iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.55252

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

A Survey on HealthCare Claim Prediction

Amogh Naik[1], Kush Parihar[2]
MIT ADT University, Loni Kalbhor

Abstract: ***The life insurance industry has experienced a significant increase in claims in recent years. To cope with the large volume of claims, insurance companies are turning to machine learning models for predicting life insurance claims. This paper presents a machine-learning model for predicting life insurance claims based on various demographic and health-related features. The proposed model will be evaluated on a real dataset and will achieve a high accuracy, precision, recall, and F1 score. The results will show that the machine learning model can effectively predict life insurance claims, which can improve the efficiency of claims processing and ultimately benefit insurance companies and policyholders.***
Keywords: Life insurance, machine learning, prediction, claims processing, demographic characteristics and health-related characteristics.

I. INTRODUCTION

Insurance claim price prediction is about estimating the amount of money an insurer is likely to pay for a particular claim. In the insurance industry, accurate price claim prediction is crucial for insurers to assess the risk associated with different policies, price them appropriately and allocate resources efficiently. By analyzing vast amount of data, including historical claims data, policyholder information and external factors, insurers can make more informed decisions about valuation, pricing and resource allocation. The use of machine learning algorithms will enable insurers to identify patterns and trends in claiming data, make accurate predictions about future claims prices and continuously refine their pricing models.[1]

Accurately predicting claims prices faces several challenges, including the high variability of claim data, the complexity of the insurance industry and the need for transparent and fair pricing models. Despite these challenges, recent advancements in machine learning and data analytics have led to significant improvements in the accuracy and efficiency of insurance claim price prediction.

In this paper, we will be using various machine-learning algorithms in combination to give a fast and accurate result. By staying ahead of the curve and embracing new technologies, insurers can better manage risk, improve customer satisfaction and ensure that policyholders receive fair and accurate compensation for their claims.[2]

II. LITERATURE SURVEY

A. Existing Work

1) Medical Insurance Prediction

The paper presents an attempt to scrutinize a phenomenon of medical insurance price prediction. Insurance is one of the policies that either decreases or removes loss costs incurred by various risks. Using various types of machine learning algorithms such as – Linear tree regression, Gradient Boosting Regression, and Decision Tree Regression, a model is created which takes a combination of their results based on the dataset acquired from Kaggle.

2) Model For Medical Insurance

The paper offers a synthesis of claim prediction which is one of the most important operations in the field of the insurance industry. A proper insurance claim for each risk represented by the customer differs widely from one customer to another. The datasets have missing values which can affect the result of the prediction. The proposed model consists of machine-learning algorithms such as decisions trees, Naïve Bayes, Artificial Neural Network (ANN) and Gradient Boosting Regression.

3) Data Mining Techniques for Insurance Claim

This thesis investigates how data mining algorithms can be used to predict Insurance claim payments based on the characteristics of the insured customer. The algorithms are tested on real data provided by the organizer of the competition. The data present several challenges such as high dimensionality, heterogeneity and missing variables. The model uses k-means techniques for the model's training and accurate predictions. The model uses algorithms such as SVM, Decision Trees, RandomForest.

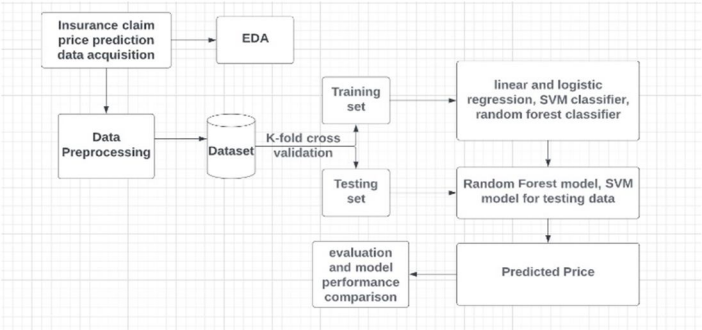
©IJRASET: All Rights are Reserved

III. PROBLEM STATEMENT

Traditional life insurance claims processing methods involve a lot of paperwork, manual review, and human error, which may lead to delays in claims processing and policyholder dissatisfaction. To overcome these challenges, insurance companies are using machine learning models to predict life insurance claims.[3] However, there is a need for a more accurate and efficient machine learning model that can predict the probability of life insurance claims based on various demographic and health-related characteristics.

IV. PROPOSED WORK

A. Arhcitechture of the Model



B. Data Collection

The first step in our methodology is to collect a dataset of demographic and health characteristics of policyholders and their beneficiaries. The dataset can be collected from various sources such as insurance companies, government agencies, and health organizations. The dataset may include characteristics such as age, gender, medical history, family history, smoking habits, occupation, and income.

C. Preprocessing the Data

The next step is to preprocess the dataset to prepare it for the machine learning model. This includes handling missing values, coding categorical variables, scaling numeric variables, and removing outliers.

D. Feature Selection

The third step is to select the most relevant features from the dataset. For this purpose, various feature selection techniques can be used, such as correlation analysis, mutual information, and feature importance assessment. The selected features are then used as inputs to the machine learning model.

E. Model Training

The fourth step is to train a machine learning model using the preprocessed dataset. The proposed model will use a combination of supervised learning algorithms, such as logistic regression, decision trees, and random forest, to predict accurate results as well as loss event. We will be testing our dataset on multiple algorithms and choose the best machine-learning algorithm amongst them.

F. Evaluation of the Model

The fifth step is to evaluate the performance of the machine learning model. This is will be done using various performance metrics such as accuracy, precision, recognition and F1 score. A confusion matrix will also be created to visualize the number of true positives, true negatives, false positives, and false negatives.

G. Deployment of the Model

The final step is to deploy the machine learning model in a real-world environment. The model will be integrated into the insurance company's claims processing system and will be able to automatically predict the probability of a claim based on the input features. The model will also be used to detect fraudulent claims and take appropriate action to prevent them.

V. CONCLUSION

In summary, insurance claim price prediction is an important task in the insurance industry that involves estimating the amount of money an insurer is likely to pay for a particular claim. Accurate prediction of claims prices is essential for insurers to manage risk, price policies appropriately, and allocate resources effectively.[3]

Machine learning algorithms are becoming increasingly popular for analyzing claims data and making predictions about future claims payments. These algorithms can learn from historical data, identify patterns and make accurate predictions, which helps insurers make informed decisions.

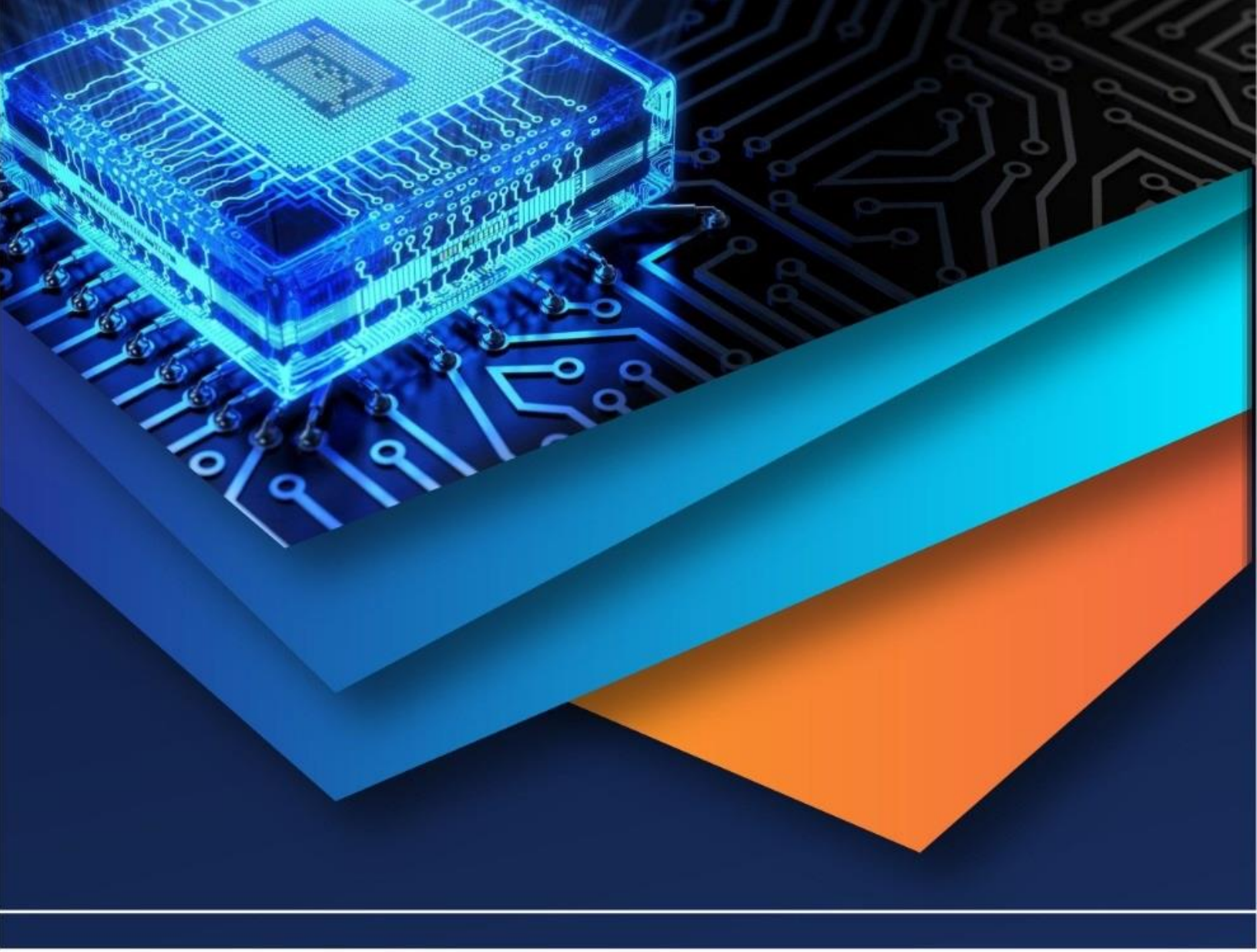
However, accurately predicting claims prices is nothing without its problems. These include challenges to the high variability of claims data, the complexity of the insurance industry, and the need for transparent and fair pricing models. To overcome these challenges, insurers must continue to invest in advanced analytics and data science technologies to improve the accuracy and efficiency of their loss claim price-prediction models.[4]

Overall, insurance premium forecasting is an important task that enables insurers to better manage risk, improve customer satisfaction and ensure that policyholders receive fair and accurate compensation for their claims. By using advanced analytics and machine learning, insurers can further improve their claims prediction capabilities and stay ahead of the competition in the dynamic and rapidly evolving insurance industry.[5]

REFERENCES

[1] Banerjee, R., & Sardar, S. K. (2019). Machine learning-based insurance claim prediction: A case study. Journal of Business Research, 102, 308-318.
[2] Bhattacharyya, S., & Dey, N. (2018). Life insurance claim prediction using machine learning algorithms. Procedia Computer Science, 132, 764-773.
[3] Cho, K., & Yoo, C. (2019). Application of machine learning techniques for life insurance claim prediction. Journal of Intelligent & Fuzzy Systems, 36(1), 563-571.
[4] Gao, X., & Chen, Q. (2021). Prediction of life insurance claim propensity based on machine learning algorithms. Journal of Computational Science, 48, 101314.
[5] Li, K., Li, Y., Li, Y., Li, C., & Li, X. (2021). Life insurance claim prediction using machine learning techniques. Symmetry, 13(6), 958.
[6] Zhang, X., Liu, J., & Feng, J. (2021). Life insurance claim prediction using machine learning methods. Mathematics, 9(5), 464.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)