages, they can forecast the likelihood that a claim will be fraudulent or calculate the claim's cost.
- **User Interface:** Via a user interface, the user is shown the predictions provided by the machine learning models. Dashboards, reports, and other visualizations may be included in this interface to assist insurance professionals in making choices..
- **Feedback Loop:** Finally, a feedback loop may be incorporated into the system to enhance the machine learning models' accuracy over time. To enhance performance, this may entail gathering more data, retraining the models, and modifying the algorithms.

IV. MACHINE LEARNING ALGORITHM

4.1 Support Vector Machine

Popular supervised machine learning algorithms for categorization and regression include Support Vector Machine (SVM). It operates by locating a hyperplane that divides the data into various classes in a high-dimensional area.SVM

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, May 2023

seeks to identify the hyperplane that maximizes the margin between the two classes in classification assignments. The distance between the nearest data points from each class and the hyperplane is referred to as the margin. Support vectors, from which the term Support Vector Machine derives, are the data points that are most closely related to the hyperplane. By utilizing various kernel functions, SVM can manage both linear and nonlinear classification tasks. The sigmoid, linear, quadratic, and radial basis function (RBF) are some common kernel functions. In regression tasks, SVM looks for a hyperplane that has a margin of tolerance and matches the data as closely as possible. The epsilon-insensitive loss function is this.In comparison to other machine learning algorithms, SVM offers a number of benefits, including the ability to manage high-dimensional data, effectiveness with small datasets, and robustness to outliers. SVM, however, can be computationally demanding and necessitates cautious hyperparameter selection.

4.2 Linear Regression Algorithm

The statistical method of linear regression is used to analyze a dependent variable (commonly denoted by "y") and one or more independent factors. (often denoted as "x").. Given some input factors, it is a common algorithm for predicting a numerical result. Finding the best-fit line that encapsulates the connection between the dependent variable and the independent variable is the aim of linear regression.(s). An expression of the form: represe

$$y = b0 + b1x1 + b2x2 + ... + bn*xn$$

where y is the dependent variable, x1, x2, ..., xn are the independent variables, b0 is the intercept, and b1, b2, ..., bn are the coefficients. The least squares estimation technique, which minimizes the sum of the squared differences between the real and predicted values of the dependent variable, is used by the linear regression algorithm to determine the coefficients. The coefficients can be used to forecast the value of the dependent variable given new values for the independent variable once they have been discovered.(s).

4.3 Gradient Boosting Algorithm

A potent machine learning approach called gradient boosting is employed for both classification and regression problems. It operates by repeatedly training weak models, with each new model fixing the mistakes caused by the prior model. Until the desired level of accuracy is attained or a predetermined number of models have been trained, this process is repeated. As an ensemble learning technique, gradient boosting combines several weak models to produce a strong model. It measures the discrepancy between the expected and actual values using a loss function, and then trains a new model to reduce the loss. To attain optimum performance, the technique requires thorough hyperparameter adjustment, which is computationally expensive. Due to its increasing popularity, gradient boosting.

4.4 Determining the best model based on MAE score

Based on the MAE score, you have determined that Gradient Boosting is the best model for your machine learning project. Gradient Boosting is a powerful ensemble learning method that combines multiple weak models to create a strong model. It works by iteratively training a series of weak models, with each subsequent model learning from the errors made by the previous models. In other words, each subsequent model is built to correct the mistakes of the previous models. This process continues until a desired level of accuracy is achieved or a specified number of models have been trained. The algorithm is computationally expensive and requires careful tuning of hyperparameters to achieve optimal performance. However, Gradient Boosting has become increasingly popular in recent years due to its ability to handle complex datasets and achieve high accuracy in predictive tasks. Overall, it sounds like you have chosen a powerful and effective model for my project.

V. CONCLUSION

In conclusion, insurance claim analysis is a valuable tool for insurance companies, providing them with insights into the efficiency and effectiveness of the insurance claims process and helping them to reduce costs, improve the customer experience, and make informed decisions about their products and overall business strategy. Addressing the challenges faced by the insurance claims management process is essential for improving the overall insurance claims experience and ensuring long term-success for insurance companies.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, May 2023

REFERENCES

[1]. Belhadji, E., G. Dionne, and F. Tarkhani, ―A Model for the Detection of Insurance Fraud, Geneva Papers on Risk and Insurance Theory.

[2]. Crocker, K. J., and S. Tennyson, Insurance Fraud and Optimal Claims Settlement Strategies: An Empirical Investigation of Liability Insurance Settlements| The Journal of Law and Economics.

[3]. KajianMuller, ―The Identification of Insurance Fraud – an Empirical Analysis Working papers on Risk Management and Insurance| no: 137, June 2013.

[4]. Tang Jincheng, Liu Lu. Applied Research on "AI + Insurance" Model in the Era of Insurance Science and Technology [J].

[5]. Uditha Balasooriya and Chan-Kee Low ,(2008), Modeling Insurance Claims With Extreme Observations: Transformed Kernel Density and Generalized Lambda Distribution