ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

Exploring the Potential of Machine Learning Techniques for Predicting Travel Insurance Claims: A Comparative Analysis of Four Models

Xiaonan Li*

Central University of Finance and Economics, Beijing, China
**Corresponding author: lxn_cufejs@126.com*

Abstract: *Travel insurance is a crucial component for any traveller as it offers protection against financial losses resulting from unforeseen events during a trip, such as trip cancellations, medical emergencies, lost luggage, and related issues. This study aims to investigate the potential of machine learning (ME) techniques for predicting the probability of travel insurance claims. In order to tackle the issue of managing extensive and intricate datasets, advanced statistical techniques were employed, including keyword extraction, feature extraction, and Chi-squared tests. Our evaluation of four popular ML models, namely balanced random forest (BRF), support vector machines (SVM), logistic regression (LR), and balanced bagging (BB), highlight that the BRF model outperforms the other models in predicting travel insurance claims. Our study emphasises the advantages of utilising machine learning algorithms in processing large datasets, producing predictions on future insurance claims, and adapting to changing circumstances, thus serving as a valuable tool for practitioners in the travel insurance industry.*

Keywords: *machine learning, travel insurance claims, BRF, LR, SVM, balanced bagging*

1. Introduction

Predicting the likelihood of travel insurance claims is a crucial task for both travel insurance companies and policyholders [1, 2]. Travel insurance claims may include medical emergencies, trip cancellations or interruptions, lost or stolen luggage, and other unforeseen circumstances [3, 4]. Accurate predictions enable insurance companies to set premiums that reflect the risk of insuring a particular traveller or trip, while also helping them maintain profitability by covering the cost of any claims made [5, 6]. This not only benefits insurance companies but also provides peace of mind to policyholders, ensuring they are adequately covered and protected during their travels [7].

Anticipating insurance claims is a crucial task for insurance companies, especially in the travel insurance industry, which has seen a surge in popularity [8-10]. In order to accurately estimate potential losses and develop effective risk management strategies, insurers must be able to navigate the intricate nature of the data involved [11, 12]. Travel insurance claims data can include a wide range of features, such as policyholder demographics, travel specifics, and claims history. However, the complexity of this data can pose a challenge for ML models in determining the most relevant features and correlations [13, 14]. In order to address the challenge of handling complex datasets, ML algorithms can leverage feature selection methods to identify critical features [15-17]. These methods enable the model to pinpoint the most relevant features that have the greatest impact on predictions while disregarding irrelevant or redundant features [18-20]. This ultimately enhances the model's accuracy and reduces the risk of overfitting or underfitting. Moreover, insurance claims data are particularly susceptible to inaccuracies and discrepancies that can further complicate the task of building accurate machine learning models [21, 22]. Data entry errors, missing values, or outliers can distort the data and result in inaccurate forecasts [23, 24]. To address this issue, ML algorithms can incorporate outlier detection and data cleaning techniques to refine prediction accuracy [25-27]. Outlier detection techniques can identify and eliminate data points that significantly deviate from the rest of the dataset, allowing the model to avoid being influenced by extreme values that could skew the predictions. Meanwhile, data cleaning techniques can detect and correct data errors, such as misspellings or missing values, resulting in accurate and reliable data that lead to more precise predictions. Overall, the use of feature selection methods, outlier detection, and data cleaning techniques can help insurance companies overcome the challenges of anticipating insurance claims [28-30]. By utilizing these tools, ML algorithms can identify critical features, minimize data complexity, and

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

refine the accuracy of predictions, ultimately enabling insurance companies to develop effective risk management strategies and reduce potential losses [31].

The objective of this investigation is to identify the most efficient ML model for anticipating claims in travel insurance. In order to address the challenges associated with data processing, multiple techniques were employed, including keyword extraction, feature reclassification, feature selection, data normalisation, and train-test splitting. Four widely recognised and effective models, namely, balanced random forest (BRF), support vector machines (SVM), logistic regression (LR), and balanced bagging (BB), were evaluated, as they are recognised for their effectiveness in various ML applications. By evaluating multiple models, the one that exhibited optimal performance on the dataset was able to be selected. This knowledge can be of significant value to insurance industry professionals seeking to construct accurate and dependable prediction models for travel insurance claims.

2. Research methodology

In this study, a ML model was proposed, implemented using Python programming language, that exhibits high accuracy in predicting travel insurance claims. Our research utilised data from a third-party travel insurance provider located in Singapore, consisting of 63,326 rows and 11 features. Each row represented a distinct traveller and their travel particulars. A diverse range of features was utilised to predict travel insurance claims, including numerical, categorical, and target variables. Specifically, numerical variables included commission, net sales, duration, and age, while the categorical variables comprised of agency, agency type, distribution channel, product name, destination, and gender. Additionally, a binary variable was utilised to represent whether a claim was filed or not. To develop the ML model, feature engineering, data pre-processing, and ML algorithms were utilised. The data were pre-processed and feature engineering was performed to transform the data into a suitable format for our model. Then the data were divided into training and testing sets to train and evaluate our model's performance. Various ML algorithms were utilised to optimise our model's accuracy and efficiency.

2.1. Data collection

The target variable for the analysis was binary, with a value of 1 indicating the occurrence of a claim and 0 denoting the absence of a claim. As illustrated in Figure 1, the dataset exhibits a significant class imbalance, with a significantly greater number of observations falling under the majority class (i.e., no claim) compared to the minority class (i.e., claim).

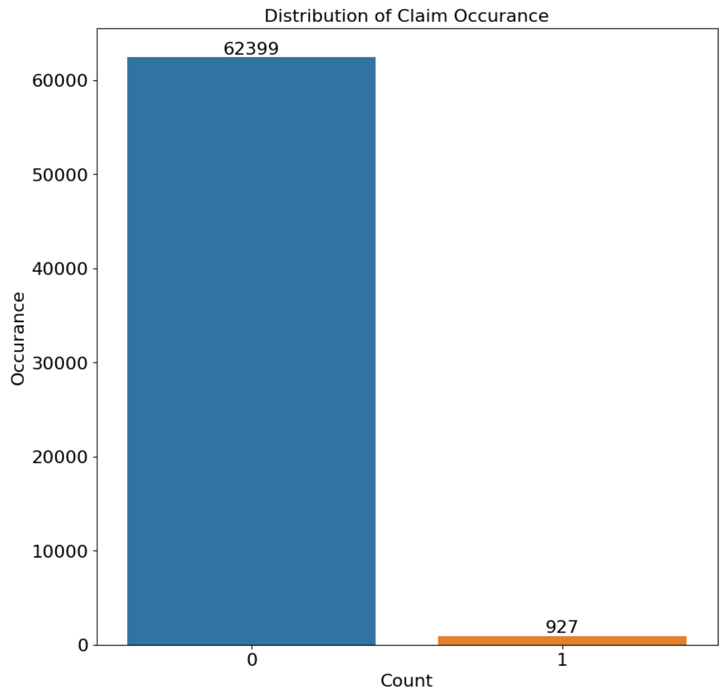

Figure 1: The target variable distribution

2.2. Data pre-processing

Evaluating the performance of a predictive model for insurance claims requires the accurate encoding of categorical features into numeric values, utilising techniques such as keyword extraction, feature reclassification, feature selection, and data normalization.

During the process of keyword extraction, it is recommended to identify and extract those keywords from the Product Name that can effectively represent the relevant product category. These keywords may include product level categories such as gold, silver, and bronze; product period categories such as 1 way, 2 way, and annual; customer type categories such as parents and child; travel transportation categories such as vehicle and cruise; and insurance type categories such as cancellation and comprehensive. Once these relevant keywords have been identified, they can be used as a new feature and encoded using the one-hot encoding method. Following the process of keyword extraction, the reclassification of features is recommended. Specifically, the 'destination' feature comprises of country names that can result in many possible values. To overcome this issue, it is suggested to reclassify the values based on their geographic location. Examples of such classifications include East Asia, West Asia, Southeast Asia, Eastern Europe, Western Europe, Central Europe, and other similar categories. Once this reclassification has been completed, the one-hot encoding technique can be applied to this feature. After performing keyword extraction and feature reclassification, the resulting dataset consisted of 63,326 rows and 65 columns. The next step was feature selection, where the Chi-squared (χ^2) test was used to assess the independence between categorical variables. This statistical method measures the degree of association between two categorical variables and helps to identify whether any relationship exists between them. The Chi-squared value for each feature was calculated using the widely used scikit-learn library. Following this step, categorical features were filtered based on their Chi-squared score, with only those features having a score greater than 20 being retained for further analysis. The resulting Chi-squared scores for these selected features are presented in Table 1. Ultimately, the final dataset comprised 63,326 rows and 24 columns. Following the feature selection process, the next step was data normalisation. In this step, all the retained features were normalised using Standard Scaler. Standard Scaler is a commonly used normalisation technique that scales features so that they have a mean of zero and a standard deviation of one. This technique helps to reduce the impact of outliers and can improve the accuracy of ML models by ensuring that all features contribute equally to the analysis. To facilitate accurate modelling and analysis, the dataset was partitioned into training and testing sets using the StratifiedKFold method, with a standard split ratio of 80% for training and 20% for testing. Specifically, the resulting dataset was comprised of 50,660 rows and 24 columns for training data, and 12,666 rows and 24 columns for test data.

Table 1: Chi-Squared scores of categorical features selected for analysis.

Categorical features	Chi Squared Score
Agency_C2B	1521.751657
Silver	1088.685331
Annual	1052.137490
Agency Type_Airlines	446.895340
Bronze	389.000439
Gender_F	365.008401
Gender_unknown	213.034542
Agency_EPX	200.998646
Cancellation	194.664511
Gender_M	180.904644
Gold	176.791891
Agency Type_Travel Agency	170.081143
Agency_LWC	67.570642
Destination_South-East Asia	43.068297
Agency_JZI	41.629366
Basic	41.269988
Destination_East Asia	35.223065
1 way	32.903795
Comprehensive	31.771393
Single	23.736891

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

2.3. Training and evaluating BRF, SVM, LR, and BB

In this study, four different models were employed to predict travel insurance claims.

The utilisation of the BRF model can be a highly advantageous approach for analysing travel insurance claims. With class imbalance being a common problem in this field due to the low frequency of claim occurrence, the BRF model can effectively address this issue through a combination of undersampling and bootstrap aggregating techniques. This enables the model to balance the dataset and reduce the impact of noise and outliers, leading to improved model performance, particularly in critical metrics such as recall and F1-score. Moreover, the BRF model is capable of handling high-dimensional datasets and non-linear relationships, which makes it a promising solution for complex insurance claim analysis problems. As such, the use of the BRF model can greatly enhance the accuracy and efficiency of travel insurance claim analysis, providing valuable insights for both insurers and customers alike.

In addition to the BRF model, another reliable ML algorithm for insurance claim analysis is the SVM model. The SVM model was trained using a carefully selected kernel function and a regularisation parameter and was evaluated using multiple metrics. By feeding the model various features such as policyholder data and incident details, the SVM model was able to predict the legitimacy of an insurance claim, providing a robust and automated solution for insurance companies seeking to assess the validity of claims.

Another model that was employed in this study is LR, which is a commonly used statistical method for analysing datasets where one or more independent variables determine an outcome. LR is like linear regression, but differs in that it computes the probability of the predicted instance being positive, as illustrated in Equation (1):

$$P = \frac{e^{b_0+b_1X}}{1+e^{b_0+b_1X}} \quad (1)$$

where P is the predicted probability of the instance, it is positive, e is the mathematical constant, b_0 is the intercept term, b_1 is the coefficients, and X is the feature values vector.

The linear function of X activated by sigmoid function. The sigmoid function is used to scale P to [0, 1]. As a result, the algorithm utilises both the P value and the threshold value to make predictions, as shown in equation (2).

$$\hat{y} = \begin{cases} 1 & if\ P \geq t \\ 0 & if\ P < t \end{cases} \quad (2)$$

where $\hat{y}$ is the predicted class, 1 represents the instance is predicted as positive (claim will occur), 0 represents the instance is predicted as negative (claim will not occur), and t is the threshold.

The logistic regression (LR) model was also explored. Prior to utilising LR, the synthetic minority oversampling technique (SMOTE) was applied to address the issue of imbalanced data and transform it into a balanced dataset. The LR model is a powerful tool that enables the computation of the probability of the dependent variable, which could either be a claim or no claim, being equal to one. This probability serves as a basis for predicting whether a claim will be made, and it is commonly determined by employing a threshold value of 0.5. LR is particularly well-suited for binary outcomes, such as car insurance claims, owing to its simplicity of use and exceptional predictive accuracy. One of the significant benefits of LR is that it allows for the interpretation of model parameters in terms of odds ratios, which can provide valuable insights into the impact of various features on the outcome. By analysing these odds ratios, researchers and practitioners can identify the factors that are most influential in determining whether a claim will be made.

The fourth test model is the BB algorithm, which is recognised as a highly effective ML technique for analysing travel insurance claims. This algorithm is a type of ensemble learning method that leverages the strength of multiple classifiers to produce more accurate predictions. In the context of predicting travel insurance claims, the BB model can enhance the accuracy of predictions by considering various factors that may impact whether a claim is filed or not. The model can more effectively identify patterns and make informed predictions, ultimately leading to more accurate and reliable results. As such, the BB model represents a promising approach for enhancing the efficiency and effectiveness of travel insurance claim analyses.

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

2.4. Measurements

Accuracy is a commonly used measure to evaluate the effectiveness of classification models. However, it is not suitable for models trained on imbalanced datasets. In such cases, metrics like precision, recall, AUC-ROC, and F1-score are more appropriate to comprehensively assess the model's performance. To further evaluate the model's effectiveness on imbalanced data, specificity and sensitivity plots are often created. In travel insurance occurrence prediction, a confusion matrix can be employed to evaluate the model's performance, consisting of four entries representing true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions made by the classifier.

● True Positives (TP), indicating the number of instances where the classifier correctly predicted the occurrence of a claim.

● False Positives (FP), representing the number of instances where the classifier incorrectly predicted the occurrence of a claim that did not actually occur.

● True Negatives (TN), indicating the number of instances where the classifier correctly predicted the absence of a claim.

● False Negatives (FN), representing the number of instances where the classifier incorrectly predicted the absence of a claim that did in fact occur.

Using the values in the confusion matrix, it is possible to calculate several key performance metrics, including accuracy, precision, recall, specificity, and F1-score. The formulas for computing these metrics are presented in Table 2.

Table 2: Metrics and computing method

Metrics	Computing method
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
F1-score	$\frac{2\times TP}{2\times TP+FP+FN}$

3. Results and discussion

The usefulness of the confusion matrix in assessing the performance of binary classifiers is presented in Table 3, where models (BRF, SVM, LR, and BB) are evaluated. The matrix displays the number of predictions made by the classifier for each class (either negative or positive) in each cell. The diagonal elements of the matrix correspond to the number of TP and TN, while the off-diagonal elements indicate the number of FP and FN.

Table 3: Confusion matrix for binary classifier models.

Models		Predicted Negative	Predicted Positive
BRF	Truth Negative	9339	3147
	Truth Positive	39	141
SVM	Truth Negative	9680	2806
	Truth Positive	49	131
LR	Truth Negative	9814	2672
	Truth Positive	47	133
BB	Truth Negative	10516	1970
	Truth Positive	65	115

In classification tasks, evaluation of models is typically based on the accuracy for each class, as computed according to the method presented in Table 2. Table 4 presents the precision, recall, F1 score, and ROC-AUC metrics for the models, specifically with respect to their performance on predictions related to the minority class.

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

Table 4: Evaluation of Minority Class Predictions: Precision, Recall, F1-Score and ROC-AUC.

Model	Precision	Recall	Specificity	F1-Score	Accuracy	AUC-ROC
BRF	4.29%	78.33%	74.80%	8.13%	74.85%	76.56%
SVM	4.46%	72.78%	77.53%	8.41 %	77.46%	75.15%
LR	4.74%	73.89%	78.60%	8.91%	78.53%	76.24%
BB	5.52%	63.89%	84.22%	10.15%	83.93%	74.06%

Table 4 provides clear evidence that the BB model excels compared to the other models in precision (5.52%), indicating that it has the capability to accurately predict a large percentage of true positives. However, the BB model's relatively low recall (63.89%) suggests that it fails to identify a significant number of positive instances, leading to a high number of false negatives. In contrast, the BRF model displays the highest recall (78.33%), successfully detecting a considerable number of claims that will occur. However, its precision is relatively low, indicating that it tends to produce a substantial number of false positives. In terms of specificity, it is evident that the BB model has the highest score (74.80%), indicating its ability to accurately identify a significant number of claims that will not occur. However, its relatively low recall implies that it is likely to produce a considerable number of false negatives, mistakenly classifying positive cases as negative ones. The F1-score, a metric that considers both precision and recall, confirms that the BB model obtains the highest score (83.93%), signifying that it has a good balance between the high precision and recall. Conversely, the BRF model yields the lowest F1-score (74.85%), indicating that it has the weakest capacity to balance the trade-off between the precision and recall. The BB model demonstrated the highest accuracy at 83.93%, but it is important to note that its performance was largely due to the high number of true negatives in the confusion matrix. Conversely, while the BRF model had the lowest accuracy among all models, its recall was the highest. The lower accuracy of the BRF model could be attributed to the significant number of negative instances in the dataset, leading to a higher number of negative samples being incorrectly predicted as positive. It is worth noting that when dealing with imbalanced datasets, accuracy may not be a sufficient metric for evaluating a model's performance. As such, it is important to consider other evaluation measures such as recall, precision, and F1-score, which provide a more comprehensive assessment of a model's ability to predict both positive and negative instances accurately. The AUC-ROC is a well-established evaluation metric for binary classifiers. Figure 2 displays the AUC-ROC curve of four models on both training and testing datasets. The graph indicates that the BRF model has the most superior AUC-ROC score, suggesting that it outperforms the other models. The accuracy of the models ranges between 73.06 % to 76.56%.

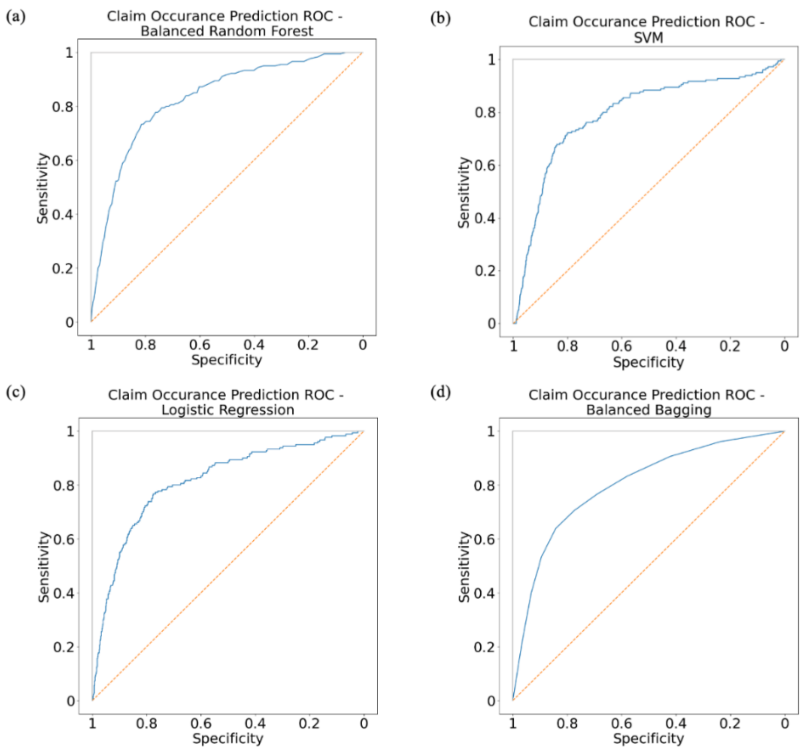

Figure 2: Comparison of AUC-ROC Performance of four Models on Training and Testing Datasets: (a) BRF, (b) SVM, (c) LR, and (d) BB.

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

In summary, selecting the best model requires careful consideration of the business constraints and costs associated with false positives and false negatives. Based on our analysis, the BRF and LR models demonstrate a desirable balance of high precision and recall, making them ideal for the given task. However, when considering the overall performance, the BRF model is recommended as it outperformed the other models with the highest AUC-ROC score. Therefore, it is crucial to evaluate these factors thoroughly to ensure the selected model aligns with the specific requirements of the business and maximises its performance.

4. Conclusion

In conclusion, this study has developed and implemented a highly accurate ML model using Python programming language to predict travel insurance claims. The model utilises a diverse range of features, including numerical, categorical, and target variables, and was trained and evaluated using data from a reputable third-party travel insurance provider based in Singapore. The study demonstrates the effectiveness of the confusion matrix in evaluating the performance of binary classifiers, as four models (BRF, SVM, LR, and BB) were evaluated using precision, recall, F1 score, and ROC-AUC metrics for predicting the minority class. The following are the findings of the study:

1) The data pre-processing, feature engineering, and various ML algorithms were employed to optimise the model’s accuracy and efficiency.

2) The results indicate that the BB model excels in precision, while the BRF model has the highest recall. The BB model also has the highest specificity and F1 score, demonstrating a good balance between the precision and recall.

3) The BRF model outperforms the other models with the highest AUC-ROC score, indicating its superiority in binary classification.

4) A thorough evaluation of these factors is necessary to ensure the selected model aligns with the specific requirements of the business and maximises its performance.

References

[1] M. A. Rubi, M. H. I. Bijoy, S. Chowdhury, and M. K. Islam, "Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior," in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022: IEEE, pp. 1-5.

[2] D. A. Hamzah, "Predicting travel insurance policy claim using logistic regression," Applied Quantitative Analysis, vol. 1, no. 1, pp. 1-7, 2021.

[3] P. A. Leggat and F. W. Leggat, "Travel insurance claims made by travelers from Australia," Journal of Travel Medicine, vol. 9, no. 2, pp. 59-65, 2002.

[4] J. T. Gamaliel and H. Murfi, "Regularization learning network for insurance claim prediction in travel insurance," Journal of Advanced Research in Dynamical and Control Systems, vol. 12, no. 4 Special Issue, pp. 1496-1503, 2020.

[5] C. Leboeuf, P. Gagnon, and M. M. Hébert, "Travel Insurance Claims modeling using Statistical Learning Mitacs Acceleration Research Project," 2020.

[6] Z. Quan and E. A. Valdez, "Predictive analytics of insurance claims using multivariate decision trees," Dependence Modeling, vol. 6, no. 1, pp. 377-407, 2018.

[7] G. Kerr and L. Kelly, "Travel insurance: the attributes, consequences, and values of using travel insurance as a risk-reduction strategy," Journal of Travel & Tourism Marketing, vol. 36, no. 2, pp. 191-203, 2019.

[8] I. M. Njoh-Paul, "A Comparative Study of Ensemble Techniques and Individual Classifiers in Predicting Insurance Claim," Dublin, National College of Ireland, 2020.

[9] A. Al Mamun, M. K. Rahman, Q. Yang, T. Jannat, A. A. Salameh, and S. A. Fazal, "Predicting the willingness and purchase of travel insurance during the COVID-19 pandemic," Frontiers in public health, vol. 10, 2022.

[10] R. F. Grace and D. Penny, "Travel insurance and medical evacuation: view from the far side," Medical journal of Australia, vol. 180, no. 1, pp. 32-35, 2004.

[11] N. Zaitseva and L. Chernikova, "Features and Prospects in the Development of the Services Provided in the Field of Travel Insurance," Middle East Journal of Scientific Research, vol. 16, no. 7, pp. 996-1002, 2013.

[12] L. Jia-lan, W. Xiao-yu, Y. Wan-jun, W. Zi-chen, Z. Huai-lin, and C. Nai-meng, "Research and design

ISSN 2616-5775 Vol. 6, Issue 4: 118-125, DOI: 10.25236/AJCIS.2023.060416

of travel insurance system based on blockchain," in 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2019: IEEE, pp. 121-124.

[13] F. B. Cappon, "Travel Insurance: The Urgent Need For Improved Regulation," 2014.

[14] Z. Rybák, "Analysis of the individual travel insurance in the Czech Republic," European Financial and Accounting Journal, vol. 13, no. 2, pp. 5-30, 2018.

[15] S. Rawat, A. Rawat, D. Kumar, and A. S. Sabitha, "Application of machine learning and data visualization techniques for decision support in the insurance sector," International Journal of Information Management Data Insights, vol. 1, no. 2, p. 100012, 2021.

[16] S. K. Vandrangi, "Predicting The Insurance Claim By Each User Using Machine Learning Algorithms," Journal of Emerging Strategies in New Economics, vol. 1, no. 1, pp. 1-11, 2022.

[17] P. C. Verhoef and B. Donkers, "Predicting customer potential value an application in the insurance industry," Decision support systems, vol. 32, no. 2, pp. 189-199, 2001.

[18] A. Panchapakesan, R. Abielmona, R. Falcon, and E. Petriu, "Prediction of container damage insurance claims for optimized maritime port operations," in Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings 31, 2018: Springer, pp. 265-271.

[19] C. A. Duah, "Predictive Modeling of Insurance Claims Using Reversible Jump Markov Chain Monte Carlo Methods," 2017.

[20] N. Nawaratana, "Analysis of distributions for insurance claims data," School of Mathematics Institute of Science Suranaree University of Technology, 2019.

[21] J.-M. Kim, J. Kim, and I. D. Ha, "Application of Deep Learning and Neural Network to Speeding Ticket and Insurance Claim Count Data," Axioms, vol. 11, no. 6, p. 280, 2022.

[22] A. J. Dey and H. K. D. Sarma, "A Survey on application of machine learning in property and casualty insurance," in Contemporary Issues in Communication, Cloud and Big Data Analytics: Proceedings of CCB 2020, 2022: Springer, pp. 307-314.

[23] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," Computational and structural biotechnology journal, vol. 15, pp. 104-116, 2017.

[24] R. P. Bellapu, S. Mylsamy, S. R. Krishna, and P. Kundu, "Developing and evaluation of machine learning models in the insurance sector," 2021.

[25] N. Sharma, S. K. Gautam, A. A. Henry, and A. Kumar, "Application of big data and machine learning," Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications, pp. 305-333, 2020.

[26] M. Riikkinen, H. Saarijärvi, P. Sarlin, and I. Lähteenmäki, "Using artificial intelligence to create value in insurance," International Journal of Bank Marketing, 2018.

[27] S. Tober, "Tree-based Machine Learning Models with Applications in Insurance Frequency Modelling," ed, 2020.

[28] K. P. Sinha, M. Sookhak, and S. Wu, "Agentless Insurance Model Based on Modern Artificial Intelligence," in 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), 2021: IEEE, pp. 49-56.

[29] S. Meng, Y. Gao, and Y. Huang, "Actuarial intelligence in auto insurance: Claim frequency modeling with driving behavior features and improved boosted trees," Insurance: Mathematics and Economics, vol. 106, pp. 115-127, 2022.

[30] M. Eling, D. Nuessle, and J. Staubli, "The impact of artificial intelligence along the insurance value chain and on the insurability of risks," The Geneva Papers on Risk and Insurance-Issues and Practice, pp. 1-37, 2021.

[31] B. Ljubic, Advanced Machine Learning Models in Prediction of Medical Conditions. Temple University, 2021.