IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# Enabling The Power Of ML In Insurance Pricing

MS. POONAM RAKIBE[1], PARAG KSHIRSAGAR[2], VEDANT MODAK[3], HARSH MAKODE[4] AND SAMRUDDHI BHOSALE[5]

***Abstract:*** Innovative platform Ensure revolutionizes insurance valuation by offering predictive analytics for a range of commodities. To reliably predict future valuations, our website uses cutting-edge machine learning algorithms, such as time series analysis and regression. Our platform takes in a wide variety of data, processes it using our unique algorithms, and produces accurate forecasts of future values. Users can price their goods more properly, evaluate their options well, and eventually increase profitability because of this predictive capability. To maintain the highest level of accuracy, our machine learning models are regularly updated with fresh data and trained on old data. Regression models allow us to quantify the relationship whereas time series analysis helps us analyze trends and patterns over time. It’s a comprehensive solution for insurance companies looking to stay ahead in the competitive market. Join us in redefining insurance valuation and embarking on a journey toward data- driven decision-making.

***Index Terms*** - Predictive Analytics, Machine Learning, Regression, Time- Series Analysis, Valuation and Pricing.

## 1.INTRODUCTION

We recognize the significance of quick and accurate insurance claim assessments in a world where uncertainties are a part of daily life. The way insurance claims are valued will be revolutionized by our platform, which will make it quicker, fairer, and more dependable than before the world of insurance, it is crucial to make sure that after a loss, policyholders receive a just and equitable settlement. Traditional claim valuation techniques sometimes include labor-intensive manual inspections and protracted processing timeframes, which aggravate policyholders and drive up operating expenses for insurance providers. Our machine learning-powered platform fills this gap, bringing efficiency and innovation to the insurancesector.

Important Elements of Our Platform

a) Machine learning algorithms to provide the most accurate value, our platform uses cutting-edge Machine Learningalgorithms to examine a variety of data points, from property damage to medical records.

b) Quickness and effectiveness: Claim valuations are finished utilizing our automated approach in a tenth of the time it takes with conventional techniques.

c) Accuracy and Reliability: Machine learning eliminates humanbiases and assures reliable, unbiased claim evaluations, fosteringjustice and trust in the insurance sector.

## 2.MOTIVATION

The insurance industry has a huge market capitalization in India.This evergreen industry has been traditionally dominated by large-cap Insurance companies. Associated Surveyors and Valuers are supposed to evaluate the claim and assess the actualloss against the claim that is being made. Although this industryhas some problems which have to be addressed.

Customers often face several challenges when claiming insurance: Documentation and Proof, Delays in Processing, Disputes and Denials, and Complexity of the Claim. Insurance surveyors, on the other hand, face their own set of challenges during claim valuations:

a) Accuracy of Valuations: Surveyors must accurately assess theextent of damage or loss to determine the claim amount. This requires expertise in various fields such as automotive, construction, or healthcare, depending on the type of insurance.

b) Coordination: Coordinating with multiple parties, including the claimant, witnesses, repair shops, and healthcare providers, can be complex and time-consuming.

c) Fraud Detection: Surveyors must be vigilant for potential fraudulent claims, which can be difficult to detect but can significantly impact the insurance company's finances.

d) Legal and Ethical Considerations: Surveyors must adhere to legal and ethical guidelines while conducting valuations to ensure fairness and compliance with industry regulations.

Both customers and insurance surveyors should benefit from streamlined and transparent processes, improved communication, and the use of technology to expedite and simplify the claims and valuation processes. The project aims tobridge the existing gaps in accurate valuation and lossassessment. By doing so, it strives to create a safer, more transparent, and secure marketplace for consumers and businesses alike.

## 3.REVIEW OF LITERATURE

**Saraswat, B. K., Singhal, A., Agarwal, S., & Singh, A. (2023).Insurance Claim Analysis Using Traditional Machine Learning Algorithms. In 2023 International Conference on Disruptive Technologies (ICDT) (pp. 623-628). IEEE.** This paper proposes a machine learning model to predict health insurance claims based on employee information. It compares various classification algorithms and evaluates their performanceusing accuracy, precision, recall, and F1-score.[1]

**Burri, R. D., Burri, R., Bojja, R. R., & Buruga, S. R. (2019). Insurance Claim Analysis Using Machine Learning Algorithms. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(6S4), 1-5**. This paperexplores the use of machine learning algorithms to analyzeinsurance data sets and provide insights for decision-making. It covers different types of insurance, such as health, life, property,and vehicle insurance, and discusses the potential applications ofmachine learning in each domain.[2]

**Nithya, B., & Ilango, V. (2017). Predictive analytics in healthcare using machine learning tools and techniques. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 492-499). IEEE.** This paper reviews various machine learning tools and techniques that can be used for predictive analytics in health care. It also discusses some of the applications and challenges of machine learning in healthcare domains, such as disease diagnosis, treatmentselection, and patient monitoring[3]

**Baudry, M., & Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance.**

**Applied Stochastic Models in Business and Industry, 35(6), 2019-2034.** This paper presents a random forest method for estimating outstanding liabilities using individual claims data and covariates. It compares its performance with some benchmark methods using a real data set of motor third-party liability claims.[7]

**Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2020). Boosting insights in insurance tariff plans with tree- based machine learning methods. North American Actuarial Journal, 24(4), 531-561** This paper applies tree-based machine learning methods, such as regression trees, random forests, and gradient boosting, to model insurance tariff plans and provide insights into the risk factors and premium structures. It compares the performance and interpretability of these methods with generalized linear models using a real data set of automobile insurance claims.[8]

**Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data— XGBoost versus logistic regression. Risks, 7(3), 70.** This paper compares the performance of two machine learning methods, XGBoost and logistic regression, for predicting motor insurance claims based on telematics data. It uses a real data set of driving behavior and claims frequency from a Spanish insurance company, and evaluates the methods in terms of accuracy, sensitivity, specificity, and ROC curve. [15]

**Rajitha, C., & Sakthivel, K. M. (2017). Artificial intelligence for estimation of future claim frequency in non-life insurance. Global Journal of Pure and Applied Sciences, 13, 10-17** This paper develops a procedure for predicting the future claim frequency of insurance portfolios in general insurance using artificial neural networks with Bayesian credibility inputs. It claims that its method can produce exact and reliable predictions of future claim frequencies, unlike conventional statistical methods.

**Albrecher, H., Bommier, A., Filipović, D., Koch-Medina, P., Loisel, S., & Schmeiser, H. (2019). Insurance: Models, digitalization, and data science. European Actuarial Journal, 9(3-4), 349-360**. This paper discusses the current and future challenges and opportunities for the insurance industry in the context of models, digitalization, and data science. It covers various topics, such as risk management, regulation, customer behaviour, product design, and social responsibility.[4]

Table No-1

| Paper No. | Model used | Accuracy |
|---|---|---|
| [1] | Decision Tree | 81.23% |
| [2] | Naïve Bayes | 81.60% |
| [3] | Logistic Regression | 75.60% |
| [4] | Random Forest | 79.42% |
| [5] | Regression Tree | 83.66% |
| [6] | XGBoost | 81.00% |
| [7] | ANN | 82.40% |

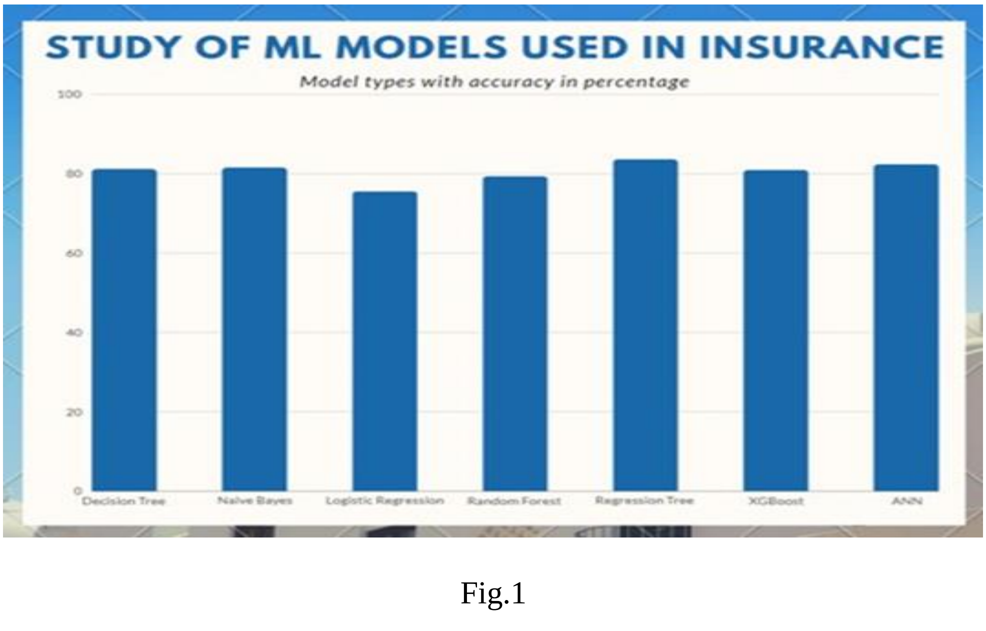


Fig.1

Table No-2

| Paper No. | Model Used | Data Samples | Features |
|---|---|---|---|
| [1] | Decision Tree | 1000 | 10 |
| [2] | Naïve Bayes | 1240 | 8 |
| [3] | Logistic Regression | 768 | 5 |
| [4] | Random Forest | 632 | 7 |
| [5] | Regression Tree | 982 | 6 |
| [6] | XGBoost | 1937 | 9 |
| [7] | ANN | 1000 | 4 |

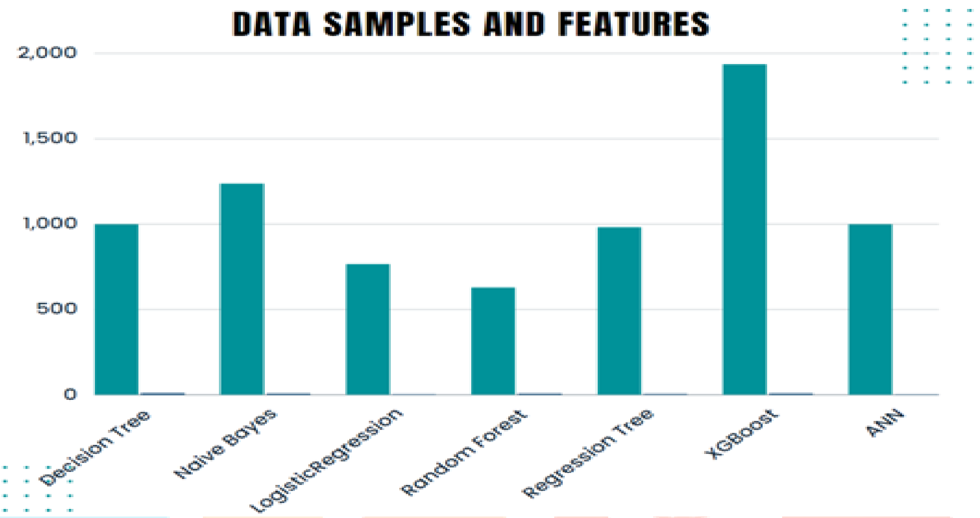


Fig.2

Table No-3

| Paper No. | Model used | Advantages |
|---|---|---|
| [1] | Decision Tree | Performs well with large datasets |
| [2] | Naïve Bayes | Gives reasonable precision |
| [3] | Logistic Regression | Performs better with small datasets |
| [4] | Random Forest | Generalizes well to new data |
| [5] | Regression Tree | Determines the best decision |
| [6] | XGBoost | Computationally efficient |
| [7] | ANN | Fault tolerance and parallel processing |

Table No-4

| Paper No. | Model used | Drawbacks |
|---|---|---|
| [1] | Decision Tree | Lacks dealing with relationships of variables |
| [2] | Naïve Bayes | Lack of data and fewer features |
| [3] | Logistic Regression | Sector specific (Healthcare) model |
| [4] | Random Forest | Target variables are censored |
| [5] | Regression Tree | Needs Gradient Boosting |
| [6] | XGBoost | Sector specific (Automobile) model |
| [7] | ANN | Issues with linearity and normality |

## 4. METHODOLOGIES

**Insurance Valuation**: Accurate assessment of individual insurance claims using standard industry practices.
**Data Integration**: Seamless integration with internal and external data sources for comprehensive analysis.
**Actuarial Modeling**: Advanced modeling for predicting commodity prices using time series analysis and decision tree.
**Reporting and Visualization**: Customizable reports and dashboards with clear graphics.
User Input consists of the type of commodity, the year in which the commodity was bought, and the cost at which the commoditywas bought. The Calendar Index is a Standard Index data used by insurance surveyors to evaluate insurance claims. The valuation model refers this data of this commodity data and computes the valuation against the user inputs. The Arima model computes a time series analysis of the commodity and predicts the expected index price of the commodity in the upcoming years.. It is used to forecast a time series using the series past values. The ARIMA model uses a combination of Auto- Regression and Moving averages to predict future data points. Although the primary requirement of the ARIMA model is the data has to be stationary.

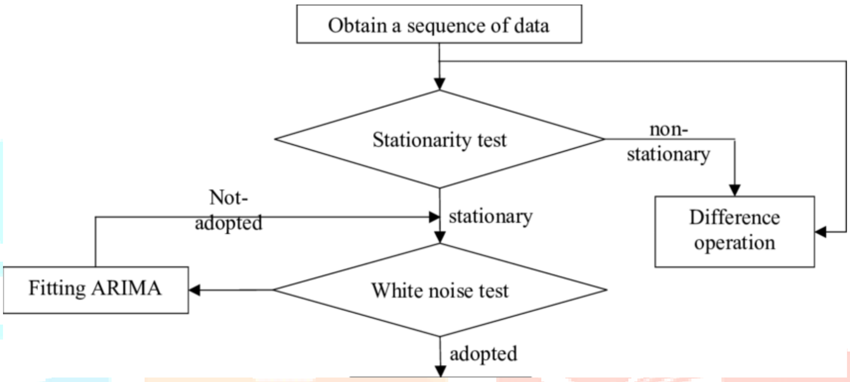


Fig.3

The data used has 6 features. Raw material cost, currency exchange rate, global demand, labour cost, interest rates, and consumer price index. All of these variables are estimated from the time series model. The target variable 0 or 1 determines whether the insurance will be claimed or not. The Decision Tree[1] model computes the input to give either of these outputs.Insurance valuers can determine the premium that is supposed tobe charged to the customer based on this result.

## 5.PROPOSED TECHNOLOGY

The Insurance Valuation and Analytics Project aims to develop a comprehensive software solution that empowers insurancecompanies, surveyors as well as clients to efficiently and accurately assess the value of their claims, leveraging advanced data analytics and actuarial methodologies.

**Insurance Valuation**: Accurate assessment of individualinsurance claims
**Data Integration**: Seamless integration with internal andexternal data sources for comprehensive analysis.
**Actuarial Modeling**: Advanced modeling for predictingcommodity prices using time series analysis.
**Reporting and Visualization**: Customizable reports anddashboards with clear graphics.
**Regulatory Compliance**: Tools to ensure compliance withindustry regulations.
**End Users**: These are individuals who use the system for evaluating their claims. They may include professionals in the insurance industry, researchers, etc.
**Administrators**: Administrators have control over system maintenance, dataset updates, and model retraining. They need adeeper technical understanding of the system.
**Developers**: Developers may interact with the system to integrate it into other applications or services.
Our system runs on web platforms. Ensure compatibility with common web browsers. It will also be cloud-ready, as we mightleverage cloud resources for model training and hosting.
regression of the excess returns on the market risk premium are excluded. And Shares are grouped by alphabetic order into group of 30 individual securities (Roll and Ross, 1980).

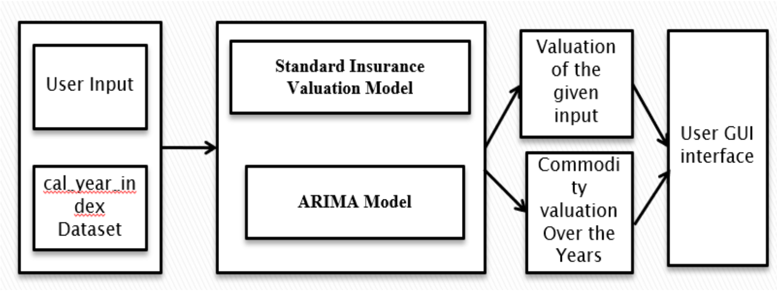


Fig.4

## 6.RESULTS

The final outcome of the project intends to introduce a fully operational web application tailored specifically for insurance surveyors and clients. This application will feature an intuitive login interface for entering user credentials such as name and phone number, thereby facilitating a smooth login or signup process. Upon successful authentication, users will gain access to a user-friendly homepage offering multiple functionalities. This application includes having estimates of commodity indexes in the near future, calculating valuations of claims, and a data-driven decision-making model to generalize the premium categories for the clients implementing principles from Human- Computer Interaction (HCI). This enhancement targets granting insurance surveyors, independent and seamless interaction within the digital realm using advanced data analytics, thereby fostering a more inclusive and accommodating web experience.

## 7.FUTURE SCOPE

Reaching the ideal solution for the problems in insurance pricingand valuation can be challenging due to various factors, but it isnot insurmountable. The difficulty of reaching an ideal solution depends on several factors:

**Complexity of the Insurance Industry**: The insurance industryis inherently complex, with various types of policies, risks, and regulations. This complexity can make it challenging to developstandardized solutions that work across the board.

**Data Quality and Availability**: Accurate and comprehensive data is crucial for pricing and valuation models. Ensuring data quality and availability, especially for historical claims data, canbe a significant challenge.

**Changing Regulatory Landscape:** Insurance regulations can vary by region and change over time. Compliance with these regulations is essential, and keeping up with regulatory changescan be complex.

**Technological Advancements:** Keeping pace with technological advancements in data analytics, machine learning,and AI is crucial for improving pricing and valuation processes.Integrating new technologies can be challenging for some organizations.

**Fraud Detection:** Detecting insurance fraud is a constant challenge. Fraudsters are continually adapting their tactics, making it difficult to develop foolproof fraud detection Customer Expectations: Meeting customer expectations for quick and fair claim settlements is essential for customer satisfaction. Balancing these expectations with the need for accurate valuations can be a delicate task.

**Resource Constraints**: Smaller insurance companies may have limited resources to invest in advanced technology and data analytics, making it more challenging for them to reach ideal solutions.

**Market Competition:** In a competitive market, insurance companies need to strike a balance between offering competitivepremiums and maintaining profitability, which can be a difficulttask.

While reaching the ideal solution may be challenging, it is not impossible. Many insurance companies are actively investing inresearch, technology, and process improvements to address thesechallenges. Collaborative efforts within the industry, advancements in data analytics and AI, and adherence to best practices can contribute to the development of more effective pricing and valuation solutions over time. Additionally, regulatory bodies and industry associations often provide guidelines and support to help insurers navigate these challenges while maintaining compliance and fairness in their operations

## 8.CONCLUSION

In conclusion, the decision tree-based machine learning model has demonstrated superior performance compared to other models. The availability of a larger dataset and a wider range of features has significantly contributed to this outcome. The model's accuracy 85.67%, which surpasses that of its counterparts, underscores its effectiveness. This suggests that decision tree algorithms, when paired with ample data and diverse features, can yield highly accurate predictions, thereby serving as a robust tool in the field of machine learning. Future work could explore further optimization techniques to enhance the model's performance even more by using neural networks. The Decision tree model used is computationally more intensive. Also, the data used has more samples with an appropriate number of features which increases its accuracy.


## ACKNOWLEDGMENT

We would like to extend our heartfelt gratitude to our project guide Prof. Ms. Poonam Rakibe, whose guidance and technical expertise have been instrumental in the successful completion of this research. Her profound knowledge and unwavering support have been a constant source of inspiration for us. We are also immensely grateful to the Head of the Department Dr. Prof. Mrs.Sarita Deshpande, whose leadership and vision have significantly contributed to our work. Her continuous encouragement and faith in our abilities have been pivotal in shaping this research. Their collective wisdom and support have been invaluable to us, and we thank them for their trust and guidance.


## REFERENCES


[1] B. K. Saraswat, A. Singhal, S. Agarwal and A. Singh, "Insurance Claim Analysis Using Traditional Machine Learning Algorithms," 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 623-628, doi: 10.1109/ICDT57929.2023

[2] Rama Devi Burri, Ram Burri, Ramesh Reddy Bojja and Srinivasa Rao Buruga, "Insurance Claim Analysis Using Machine Learning Algorithms", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 6S4, [online] Available: https://www.ijitee.org/wp-content/uploads/papers/v8i6s4/F11180486S419.pdf, ISSN 2278-3075

[3] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 492-499, June 2017, [online] Available: https://ieeexplore.ieee.org/document/825077.

[4] Albrecher, Hansjörg, Antoine Bommier, Damir Filipovi´c, Pablo Koch-Medina, Stéphane Loisel, and Hato Schmeiser. 2019. Insurance: Models, digitalization, and data science. European Actuarial Journal 9: 349–60.

[5] Ayuso, Mercedes, Montserrat Guillén, and Ana María Pérez- Marín. 2014. Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. Accident Analysis & Prevention 73: 125–31.ChengbinJin, ShengzheLi, Trung Dung Do, Hakil Kim, ”Real-Time Human Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras”, Information and Communication Engineering, Inha University, Incheon, Korea

[6] Barry, Laurence, and Arthur Charpentier. 2020. Personalization as a promise: Can big data change the practice of insurance? Big Data & Society 7. E. Mohammadi, Q.M. Jonathan Wu, M. Saif, ”Human Activity Recognition Using anEnsemble of Support Vector Machines”, 2016 IEEE International Conference on HighPerformance Computing Simu- lation (HPCS) , July 2016

[7] Baudry, Maximilien, and Christian Y. Robert. 2019. A machinelearning approach for individual claims reserving in insurance.
Applied Stochastic Models in Business and Industry

[8] Henckaerts, Roel, Marie-Pier Côté, Katrien Antonio, and Roel Verbelen. 2020. Boosting insights in insurance tariff plans with tree-based machine learning methods. North American Actuarial Journal 1–31.C.

[9] Bothwell, Peter T., Mary Jo Kannon, Benjamin Avanzi, Joseph Marino Izzo, Stephen A. Knobloch, Raymond S. Nichols, James L.Norris, Ying Pan, Dimitri Semenovich, Tracy A. Spadola, and et al.2016. Data & Technology Working Party Report. Technical Report.
Arlington: Casualty Actuarial Society

[10] Boucher, Jean-Philippe, Steven Côté, and Montserrat Guillen.2017. Exposure as duration and distance in telematics motor insurance using generalized additive models. Risks 5: 54.Z.

[11] Castignani, German, Thierry Derrmann, Raphaël Frank, and Thomas Engel. 2015. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent Transportation Systems Magazine 7: 91–102

[12] Chen, Kuan-Ting, and Huei-Yen Winnie Chen. 2019. Drivingstyle clustering using naturalistic driving data. Transportation Research Record 2673: 176–88.

[13] Friedman, Jerome H., Bogdan E. Popescu. 2008. Predictive learning via rule ensembles. The Annals of Applied Statistics 2: 916–54.

[14] Gabrielli, Andrea. 2020b. A neural network boosted double overdispersed Poisson claims reserving model. ASTIN Bulletin: The Journal of the IAA 50: 25–60.

[15] Pesantez-Narvaez, Jessica, Montserrat Guillen, and Manuela Alcañiz. 2019. Predicting motor insurance claims using telematicsdata—XGBoost versus logistic regression. Risks7:70.