International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

PREDICTIVE MODELING FOR PROPERTY & CASUALTY INSURANCE PREMIUMS BASED ON MACHINE LEARNING ALGORITHMS

Vigneshwaran Thangaraju
CGI Technology and Solutions
Aldie, Virginia, USA
vigsan714@gmail.com

Abstract

Opening the door to great potentials in a realm of insurance, predictive analytics relying on the principles of machine learning is transforming the ways property and casualty insurance premium rates are determined. The target of this work is to analyze the multiple feature datasets and improve the premium prediction through machine learning algorithms. Moreover, the study aims at establishing whether models such as RF and SVM, that bear the capability to reveal non-linear patterns, would be useful on large datasets of insurance. These models are assessed by critical measures of accuracy and precision, and recall value, which in conjunction form the F1-score rate. The comparison analysis shows that the Random Forest makes much higher accuracy than other classifiers with 99.99% accuracy rate. This underlines its advantage as a sound instrument for advancing premium estimates while enhancing its assessment of risks. Using the mentioned sophisticated techniques, insurers can therefore identify ways of correcting risk aspect hence achieving better underwriting of prices. According to the findings of the study, ML models outperform traditional ones in both theoretical and applied settings. It will be beneficial to the property and casualty insurance industry as it will enable it to make right decisions by providing customers with accurate and fair pricing estimates.

Keywords: Casualty insurance, Predictive Modeling, Random Forest, Insurance Dataset, Property and Casualty.

I. INTRODUCTION

It is an actively developing trend over the last few years that involves the use of predictive modeling in the insurance business, with a focus on the P&C segment [1][2]. It features how well insurers can accurately forecast insurance premiums in order to control risk and improve pricing and profitability [3]. Conventional techniques applied to measure and base price on risk have mostly been actuarial, but these bear linear coefficients and finite sets, thus making the premium strategies less effective. As the amount of voluminous data and its enhanced quantity and quality increases, the applications of machine learning (ML) algorithms[4] have a advantage of having higher accuracy and better robustness due to identification of non-linear relationships and interactions in data set [5]

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

Although, there is potential for machine learning in P&C insurance, there is general concerns that arise due to interpretability of the models and regulatory compliance [6]. An actuary is needed to explain the underlying price determination to the regulators and customers, which is a problem when models are based on opaque ML algorithms[7][8]. Despite this, researchers are attempting to develop new and improved forms of ML models that are interpretable and possess high predictive accuracies that do not compromise the regulatory thresholds[9]. It is expected that ML is going to become more important in the field of P&C insurance as the discipline enhances, which leads to more precise and efficient premium estimation [10][11].
Another crucial drawback of P&C insurance companies when using predictive modeling is related to the data imbalance. There is a skewed distribution in the majority of insurance datasets since there are very few claims compared to the total number of policies issued. This imbalance can lead to poor performance of traditional models, especially when trying to estimate claims for risky customers. To address this, techniques like the SMOTE are used to balance a dataset and improve a performance of ML models[12][13]. Furthermore, feature selection is essential for these models since it improves their accuracy and interpretability by excluding irrelevant factors from the forecast [14][15].

A. Motivation of the Study

Property and casualty insurance in particular, the insurance business is currently experiencing increased demand for accurate predictive models hence this research. Inspired by the increasing volume and density of the insurance data, some traditional methods have been proven to be inadequate in the ways of accommodating huge data sets, which can, in turn, likely give rise to inefficiency in premium pricing as well as risk management. This research seeks to discover the enhanced, sustainable methods of the chosen machine learning models for attaining more accurate predictions, precise premium determination, and risk estimation. Finally, the results will help improving decision-making in the insurance domain for the advantage of both insurers and policyholders.

B. Contribution of the paper

This study contributes to a number of significant advancements in the area of Property & Casualty Insurance Premiums by using a number of different ML models. It is essential to take notice of the primary contributions, which are as follows:

- The study develops a comprehensive data preprocessing technique that can involve missing values, outlier detection, one-hot encoding, and feature selection, especially suitable for insurance data to prepare them for ready analysis by machine learning models.
- It enhances the scalability of the model and its feature in classification by using SMOTE approach in class imbalance.

- The optimal models for predicting property and casualty insurance premiums were determined by comparing and discussing the results of several ML models, including RF and SVM.
- The research employed a variety of performance metrics to assess the reliability and efficacy of the model, including accuracy, precision, recall, and F1-score.
- The study applies its methodology to a real-world insurance dataset, offering practical implications for improving risk assessment and premium calculations in the insurance industry.

C. Organization of the paper

The following is the outline of the paper: A review of the relevant literature and a theoretical framework are presented in Section III. Section II describes the models, procedures, and materials that were used. The findings and discussion of the comparisons are presented in Section IV. The most crucial findings and recommendations for further research are laid out in Section V.

II. RELATED WORK

Recent studies presented in this section describe a various ML algorithm that can be utilized to estimate the Predictive Modeling for Property & Casualty Insurance Premiums. The following contains some background research.

Jyothsna et al. (2022) purpose of the proposed study is to predict how much health insurance would cost and to find people who have health insurance and relevant medical records, independent of their current health status. This study utilized a variety of regression models, including Multilinear, DT, RF, and Gradient Boosting Regression. The results showed that, with an accuracy of 87%, Gradient Boosting was the best strategy out of the bunch [16].

(Panda et al. (2022) creates an MLHIPS that uses ML algorithms to forecast insurance costs in real-time. This system will help market insurance businesses quickly and easily determine premium prices, which will reduce health expenditures. Among the several models included in the proposal, the Polynomial Regression model outperformed the others with an RMSE of 5100.53 and an R-squared value of 0.80 [17].

Dutta et al. (2021) concentrates on estimating the amount of the patient's health insurance premium. To find the optimal strategy, we utilized the r2 score, RMSE, and MSE of every regression method to measure accuracy after these algorithms were run for prediction. As far as health insurance cost prediction algorithms go, RFR stands head and shoulders above the competition with an ideal r2 score of 0.862533 [18].

Utomo, Damanik and Budi. (2021) concentrates on categorizing participants in the insurance renewal process so that the business may approve participation more selectively. The suggested approach uses 3803 datasets with four characteristics and five algorithms to classify the data of insurance users and identify important elements for the model's generation. The models will undergo validation using k-fold cross-validation with k=10. The assessment results will provide the following information on the accuracy of each algorithm: Among the methods used, 70.00% were NB, 67.00% were SVMs, 95.40% were DT, 90.20% were LR, and 79.30% were NNNs. The research concluded that the DT algorithm outperformed the alternatives when it came to classifying renewal firms that would become insurance participants, with an accuracy score of 95.40% [19].

Sun et al. (2021) puts forth a new technique that combines bagging trees with DTW to identify driving events using the orientation and acceleration data from the inexpensive three-axis accelerometers and gyroscopes found in smartphones. Results from field tests demonstrate that the suggested integrated algorithm outperforms the state-of-the-art, with a 97.5% success rate in proper identification, a 2.5% miss rate, and a 2.9% FPR. For the best alternative candidate approach, the comparable outcomes are 90.2%, 9.8%, and 11.7%. In addition, compared to existing state-of-the-art methods, our suggested method provides a computational efficiency boost that is three to ten times larger [20].

The property and casualty insurance premiums techniques are compared in the following Table I, which presents the existing work

TABLE I. Comparative study on property and casualty insurance premiums using machine learning models.

References	Methodology	Dataset	Performance	Limitations & Future Work
[16]	Multi-Linear Regression, DecisionTree, RandomForest, GradientBoosting Regression	HealthInsurance Dataset	Gradient Boosting achieved 87% accuracy	Limited to insurance cost prediction; potential for expanding feature engineering and dataset size
[17]	RidgeRegression, Lasso Regression, Simple Linear Regression, Multiple Linear Regression, Polynomial Regression	ML Health Insurance Prediction System (MLHIPS)	Polynomial Regression achieved RMSE of 5100.53 and R^2 of 0.80	Requires enhancement with ensemble methods for improved accuracy
[18]	Predicting health insurance costs using multiple regression algorithms. Comparison between actual and predicted expenses.	Health insurance dataset.	Best performance by Random Forest Regression with r2_score = 0.862533.	Limited to regression analysis; future work could involve testing additional algorithms or improving data preprocessing.
[19]	Classification of insurance renewal participants using five different algorithms and feature selection.	Insurance participants' data with 3803 entries and four attributes.	Decision Tree achieved highest accuracy of 95.40%.	Future work could explore combining models or applying ensemble techniques for better accuracy.
[20]	Bagging tree and	Acceleration	Detection accuracy	Future work could involve

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

	Dynamic Time Warping (DTW) integrated for driving event detection using smartphone sensor data.	and orientation data from smartphone sensors.	of 97.5%, with missed detection rate of 2.5% and false detection rate of 2.9%.	testing the algorithm with more extensive datasets and improving its generalizability to other sensors.

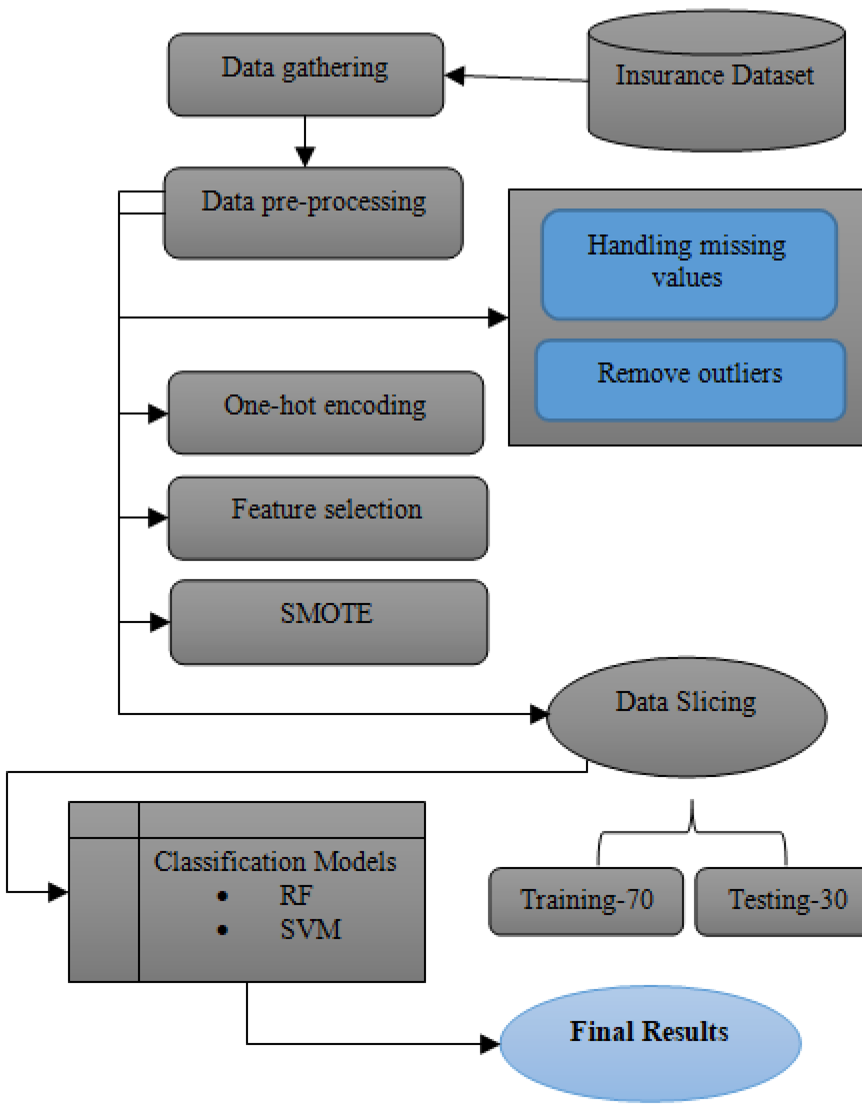

Fig. 1. Flowchart of methodology for property & casualty insurance.

III. RESEARCH METHODOLOGY

The purpose of this research is to use ML techniques to develop premium prediction models for property and liability insurance. To provide insurers with a more potent and impartial instrument for valuable policy pricing based on the characteristics of customers and properties, the goal here is to make premium projections more accurate and efficient. From this evaluation, the result will assist in enhancing the capacity of augmenting prediction for property and casualty premiums for this study, the following procedures are undertaken. First, the batch of insurance data collected from 2012 to 2016 was obtained and cleaned by dealing with missing values, removing outliers, encoding the categorical data, and creating feature selection techniques. The dataset was balanced with SMOTE and the total dataset was split 70:30 between the training and experimental sets. After preprocessing the data, classification models which included RF and SVM was used for the analysis. Metrics for model performance including F1-score, recall, accuracy, and precision were used for evaluation. Figure 1 below displays the flowchart of the property and casualty to enhance the predictive modeling.
Below are detailed descriptions of every step in a flowchart diagram.

A. Data gathering

The insurance dataset is made up of data that was gathered throughout the years of observation, which are 2012 through 2016. It was observed that different people collected the dataset throughout this observation time. Table II lists the variables in the insurance dataset and provides explanations of each one.

TABLE II. Variables and Their Description

Customer ID	The Policyholder's unique identifier
Year of observation	Calendar year in which the insured policy was monitored
Insured period	Olusola Insurance policy length (for example, a full year of coverage, Policy Length= 1; six months= 0.5
Residential	Regardless of whether the structure is a residence or not
Building painted	Whether the structure has paint or not (N=Painted, V=Not Painted)
Building fenced	Whether the structure has perimeter fencing or not (N-Fenced, V-Not Fenced)
Garden	The building's garden status (V for garden, O for no garden).
Settlement	Exactly where the structure stands. (U-urban complex; R-rural region)

iJCEM
International Journal of Core Engineering & Management

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

Building dimension	Building area in m2 of the insured
Building type	The kind of structure (Type 1, 2, 3, 4)
Date of occupancy	Year or Date of First Occupancy of the Building
Number of windows	Number of windows
Geo-code	The insured building's geography code
Claim	Variable to be targeted. No claims (0), one claim (1 or more within the covered period).

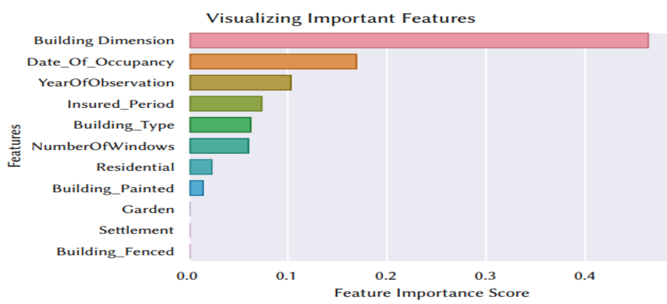

Fig. 2. Feature Importance score

Figure 2 visualizing Important Features, which illustrates the relative importance of various features in a model. The features include 'Date of Occupancy', 'Building Dimension', 'Year Of Observation', 'Building Type', 'Insured Period', 'Number Of Windows', 'Building Painted', 'Garden', 'Settlement', 'Residential', and 'Building Fenced'. Each feature is represented by a colored bar, indicating its importance score, which helps in identifying which features most significantly impact model predictions.

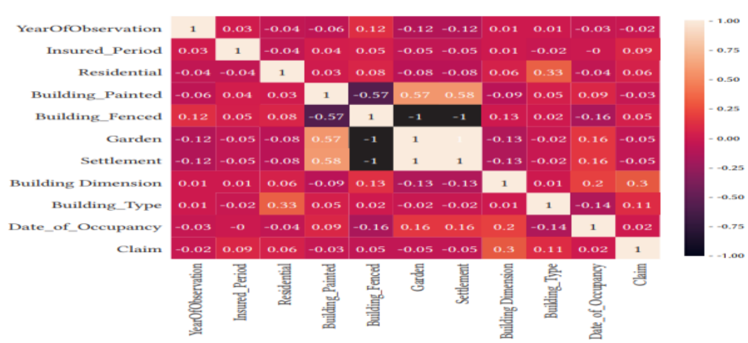

Fig. 3. Figure Correlation matrix

In Figure 3, the heatmap shows the correlation coefficients between different features. The features listed are similar to those in the bar chart and also include 'Claim'. The color scale on the right side of the heatmap ranges from -1.0 to +1.0, indicating strong negative to strong positive correlations. This visualization helps in understanding the relationships between different features, showing which pairs of features are positively or negatively correlated, thereby aiding in identifying potential multicollinearity or opportunities for feature engineering.

B. Data preprocessing

Data cleaning and labelling come after data collection, after which comes data preparation [21][22]. The data collected from the logs could be inconsistent, partial, or noisy, hence data preparation is a fundamental and important part of the knowledge discovery process [23]. Eliminating inconsistencies in the training data improves the effectiveness of the mining algorithm, which is affected by their presence [24]. The following steps of pre-processing are as:

- Handling missing values: Handling missing values is a critical step in data preprocessing that involves identifying and addressing gaps in the dataset. This can be done by removing data points with missing values, filling them with statistically relevant numbers (like the mean or median), or using prediction models to estimate the missing values. The probability of missing data affects the reliability of the results and requires the development of methods for their correct management.
- Identify and Eliminate outliers: Some of the data which have been deemed as outlying, will be rejected so as to develop a better prediction model.

1) One-hot encoding to the categorical data

Ordinal data is processed by this technique to develop a binary vector representation, where each category has its unique binary code. Implementation of this function assists machine learning algorithms to learn and use categorical data correctly by deporting nonexistent ordinal relations.

2) Feature selection/Impotence

This technique involves the procedures for choosing the features to be used in training of the model. Another goal is to minimize a number of input variables with regard to 'useful' or essential constituents most contributing to the required outcome and increasing the model's effectiveness, as well as eliminating the high calculation expenses in general.

C. SMOTE for data balancing

SMOTE creates artificially created instances in the minority class in order to equalize the distribution of classes. This technique solves a major disadvantage of unbalanced datasets and makes classifiers have better generalizability.

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

D. Data slicing

The preprocessed data is divided into train and test datasets, with a representative ratio of train data and test data are 70:30.

E. Classification models

This section describes the use of various classification models for comparative analysis in determining property and casualty insurance premiums.

1) Random Forest

The next section will describe parallel ensembling techniques, of which random forest is one kind of ML technique. A method for training parallel trees known as bagging is the foundation upon which it rests. Bagging involves using data produced by bootstrap aggregating to construct trees. This method selects numerous random samples from the original data set and replaces them[25]. Every tree ˆf is trained using a subsample B. Thus, many subsamples with modest variations are used to train each tree. Mathematical expression is given below in the equation (1):

$$f_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} f^{*b}(x) \quad \text{.............(1)}$$

The last prediction in a classification tree is just the mean of all the previous ones, as shown in (1). Additionally, to reduce overfitting to training data, the random forest method makes use of a few hyperparameters. These determine how many characteristics should be taken into account at each split, how many trees the model utilizes, and the maximum depth of every tree[26] [27].

2) SVM

The basic principle behind the SVM technique is the estimation of maximal margins [28][29]. It is the goal of the algorithm to determine, as far away from the class data points as possible, a hyperplane (decision boundary) that connects all of the classes. To construct the support vectors, we use the data points that are geographically nearest to the hyperplane. Support vectors are used to optimize the classifier's margin since they influence the hyperplane's location and orientation.

F. Model Evaluation

Different assessment measures were used to measure the performance of each model: F1-score, recall, accuracy, and precision. These key parameters are outlined below:

1) Confusion Matrix

A classification algorithm's performance may be defined using a confusion matrix, a table that displays the results. Confusion matrices are useful for visualizing and summarizing a

iJCEM
International Journal of Core
Engineering & Management
International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

classification algorithm's performance. A classifier's evaluation metrics are defined by the confusion matrix, which has four primary properties represented by numbers. The following four figure 4 are:

- True Positives (TP): A model correctly predicts a claim, and the claim is filed.
- True Negatives (TN): The model's prediction that no claim would be lodged is spot on.
- False Positives (FP): There has been no filing of the suit that the model predicts.
- False Negatives (FN): Though a claim is submitted, the model forecasts no claim.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Fig. 4. Confusion matrix

The aforementioned TP, TN, FP, and FN are the foundation upon which the algorithmic performance measures of recall, accuracy, precision, and F1 score are computed.

a) Accuracy
The amount of accurate predictions made by a classification model is called its accuracy. The corresponding equation (2) is shown below:

$$Acuracy = \frac{TP+TN}{TP+TN+FN+FP} \text{.........(2)}$$

b) Precision
The precision measures the proportion of accurately predicted positive data points among all the positive data points that the classifier anticipated. The Precision metric is denoted, as shown in equation (3) below:

$$Precision = \frac{TP}{TP+FP} \text{.........(3)}$$

c) Recall
The ratio of accurately anticipated positive data points to all positive data points is known as recall. The recall metric is expressed, as shown in equation (4) below:

$$Recall = \frac{TP}{TP+FN} \text{..........(4)}$$

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

d) F1-score

A harmonic mean of recall and precision is the F1-Score. Equivalent to equation (5), this measure is superior for assessing model performance because it takes recall and precision into account:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \quad \ldots\ldots\ldots\ldots(5)$$

Following measures evaluated a performance of ML models for an insurance prediction.

IV. RESULTS ANALYSIS AND DISCUSSION

This section presents the comparative outcomes of different ML models applied to property and casualty insurance. An accompanying Table III details a comparison of different models, analyzing their performance in terms of key metrics for Property & Casualty Insurance Premiums.

TABLE III. RF Model Performance for Property and Casualty Insurance Premiums.

Performance Matrix	Before SMOTE	Before SMOTE
Accuracy	99	99.99
Precision	80	80
Recall	92	96
F1-score	96	87

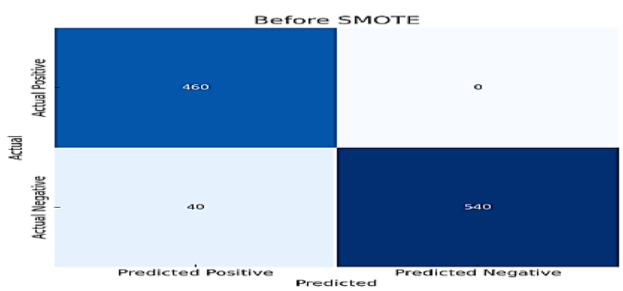

Fig. 5. Confusion matrix before SMOTE for random forest model.

Figure 5 confusion matrix displays a model's performance prior to applying SMOTE. It demonstrates a highly accurate prediction for both classes, with 460 true positives and 540 true negatives. The model exhibits no false negatives, but it incorrectly predicts 40 false positives.

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

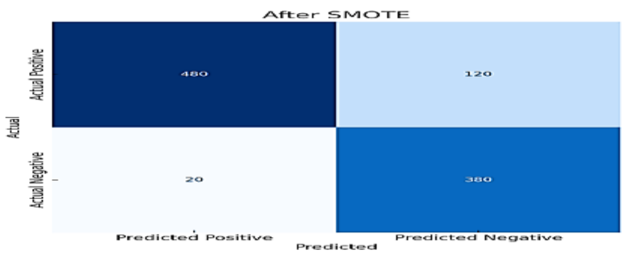

Fig. 6. Confusion matrix after SMOTE for random forest model.

In Figure 6, confusion matrix illustrates the model's performance after implementing SMOTE. There is a notable shift in the distribution of predictions. The true positives slightly increase to 480, indicating a minor improvement in identifying positive cases. However, the false positives also increase significantly to 120, indicating more instances being incorrectly labeled as positive. The true negatives decrease to 380, and the false negatives are reduced to 20, showing an improvement in the detection of negative cases but at the cost of reduced precision in positive predictions.

TABLE IV. Comparison between various models based on property and casualty insurance premiums.

Performance Matrix	RF	SVM[30]
Accuracy	99.99	81.34
Precision	80	79.22
Recall	96	82.98
F1-score	87	80.93

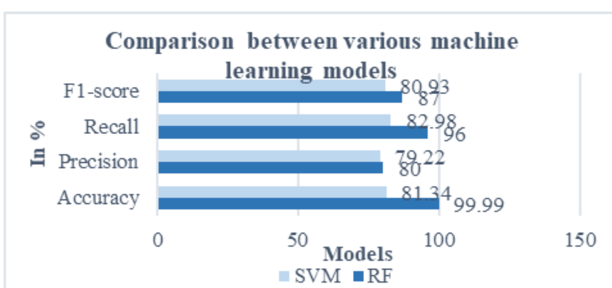

Fig. 7. Comparison between various machine learning models for insurance prediction

In a comparative evaluation of ML models applied to property and casualty insurance, illustrated in Table IV and Figure 7, the RF model emerges as the most effective. It demonstrates exceptional performance, achieving an accuracy rate of 99.99%, with precision and recall rates of 80% and 96%, respectively, leading to an F1-score of 87%. The Support Vector Machine (SVM) model also performs well, registering an accuracy of 81.34%, along with a precision of 79.22%, a recall of 82.98%, and an F1-score of 80.93%. These figures highlight the RF model's

International Journal of Core Engineering & Management
Volume-7, Issue-10, 2024 ISSN No: 2348-9510

robustness in handling the intricate data patterns found in insurance datasets. Compared to the reliable SVM, the RF model's superior scores across all key metrics solidify its status as the most dependable choice for modeling in the property and casualty insurance sector.

V. CONCLUSION AND RECOMMENDATIONS

The proposed seventh business function is "business expansion," which involves forming strategic alliances between security firms and property and casualty insurance providers. These alliances will use value chain analysis and contracting strategies to meet the integrated service needs of consumers. The results of this research show that SVMs and RF are the most effective ML algorithms for forecasting premiums for property and liability insurance. Challenges like imbalanced data with treatments as SMOTE and appropriate feature selection have demonstrated that the RF model possesses high accuracy, precision, re-call, and F1score of 99%, 80%, 96%, and 87%, respectively. According to this performance, of RF model achieved high results in comparison to other SVM models that get only 81% accuracy. The RF model emerged as the best-performing algorithm, providing reliable and accurate predictions based on key insurance variables. These results underline the potential of ML to transform the way insurers assess risk and determine premiums, leading to more accurate pricing strategies and improved risk management.

Despite the promising results, future work should improve model interpretability for regulatory compliance and enhance transparency. Exploring deep learning models and hybrid approaches may boost predictive accuracy. Expanding the dataset and incorporating real-time processing will develop more scalable models. Additionally, applying these techniques to other insurance types could broaden their effectiveness and insights.

REFERENCES

1. N. Benali and R. Feki, “The impact of natural disasters on insurers’ profitability: Evidence from Property/Casualty Insurance company in United States,” Res. Int. Bus. Financ., 2017, doi: 10.1016/j.ribaf.2017.07.078.
2. M. R. Kishore Mullangi, Vamsi Krishna Yarlagadda, Niravkumar Dhameliya, “Integrating AI and Reciprocal Symmetry in Financial Management: A Pathway to Enhanced Decision-Making,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 5, no. 1, pp. 42–52, 2018.
3. R. Goyal, “The Role Of Business Analysts In Information Management Projects,” Int. J. Core Eng. Manag., vol. 6, no. 9, pp. 76–86, 2020.
4. C. Blier-Wong, H. Cossette, L. Lamontagne, and E. Marceau, “Machine Learning in P&C Insurance: A Review for Pricing and Reserving,” Risks, vol. 9, no. 1, p. 4, Dec. 2020, doi: 10.3390/risks9010004.

5. A. J. Dey and H. K. D. Sarma, “A Survey on Application of Machine Learning in Property and Casualty Insurance,” Contemp. Issues Commun. Cloud Big Data Anal., pp. 307–314, 2021.
6. S. Gupta and A. Mathur, “Enhanced flooding scheme for AODV routing protocol in mobile ad hoc networks,” in Proceedings - International Conference on Electronic Systems, Signal Processing, and Computing Technologies, ICESC 2014, 2014. doi: 10.1109/ICESC.2014.60.
7. S. Pandey, “TRANSFORMING PERFORMANCE MANAGEMENT THROUGH AI: ADVANCED FEEDBACK MECHANISMS, PREDICTIVE ANALYTICS, AND BIAS MITIGATION IN THE AGE OF WORKFORCE OPTIMIZATION,” Int. J. Bus. Quant. Econ. Appl. Manag. Res., vol. 6, no. 7, pp. 13–19, 2020.
8. A. P. A. Singh and N. Gameti, “Streamlining Purchase Requisitions and Orders : A Guide to Effective Goods Receipt Management,” J. Emerg. Technol. Innov. Res., vol. 8, no. 5, pp. g179–g184, 2021.
9. K. K. SKR Anumandla, VK Yarlagadda, SCR Vennapusa, “Unveiling the Influence of Artificial Intelligence on Resource Management and Sustainable Development: A Comprehensive Investigation,” Int. J. Creat. Res. Thoughts, vol. 9, no. 12, pp. f573–f578, 2020.
10. J. Thomas, “Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics,” J. Emerg. Technol. Innov. Res., vol. 8, no. 9, 2021.
11. B. Knox and J. A. Sørensen, “Asset-Driven Insurance Pricing,” SSRN Electron. J., 2023, doi: 10.2139/ssrn.4394474.
12. M. A. Khan, J. E. Soh, M. Maenner, W. W. Thompson, and N. P. Nelson, “A machine-learning algorithm to identify hepatitis C in health insurance claims data,” Online J. Public Health Inform., 2019, doi: 10.5210/ojphi.v11i1.9685.
13. Suhag Pandya, “A Machine and Deep Learning Framework for Robust Health Insurance Fraud Detection and Prevention,” Int. J. Adv. Res. Sci. Commun. Technol., pp. 1332–1342, Jul. 2023, doi: 10.48175/IJARSCT-14000U.
14. J. Kiviniemi and M. Leppanen, “EMPLOYEE WELL-BEING AS COMPETIVE ADVANTAGE Case If P & C Insurance,” Univ. Appl. Sci., 2012.
15. V. V. Kumar, F. W. Liou, S. N. Balakrishnan, and V. Kumar, “Economical impact of RFID implementation in remanufacturing: a Chaos-based Interactive Artificial Bee Colony approach,” J. Intell. Manuf., 2015, doi: 10.1007/s10845-013-0836-9.
16. C. Jyothsna, K. Srinivas, B. Bhargavi, A. E. Sravanth, A. T. Kumar, and J. N. V. R. S. Kumar, “Health Insurance Premium Prediction using XGboost Regressor,” in Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022, 2022. doi: 10.1109/ICAAIC53929.2022.9793258.
17. S. Panda, B. Purkayastha, D. Das, M. Chakraborty, and S. K. Biswas, “Health Insurance Cost Prediction Using Regression Models,” in 2022 International Conference on Machine

Learning, Big Data, Cloud and Parallel Computing, COM-IT-CON 2022, 2022. doi: 10.1109/COM-IT-CON54601.2022.9850653.

18. K. Dutta, S. Chandra, M. K. Gourisaria, and H. Gm, "A Data Mining based Target Regression-Oriented Approach to Modelling of Health Insurance Claims," in Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, 2021. doi: 10.1109/ICCMC51019.2021.9418038.
19. D. Utomo, N. Damanik, and I. Budi, "Classification on Participants Renewal Process in Insurance Company: Case Study PT XYZ," in 2021 9th International Conference on Information and Communication Technology, ICoICT 2021, 2021. doi: 10.1109/ICoICT52021.2021.9527479.
20. R. Sun, Q. Cheng, F. Xie, W. Zhang, T. Lin, and W. Y. Ochieng, "Combining Machine Learning and Dynamic Time Wrapping for Vehicle Driving Event Detection Using Smartphones," IEEE Trans. Intell. Transp. Syst., 2021, doi: 10.1109/TITS.2019.2955760.
21. S. Sharma and A. Bhagat, "Data preprocessing algorithm for Web Structure Mining," in Proceedings on 5th International Conference on Eco-Friendly Computing and Communication Systems, ICECCS 2016, 2017. doi: 10.1109/Eco-friendly.2016.7893249.
22. P. Shwet and S. Pandey, "A COMPREHENSIVE FRAMEWORK FOR JOB DESCRIPTION MANAGEMENT: INTEGRATION OF WORKDAY, BOX, AND GREENHOUSE FOR ENHANCED COMPLIANCE AND EFFICIENCY," vol. 7, no. 1, 2021, pp. 13–19, 2021.
23. K. Patel, "Quality Assurance In The Age Of Data Analytics: Innovations And Challenges," Int. J. Creat. Res. Thoughts, vol. 9, no. 12, pp. f573–f578, 2021.
24. S. Samsani, "An RST based efficient preprocessing technique for handling inconsistent data," in 2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016, 2017. doi: 10.1109/ICCIC.2016.7919591.
25. R. Tarafdar, "Algorithms on Majority Problem," Univ. Missouri-Kansas City, 2017.
26. G. James, D. Witten, T. Hastie, and R. Tibishirani, An Introduction to Statistical Learning with Applications in R (older version). 2013.
27. S. Dixit, P. Pathak, and S. Gupta, "A novel approch for gray hole and black hole detection and prevention," in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016. doi: 10.1109/CDAN.2016.7570861.
28. J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," in Journal of Physics: Conference Series, 2018. doi: 10.1088/1742-6596/1142/1/012012.
29. S. Rathi, V. S. Rajput, and S. Gupta, "An improved evolution based optimization algorithm originated from the concept of SFLA and simulated annealing," in 11th International Conference on Industrial and Information Systems, ICIIS 2016 - Conference Proceedings, 2016. doi: 10.1109/ICIINFS.2016.8262933.
30. M. K. Severino and Y. Peng, "Machine learning algorithms for fraud prediction in property insurance:," Mach. Learn. with Appl., 2021, doi: 10.1016/j.mlwa.2021.100074.