International Journal of Recent Innovations in Academic Research

E-ISSN: 2635-3040; P-ISSN: 2659-1561
Homepage: https://www.ijriar.com/
Volume-9, Issue-3, July-September-2025: 55-66

Research Article

Predictive Modeling for Property Insurance Premium Estimation Using Machine Learning Algorithms

[*a]Ram Mohan Polam, [b]Bhavana Kamarthapu, [c]Ajay Babu Kakani, [d]Sri Krishna Kireeti Nandiraju, [e]Sandeep Kumar Chundru and [f]Srikanth Reddy Vangala

[a&d]University of Illinois at Springfield, The United States of America
[b]Fairleigh Dickinson University, The United States of America
[c]Wright State University, The United States of America
[e]University of Central Missouri, The United States of America
[f]University of Bridgeport, The United States of America
*Corresponding Author Email: ramreddy.polam@gmail.com

Received: June 17, 2025 **Accepted:** July 06, 2025 **Published:** July 13, 2025

Abstract
Accurate premium estimation is a cornerstone of effective property insurance, ensuring both financial stability for insurers and fair pricing for policyholders. Traditional actuarial models, though foundational, often fall short when dealing with high-dimensional, nonlinear insurance data. This study presents a machine learning-based approach for accurately predicting property insurance premiums using the Extreme Gradient Boosting (XGBoost) model. Leveraging a comprehensive home insurance dataset with 42 attributes, the proposed methodology involves data preprocessing, feature engineering, training the model, and evaluating its performance using conventional regression measures. The XGBoost model demonstrated superior predictive performance, achieving RMSE = 52.99, MSE = 28.07, and MAE = 30.64 is the calculated value for the mean squared error. These results indicate its strong capability in modeling complex relationships and estimating premiums with high accuracy. Comparative analysis against the Linear Regression model further highlights the effectiveness of the proposed XGBoost model, making it a robust and scalable solution for real-world property insurance pricing systems.
Keywords: Property Insurance, Insurance Premium Estimation, Predictive Modeling, Machine Learning, XGBoost.

I. Introduction

The purposes of property insurance as a risk manager involves the financial transfer of risk associated with losses or damages to real estate physical properties such as residences, commercial buildings, and other real estate property [1]. One of the most important problems with property insurance is how to estimate premiums on property insurance. Premiums are the amount a policyholder pays to transfer risk to an insurer, and this can have significant implications for an insurer's difference between profit and loss, insolvency and viability as well as the customers' price to purchase risk transfer channel in a specific risk-adjusted context. In practice, premium calculations have relied on actuarial models and expert judgment. Traditionally, these are the foundational methods of premium calculation, but they often depend on linear relationships, are dependent on historical claims data, and are based on subjective judgments [2]. Premium calculations based on these methods typically cannot incorporate the increasing volume, complexity, and heterogeneity of datasets associated with modern insurance [3]. Furthermore, there are additional sources of dynamic risk associated with climate change, construction and modern features of structural properties [4], and changes in behavior of current and future policyholders, which also cannot be modeled using traditional premium calculation methods.

As structured and unstructured insurance data (from geospatial risk indicators to detailed claims history) is rapidly increasing, the demand for more data-driven and flexible methods is growing. In this context, mold a machine learning (ML) approach is a powerful alternative [5-7]. ML algorithms are strong performers when it comes to model complex, non-linear associations within considerably large datasets, and they do so without strict statistical assumptions. ML techniques have been applied successfully to many different insurance applications in recent years, including fraud detection, risk calculations, and claims management.

A. Motivation and Contribution

In an increasingly complicated risk environment, there is a pressing need for more precise and trustworthy ways to estimate property insurance rates, which is what motivated this work. Traditional pricing models often struggle to capture the intricate, non-linear relationships present in large insurance datasets, leading to less precise premium calculations that can affect both insurers and policyholders. To overcome these obstacles, this work employs state-of-the-art ML methods such as XGBoost, providing a scalable, data-driven approach that improves prediction accuracy and supports fairer, more efficient pricing. Ultimately, this work seeks to enhance the decision-making process in the property insurance industry, benefiting all stakeholders through better risk assessment and premium estimation. The key contributions of this study include*:*

- Developed an accurate property insurance premium prediction model using the XGBoost algorithm, successfully extracting intricate non-linear correlations from the dataset.
- Demonstrated the model's superior predictive performance in contrast to conventional techniques like LR, confirmed by important metrics like RMSE, MSE, and MAE.
- Utilized a comprehensive home insurance dataset with 42 attributes, incorporating extensive feature engineering and preprocessing to enhance model robustness and reliability.
- Provided actionable insights through feature importance analysis, aiding insurers in understanding critical factors influencing premium pricing and improving decision-making processes.

B. Novelty and Justification of the Paper

The applicability of this study is what makes it innovative of the XGBoost algorithm to the domain of property insurance premium prediction, leveraging a comprehensive home insurance dataset with 42 diverse attributes. Unlike traditional statistical methods, this approach integrates advanced ML techniques to model complex, non-linear relationships and interactions among customer, property, and risk-related features. The study is justified by the growing demand for data-driven, scalable, and accurate insurance pricing systems in the property insurance sector. By demonstrating superior performance in terms of key evaluation metrics and offering strong generalization capabilities, the proposed model provides a significant advancement over conventional approaches, ensuring more precise premium estimation and supporting informed decision-making for insurers.

C. Organization of the Paper

The structure of the paper is as follows: Related works are reviewed in Section II. The suggested technique is described in Section III. Experimental data are shown in Section IV, and important conclusions and future research objectives are discussed in Section V.

II. Literature Review

This section explores recent advancements in property insurance premium estimation through ML methods. In order to make premium estimates more accurate and reliable, the research that were examined utilized predictive modelling approaches to analyze property records, risk factors, and market trends. The studies reviewed include:

Yan and Nettayanun (2019) examines friction costs in the property and liability insurance market in China, with an eye on metrics including market penetration, industry leverage, hedging levels, and geographic concentration. Results show a negative relationship between capital structure, group, assets, and individual hedging levels, with insurers having less incentive to hedge [8].

Ajemunigbohun *et al.,* (2019) a stable, trust-based, and rational market system depends on an efficient fraud deterrent, according to research on homeowners' insurance fraud in Nigeria. A total of 221 insurance company employees (31 companies) participated in the study, which suggested that there was a need for government intervention in an effort to drive anti-fraud strategies. It offers a contribution to present knowledge of how academic research and practice to provide an anti-fraud strategy should be continuously engaged and interact with academics, insurance professionals, information technology specialists, and other interested parties in developing a plan to combat fraud and increase insurance market penetration and density [9].

Sudarwanto *et al.,* (2019) they developed insurance for the rental property industry through the use of GLM. Among the models used to investigate the links between distribution functions in this model are Poisson-

Gamma, Poisson-Inverse Gauss, Negative Binomial-Gamma, and Negative Binomial-Inverse Gauss. Simulation results show that random effects increase standard deviation values and variance in the average estimator [10].

Varadharajan (2018) explores the concept of insurance, which involves spreading risk among insurers. The use of big data systems can streamline the risk prediction process, enabling companies to make informed decisions. The project aims to build models to quantify risk associated with insuring a business, specifically in the property and casualty business. The models to be evaluated using predictive modelling principles, including data preprocessing, scaling, training, validation, and testing. The dataset used will cover various lines of business [11].

Lima Ramos (2017) paper discusses premium calculation principles used by insurance companies to calculate clients' premium values for risk exposure. It presents practical examples and discusses the approximation of the ruin probability of insurance claims using analytical and computational methods [12].

Guici and Yiyun (2017) study examines the market for property insurance in China's Hubei Province, evaluating six different regions. They analyze the factors influencing demand, predict their orientation, and establish an analysis using quantile regression and a fixed effect model. Evidence suggests that GDP, disposable income per capita, and insurance claims all work together to boost demand, while population growth negatively correlates with ITZ [13].

Table 1 provides a consolidated overview of key research on machine learning models for property insurance premium estimation, summarizes the main contributions of each study, identifies their limitations, and highlights potential directions for future improvements in premium prediction accuracy.

Table 1. Summary of recent studies on machine learning models for property insurance premium estimation.

Author	Approach	Dataset	Main contributions	Limitations	Future work
Yan and Nettayanun, 2019	Panel data analysis	Chinese Insurance Yearbook (2008–2015)	Investigated how market structure, capital structure, and firm-level characteristics affect reinsurance purchase and hedging in P&C insurance in China	Focused on strategic risk management; not predictive in nature; lacks machine learning techniques	Extend to predictive analytics models; explore machine learning for hedging behavior prediction
Ajemunigbohun, *et al.,* 2019	Descriptive analysis and T-test	Structured questionnaire (Nigerian insurance industry)	Empirical insight into homeowner's insurance fraud and industry capacity	Non-quantitative modeling; lacks premium estimation or risk scoring	Integrate data analytics or predictive models for fraud detection and premium adjustments
Sudarwanto, *et al.,* 2019	Generalized linear models (GLM)	Simulated property Rental business data	Applied GLM to model premium estimation with multiple distribution assumptions	Used simulation; lacks real insurance datasets and modern ML methods	Incorporate real-world insurance datasets and compare with ML techniques
Varadharajan, 2018	Predictive modeling (Supervised	Iowa property and casualty	Developed risk quantification models using	Lacked granularity for line-of-	Expand to line-specific models and explore

	ML)	insurance premiums and losses	P&C insurance data; focused on model evaluation and business decision support	business differentiation; limited scope	additional ML algorithms
Lima Ramos, 2017	Analytical and computational risk models	Premium calculation	Examined premium calculation principles using models like Cramér-Lundberg and non-homogeneous poisson processes	Traditional modeling focus; lacks empirical testing or predictive component	Apply ML to traditional premium principles and integrate big data
Guici and Yiyun, 2017	Panel data, fixed effects, quantile regression	Data from Hubei and 5 other Chinese provinces	Analyzed demand drivers for property insurance in Hubei; established correlation with macroeconomic factors	Demand-side focus; doesn't predict premiums or use ML	Introduce premium prediction models using macro-and micro-level features

III. Methodology

This methodology presents a structured approach to predicting property insurance premiums using machine learning techniques in Figure 1. It begins with the acquisition of home insurance data, followed by a detailed data preprocessing phase involving data conversion, cleaning, feature engineering, encoding, and scaling to ensure high-quality inputs. The cleaned-up dataset is split into two halves, the training set and the testing set, in order to construct and validate a model. To understand the intricate correlations between characteristics and premium values, though, the training data is run via the XGBoost algorithm, which excels in regression problems. After training, the model's prediction accuracy is assessed on test data using important assessment metrics such RMSE, MAE, and MSE. These metrics allow one to quantitatively assess how well the model is able to predict the system performance, that is, how accurate and reliable the model is the model is then interpreted to derive useful insights about the model's capacity to make accurate insurance premiums estimations.

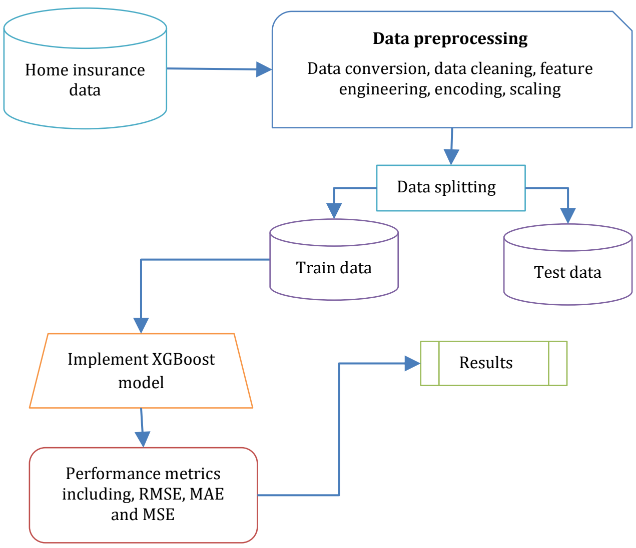

Figure 1. Workflow of predictive modeling for property insurance premium estimation.

A. Dataset Description

The dataset used is Home Insurance Data, consisting of the details of home insurance policies from 42 attributes and large number of records. It comprises the information about customers, properties, associated risks and policies offered in order to predict annual insurance premiums. Key factors considered in the dataset include building and content coverage, number of bedrooms, year of construction, ownership type, hazard-rated areas, and the customer's claim history. The Annual_Premium serves as the target variable, representing the yearly cost of the insurance policy. This rich and diverse dataset enables accurate prediction of property insurance premiums using machine learning techniques. The data visualizations are provided below:

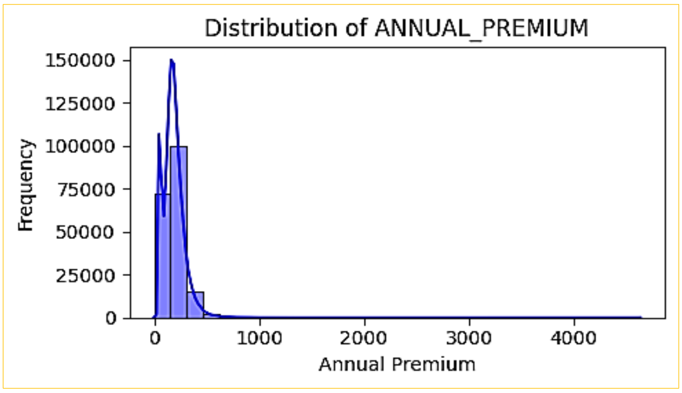

Figure 2. Distribution of the annual premium values.

Annual premium amounts are distributed as shown in Figure 2. It's heavily right-skewed, with most premiums concentrated between 0 and 500, indicated by a tall bar and a steep blue curve. Frequencies drop sharply for higher premium amounts, suggesting fewer instances of very large annual premiums.

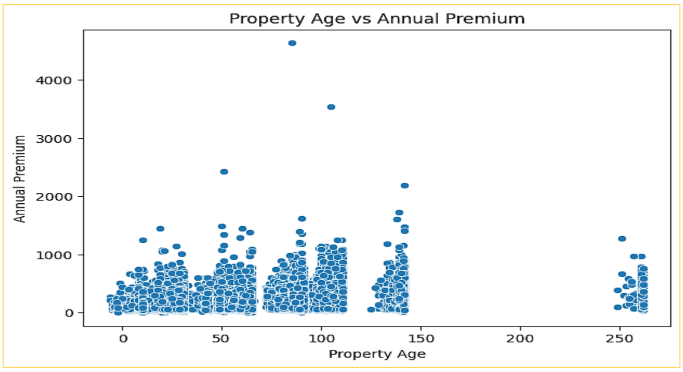

Figure 3. Property age vs. annual premium scatter plot.

In Figure 3, the x-axis represents "property age," extending to approximately 275, while the y-axis shows "annual premium," up to 4500. The data points are clustered at various property ages, particularly noticeable around 0-150 and 250-275, with some outliers showing higher premiums. There doesn't appear to be a clear linear correlation, but rather distinct groups of data points at specific property ages.

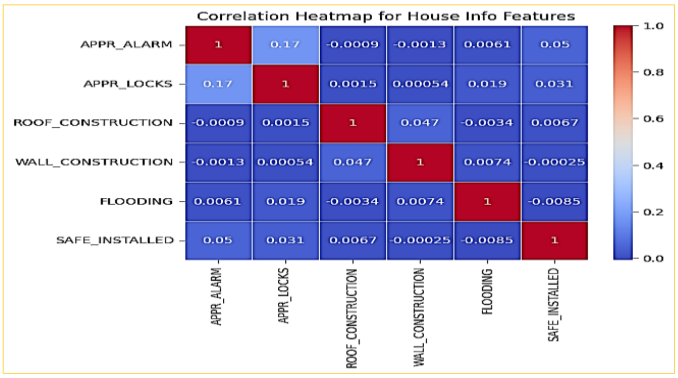

Figure 4. Correlation heatmap of the dataset.

The correlation coefficients between six characteristics are displayed in the correlation heatmap in Figure 4: Appr_Alarm, Appr_Locks, Roof_Construction, Wall_Construction, Flooding, and Safe Installed. The diagonal shows perfect positive correlation (1) for each feature with itself. The off-diagonal values, represented by with a colour gradient, show how strongly and in what direction various attributes are related to one another, with values close to zero showing weak or no correlation.

B. Data Preprocessing

Dara Preprocessing is the set of techniques applied to raw data to clean, transform, and prepare it for analysis or modeling. It ensures the data is in the right format for ML algorithms and helps enhance data quality. Here are the steps:

- **Date Conversion:** Converted categorical columns into datetime format using custom parsing to handle multiple formats. Enabled extraction of Age and Property_Age, as well as temporal grouping for trend analysis.
- **Data Cleaning:** Removed entries with Annual_Premium <= 0 to eliminate noise and invalid records. Dropped redundant date features after extraction. Ensured clean, structured input for model training.
- **Missing Value Treatment:** Removed features with a high percentage of missing data. Applied median imputation for numeric fields with moderate null values. Used the most common category to fill missing values in categorical data and applied forward fill for temporal gaps to maintain sequence integrity.

C. Feature Engineering

To enrich the dataset with meaningful inputs, new features were created based on existing data. Client age (age) was calculated from the difference between cover start and p1_dob, while property age was derived from the construction year. Additional fields that are temporal features (quote year, quote month and cover_start_year) were extracted from datetime fields. The binary categorical columns such as flooding, subsidence and buildings cover were encoded using a simple mapping. The model these enhancements increased the predictive power of the dataset by giving models more context and patterns among individual profiles and property attributes.

D. Categorical Encoding

The categorical variables were transformed using one-hot encoding, which had multiple unique classes, so that they were compatible with machine learning models. Each of the categories was turned into binary columns without any ordinal bias. The redundant columns were then dropped so as not to retain the original categorical columns. It was important for this step that all input features were numeric and in the correct format for regression algorithms so that learning capability and convergence were improved. Encoding enabled, keeping the granularity of the qualitative information, and at the same time aligning with the input requirements of the models.

E. Feature Scaling

The standard scaler is a data preprocessing technique that assigns a mean of 0 and a standard deviation of 1 to data points in a dataset that are organized so that they fit the normal distribution. As shown in Equation (1) the scaling plays a crucial role in several ML methods that are sensitive to the size of input characteristics:

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

In this case, μ represents the feature's mean and σ stands for its standard deviation.

F. Data Splitting

In this analysis, the dataset is divided using an 80/20 ratio into two separate subgroups. The ML models are trained using the bigger subset, which makes up 80% of the data, to accurately estimate premiums; the remaining 20% is set aside for testing and confirming the prediction performance of the models.

G. Implementation of XGBoost Regressor

XGBoost, developed by Chen and Guestrin, is an innovative technique for boosting gradient trees [14]. Applying a group of CART as weak learners is the initial step. Then, to improve the trees' performance, an ensemble of trees is created to minimise a regularised objective function. This approach enhanced earlier gradient tree boosting techniques by including concepts such as cache-friendly approximations for splitting point determination, sparsity-aware split finding into each tree, and effective out-of-core computation. The result is an algorithm that is both computationally fast and has good prediction ability [15]. At the t-th boosting iteration, Equation (2) expresses the basic goal function of XGBoost:

$$L^t = \sum_{i=1}^{n} l\big(y_i, F_{t-1}(x_i) + f_t(x_i)\big) + \Omega(f_t) \quad (2)$$

In Equation (3), l is a complicated loss function that distinguishes the expected i-th result $F_{t-1}(x_i)$ of the $(t-1)$ -th ensemble from the i-th outcome y_i. $\Omega\ (ft)$ is a function that penalises tree complexity with T, and it is the regularisation term that penalises model complexity.

$$\Omega(f_t) = \gamma^T + \frac{\lambda||w||^2}{2} \quad (3)$$

The regularization hyperparameter of XGBoost is denoted by w, which is the number of leaves, while the minimal loss hyperparameter is denoted by λ.

H. Performance Measurement Parameters

In order to determine how well regression models estimate property insurance premiums, the key performance indicators are detailed in this section. The model's performance may be quantitatively understood using these indicators:

Mean Absolute Error (MAE): The anticipated errors are added together to give the absolute error. By averaging all absolute errors, we get the mean absolute error [16]. Equation (4) represents the mathematical calculation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (4)$$

Mean Squared Error (MSE): The acronym for mean squared error is MSE. One measure of risk is the MSE, or estimated squared error loss. As demonstrated in Equation (5), the MSE takes into account the estimator's bias and variance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (5)$$

Root Mean Squared Error (RMSE): The average of the squared differences between the actual and anticipated values is known as the RMSE; it is shown in Equation (6).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (6)$$

A comprehensive picture of the model's performance with the predictions is provided by these assessment measures taken together.

IV. Result Analysis and Discussion

This study uses sophisticated ML algorithms to better correctly and consistently anticipate property insurance prices. Specifically, we use the XGBoost model for regression tasks owing to its efficiency and ability to harness complex nonlinear relationships that exist in data. The development and evaluation of models was made using Python in the Jupyter Notebook environment, on a standard desktop workstation with 16GB of RAM was enough to process large-scale insurance datasets. Powerful libraries such as Scikit-learn, XGBoost, Pandas as well as NumPy provided good support for the modeling process. Following the values given in Table 2, the XGBoost model has produced an RMSE of 52.99, an MSE of 28.07 and an MAE of 30.64. Results show that the model has strong predictive power and verify that the model has potential for use in real estate property insurance pricing systems as a data-driven and scalable solution for accurate premium estimation.

Table 2. Evaluation results of the XGBoost model on property insurance premium prediction.

Metrics	XGBoost model
RMSE	52.99
MSE	28.07
MAE	30.64

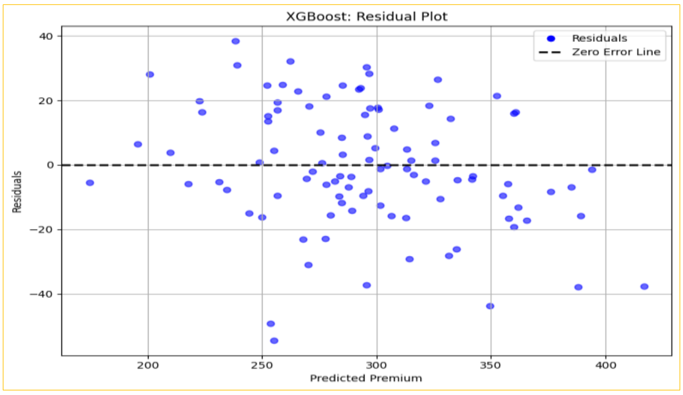

Figure 5. Residual plot for XGBoost model.

This residual plot in Figure 5, displays residuals (prediction errors) against predicted premium values. The scatter of blue data points, ideally showing no discernible pattern around the dashed "zero error line," indicates a well-performing model where errors are randomly distributed, suggesting no systematic biases in the predictions across different premium ranges.

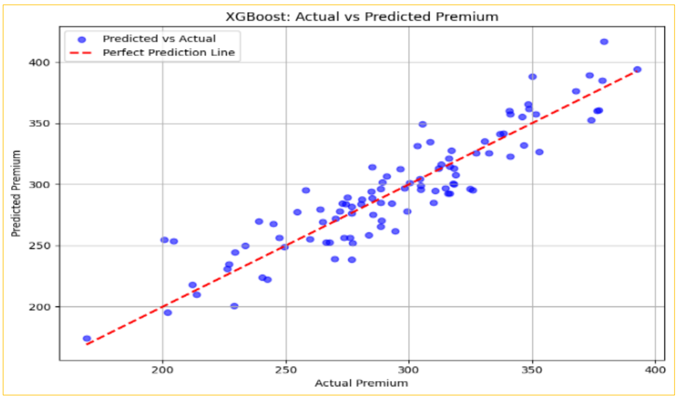

Figure 6. Actual vs. predicted premium plot for XGBoost model.

The Figure 6 displays a scatter plot, illustrating the performance of an XGBoost model. Blue dots represent individual data points, showing the predicted premium against the actual premium. The blue dashed line, which is known as the perfect prediction line, would represent the expected values. Premiums with a broad range of values may be predicted by this model, and the fact that the points are clustering around this line shows that it fits well.

A. Comparative Analysis

The article concludes with a comparison of various ML algorithms that may be used to estimate mortgage rates. The XGBoost model exhibited the best prediction accuracy of all the models evaluated, as shown in Table 3, with the lowest RMSE of 52.99. XGBoost's performance in this application is superior, and it demonstrates itself as the best model to accurately estimate insurance premiums and, in turn, the most suitable choice for property insurance pricing tasks. On the other hand, the RMSE for LR was 57.09, which is higher and hence lower in predictive precision. These results clearly show that XGboost provides better premium estimations with more reliability.

Table 3. Comparative performance of machine learning models for property insurance premium estimation.

Models	**RMSE**
XGBoost	52.99
Linear regression [17]	57.09

The high predictive accuracy of the proposed XGBoost model and its ability to capture complex, non–linear data relationships indicate that such a model can be useful for predicting the LOA of a system performing AC duties. It has a helpful built-in regularization that prevents overfitting and handles missing values well. XGBoost is a good solution for property insurance pricing tasks because it is fast, scalable and finds important features [18-47].

V. Conclusion and Future Scope

Estimation of insurance premiums is very important and equally important for both insurers and customers, as the premiums represent the amount policyholders pay annually to maintain coverage. The nature of the pricing premium is that it depends upon many parameters such as characteristics of the property, risk exposure and history of claims. In predicting property insurance premiums, an advanced XGBoost, ML approach was used in this study. The proposed XGBoost model achieved strong predictive performance by means of a complete workflow that consisted of data preprocessing, feature engineering and model evaluation. The model was able to achieve 52.99 RMSE which was far superior to the comparative linear regression model which achieved a higher RMSE. From this, we can clearly see that XGBoost is a more accurate and more reliable model to extract intricate and non-linear connections from the insurance data, making it a more suitable model for property insurance pricing tasks in the real world. Additional future work could also look at including the implementation of additional models, for instance, an ensemble learning model or DL architectures that could help improve the prediction accuracy. External data sources, which include geographic risk indices, weather data or economic indicators, can be incorporated to make models more robust. Furthermore, using the insurance premium prediction framework, interpretable AI methods will also be developed to ensure transparency and trust among industry stakeholders. This research would also be extended to its real-time deployment in production environments, and monitoring the performance of the model over time.

Declarations

Acknowledgments: We gratefully acknowledge all of the people who have contributed to this paper.
Author Contributions: All authors contributed equally to this work.
Conflict of Interest: The authors declare no conflict of interest.
Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.
Data Availability Statement: The data presented in this study are available upon request from the corresponding author.
Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Research Content: The research content of manuscript is original and has not been published elsewhere.

References

1. Antal, R.M. 2012. The premium for mandatory house insurance in Romania–considerations regarding its financial solvability. Procedia Economics and Finance, 3: 829-836.

2. Ali, S.D. 2016. Ratemaking for emerging liabilities in property and casualty insurance: Practical tools and enriching imagination. Available at SSRN 2800738.

3. Savatteri, V. 2018. The impact of IoT on the home insurance industry. Master Thesis, Copenhagen Business School.

4. Lee, H.H. and Lee, C.Y. 2012. An analysis of reinsurance and firm performance: Evidence from the Taiwan property-liability insurance industry. The Geneva Papers on Risk and Insurance-Issues and Practice, 37: 467-484.

5. Shim, J. 2011. Mergers and acquisitions, diversification and performance in the US property-liability insurance industry. Journal of Financial Services Research, 39: 119-144.

6. Garg, S. 2019. Predictive analytics and auto remediation using artificial inteligence and machine learning in cloud computing operations. The International Journal of Innovative Research in Engineering and Multidisciplinary Physical Sciences, 7(2): 1-5.

7. Kolluri, V. 2015. A comprehensive analysis on explainable and ethical machine: Demystifying advances in artificial intelligence. TIJER–International Research Journal, 2(7): a1-a5.

8. Yan, Y. and Nettayanun, S. 2019. Strategic risk management in the Chinese property and casualty insurance industry. In: Asia-Pacific contemporary finance and development (pp. 89-115). Emerald Publishing Limited.

9. Ajemunigbohun, S.S., Isimoya, O.A. and Ipigansi, P.M. 2019. Insurance claims fraud in homeowner's insurance: Empirical evidence from the Nigerian insurance industry. Facta Universitatis, Series: Economics and Organization, 16(1): 103-116.

10. Sudarwanto, S., Ambarwati, L. and Hadi, I. 2019. Development of rental property insurance models with Generalized Linear Models (GLM). Journal of Physics: Conference Series, 1402(7): 077104.

11. Varadharajan, S. 2018. Risk prediction for property and casualty insurance policies. Master's Thesis, University of Massachusetts Lowell.

12. Lima Ramos, P. 2017. Premium calculation in insurance activity. Journal of Statistics and Management Systems, 20(1): 39-65.

13. Guici, C. and Yiyun, L. 2017. The empirical analysis on demand of property insurance in Hubei Province based on panel data. In: 2017 29th Chinese control and decision conference (CCDC) (pp. 6378-6383). IEEE.

14. Chen, T. and Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).

15. Fauzan, M.A. and Murfi, H. 2018. The accuracy of XGBoost for insurance claim prediction. International Journal of Advances in Soft Computing and Its Applications, 10(2): 159-171.

16. Yunos, Z.M., Ali, A., Shamsyuddin, S.M., Ismail, N. and Sallehuddin, R.S. 2016. Predictive modelling for motor insurance claims using artificial neural networks. International Journal of Advances in Soft Computing and Its Applications, 8(3): 160-172.

17. Kitchens, F. and Harris, T. 2015. Genetic adaptive neural networks for prediction of insurance claims. International Journal of Engineering and Advanced Research Technology, 1(6): 2454-9290.

18. Polu, A.R., Narra, B., Vattikonda, N., Gupta, A.K., Buddula, D.V.K.R. and Patchipulusu, H.H.S. 2021. Evolution of AI in software development and cybersecurity: Unifying automation, innovation, and protection in the digital age. International Journal of Research in Engineering and Applied Sciences, 11(5): 1-15.

19. Buddula, D.V.K.R., Patchipulusu, H.H.S., Polu, A.R., Narra, B., Vattikonda, N. and Gupta, A.K. 2021. Integrating AI-based sentiment analysis with social media data for enhanced marketing insights. International Journal of Engineering, Science and Mathematics, 10(2): 56-68.

20. Narra, B., Buddula, D.V.K.R., Patchipulusu, H.H.S., Polu, A.R., Vattikonda, N. and Gupta, A.K. 2023. Advanced edge computing frameworks for optimizing data processing and latency in IoT networks. Journal of Emerging Trends in Scientific Research, 1(1): 1-10.

21. Buddula, D.V.K.R., Patchipulusu, H.H.S., Vattikonda, N., Polu, A.R., Narra, B. and Gupta, A.K. 2023. Predictive analytics in e-commerce: Effective business analysis through machine learning. JOETSR-Journal of Emerging Trends in Scientific Research, 1(1): 41-49.

22. Kalla, D. 2022. AI-powered driver behavior analysis and accident prevention systems for advanced driver assistance. International Journal of Scientific Research and Modern Technology, 1(12): 1-9.

23. Kuraku, D.S., Kalla, D. and Samaah, F. 2022. Navigating the link between internet user attitudes and cybersecurity awareness in the era of phishing challenges. International Advanced Research Journal in Science, Engineering and Technology, 9(12): 116-124.

24. Kalla, D., Smith, N., Samaah, F. and Polimetla, K. 2022. Enhancing early diagnosis: Machine learning applications in diabetes prediction. Journal of Artificial Intelligence and Cloud Computing, 1(1): 1-7.

25. Kalla, D., Kuraku, D.S. and Samaah, F. 2021. Enhancing cyber security by predicting malwares using supervised machine learning models. International Journal of Computing and Artificial Intelligence, 2(2): 55-62.

26. Katari, A. and Kalla, D. 2021. Cost optimization in cloud-based financial data lakes: Techniques and case studies. ESP Journal of Engineering and Technology Advancements, 1(1): 150-157.

27. Kalla, D., Smith, N., Samaah, F. and Polimetla, K. 2021. Facial emotion and sentiment detection using convolutional neural network. Indian Journal of Artificial Intelligence Research, 1(1): 1-13.

28. Chinta, P.C.R., Katnapally, N., Ja, K.M., Bodepudi, V., Boppana, S.B. and Sakuru, M. 2022. Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Studies, 10(2): 665-674.

29. Maka, S.R., Bodepudi, V., Routhu, K., Jha, K. M., Rao Chinta, P.C.R. and Sakuru, M. 2020. A deep learning architecture for enhancing cyber security protocols in big data integrated ERP systems. Journal of Artificial Intelligence and Big Data, 1(1): 1238.

30. Katnapally, N., Chinta, P.C.R., Jha, K.M., Routhu, K.K., Velaga, V. and Sadaram, G. 2021. Neural network-based risk assessment for cybersecurity in big data-oriented ERP infrastructures. Educational Administration: Theory and Practice, 27(4): 1329-1341.

31. Katnapally, N., Chinta, P.C.R., Routhu, K., Velaga, V., Bodepudi, V. and Karaka, L.M. 2021. Leveraging big data analytics and machine learning techniques for sentiment analysis of Amazon product reviews in business insights. American Journal of Computing and Engineering, 4(2): 35-51.

32. Chinta, P.C.R., Jha, K.M., Routhu, K., Velaga, V., Moore, C.S. and Boppana, S.B. 2022. Enhancing supply chain efficiency and performance through ERP optimisation strategies. Journal of Artificial Intelligence and Cloud Computing, 1(4): 1-7.

33. Sadaram, G., Sakuru, M., Karaka, L.M., Reddy, M.S., Bodepudi, V., Boppana, S.B. and Maka, S.R. 2022. Internet of things (IoT) cybersecurity enhancement through artificial intelligence: A study on intrusion detection systems. Universal Library of Engineering Technology, 2022: 01-09.

34. Moore, C.S., Boppana, S.B., Bodepudi, V., Jha, K.M., Maka, S.R. and Sadaram, G. 2021. Optimising product enhancements strategic approaches to managing complexity. American Journal of Computing and Engineering, 4(2): 52-72.

35. Jha, K.M., Bodepudi, V., Boppana, S.B., Katnapally, N., Maka, S.R. and Sakuru, M. 2023. Deep learning-enabled big data analytics for cybersecurity threat detection in ERP ecosystems. Review of Contemporary Philosophy, 22(1): 6193-6209.

36. Kalla, D. and Chandrasekaran, A. 2023. Heart disease prediction using chi-square test and linear regression. Computer Science and Information Technology, 13: 135-146.

37. Kalla, D. and Kuraku, S. 2023. Phishing website URL's detection using NLP and machine learning techniques. Journal of Artificial Intelligence, 5: 145-162.

38. Kuraku, D. and Kalla, D. 2023. Impact of phishing on users with different online browsing hours and spending habits. International Journal of Advanced Research in Computer and Communication Engineering, 12(10): 34-41.

39. Kuraku, S., Kalla, D., Samaah, F. and Smith, N. 2023. Cultivating proactive cybersecurity culture among IT professional to combat evolving threats. International Journal of Electrical, Electronics and Computers, 8(6): 1-7.

40. Kuraku, S., Kalla, D., Smith, N. and Samaah, F. 2023. Exploring how user behavior shapes cybersecurity awareness in the face of phishing attacks. International Journal of Computer Trends and Technology, 71(11): 74-79.

41. Routhu, K., Velaga, V., Moore, C.S., Boppana, S.B., Chinta, P.C.R. and Ja, K. 2023. Leveraging machine learning techniques for predictive analysis in merger and acquisition (M&A). Journal of Artificial Intelligence and Big Data, 3(1): 56–71.

42. Kuraku, S., Kalla, D., Smith, N. and Samaah, F. 2023. Safeguarding FinTech: Elevating employee cybersecurity awareness in financial sector. International Journal of Applied Information Systems, 12(42): 43-47.

43. Maka, S.R., Jha, K.M., Chinta, P.C.R., Moore, C.S., Katnapally, N. and Sadaram, G. 2023. AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnology Perceptions, 19(S1): 46-64.

44. Chinta, P.C.R., Jha, K.M., Routhu, K., Velaga, V., Moore, C.S., Boppana, S.B. and Sakuru, M. 2023. The art of business analysis in information management projects: Best practices and insights. Journal of Contemporary Education Theory and Artificial Intelligence, JCETAI-103.

45. Jha, K.M., Bodepudi, V., Katnapally, N., Maka, S.R., Karaka, L.M., Sadaram, G. and Sakuru, M. 2023. Optimising sales forecasts in ERP systems using machine learning and predictive analytics. Journal of Contemporary Education Theory and Artificial Intelligence, JCETAI-104.

46. Sadaram, G., Sakuru, M., Jha, K.M., Bodepudi, V., Katnapally, N., Maka, S.R. and Karaka, L.M. 2023. Understanding the fundamentals of digital transformation in financial services: Drivers and strategic insights. Journal of Artificial Intelligence and Big Data, 3(1): 72–83.

47. Sadaram, G., Routhu, K., Velaga, V., Boppana, S.B., Katnapally, N. and Sakuru, M. 2023. Machine learning for cyber defense: A comparative analysis of supervised and unsupervised learning approaches. Journal for ReAttach Therapy and Developmental Diversities, 6(10s) (2): 1790-1803.

Citation: Ram Mohan Polam, Bhavana Kamarthapu, Ajay Babu Kakani, Sri Krishna Kireeti Nandiraju, Sandeep Kumar Chundru and Srikanth Reddy Vangala. 2025. Predictive Modeling for Property Insurance Premium Estimation Using Machine Learning Algorithms. International Journal of Recent Innovations in Academic Research, 9(3): 55-66.