ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

Prediction of Satisfaction Accuracy on Health Insurance Policy through Machine Learning Algorithms Especially Tree Models

Received: 26 February 2023, **Revised:** 30 March 2023, **Accepted:** 21 April 2023

Sayantani Ray* and Raghunath Datta
Department of Commerce and Management
Seacom Skills University
Kendradangal, Birbhum

Keywords

Machine leaning algorithms, Tree models, Health insurance policy, Consumer satisfaction

Abstract

This study was to predict the satisfaction accuracy on health insurance policy (HIP) among participants through machine learning (ML) algorithms especially tree models. We studied the data mining based on ML classifiers by using WEKA tool, version, 3.8.5. The study was conducted through questionnaires-based survey among 385 respondents of eastern India. The predictive accuracy of data of satisfaction on HIP through ML algorithms especially 4 tree algorithms viz. decision tree (DT) J48, Random forest (RF), Random tree (RT) and Fast decision tree learner tree (REPT) along with 9 attributes viz. Facilities, Claim_coverage, Tax_benefit, Unexceptional_risk, Trust_insurer, Policy_benefit_bonus, Policy_benefit_premium_amount, Amount_claim_offered_maturity, Policy_benefit_family_production and class (poor, moderate and good response) from dataset were determined. In our study, in the poor class a maximum value of precision recall curve (PRC) value was obtained as per the ML algorithms such as RF (96%) and RT (93%) followed by REPT (91%) and DT J48 (90%). It is concluded that the valuable information of the dataset through ML algorithms especially tree models are obtained prediction accuracy higher in RF and RT and lower in REPT and DTJ48 algorithms as per cross validation (CV) test. From this study, it was predicted poor satisfaction on HIP among participants. It is suggested to validate the present predictive data.

Introduction

Generally, machine learning (ML) algorithms play vital roles in the health insurance market such as chatbots, faster claim settlements, personalised HIP, cost-effectiveness, fraud detection, faster underwriting (Kaushik et al., 2022). Moreover, the insurance related to health sponsoring procedure is prolonged and taking more time phenomenal process. Now-a-days, adoption of AI and/or ML-based predictive consideration, health insurance organizations might be prevented time and money (Kaushik et al., 2022).

Rawat et al. (2021) mentioned that a complete analysis could be helped to detect fraud in insurance industries. According to investigator, insurance fraud was reported about 40 billion dollars in the industry in every year, for this reason, the use of ML algorithms-based prediction could be helpful to alert brokers in case of fraudulence activities. They analysed predictive tools for using to gather perceptions on consumer behaviour as well as to accumulate understandings on employees to maintain valuable ability. This could be achieved by understanding the performance, profits, learning styles of the traders as well as their trade satisfaction and the potential to pay attention regarding a job someplace else (Ozbayoglu et al., 2020; Sengupta et al., 2020). Additionally, AI and/or ML both could be used for the efficient marketing of HIP.

Moreover, the usage of ML in the predictive study is spreading tremendously in the insurance organization regardless of early resistance by the business because of its clearly recursive method in predictive modelling to get better model at each recurrence (Rawat et al., 2021). In this analysis, some studies dealt related to the claim scrutiny of the insurance business (Doupe et al., 2019; Dave et al., 2021). In earlier studies, ML is used in claim investigation and managing for triaging claims, classifying outlier claims and even fraud, and automating where applicable in which it was observed a declining trend of the human interference for claim processing and producing hassle-free for whole process (Gupta et al., 2018; Kakhki et al., 2020; Rawat et al., 2021). Kar (2016) reported that the usage of ML modelling in this process helped the industry to determine the beneficiary's claim applicating pattern along with the pattern of claim acceptance, which could be applied to augment the entire process flow for policy enrolment.

However, the prediction of satisfaction data accuracy on health insurance policy (HIP) through machine learning algorithms among respondents are lacking. In this regard, it was attempted to predict the satisfaction accuracy on health insurance policy among participants through machine learning algorithms especially tree models.

Methodology

We studied the data mining based on ML classifiers by using WEKA tool, version, 3.8.5 (Frank et al., 2016). The study was conducted on primary data and questionnaires-based survey among 385 respondents of Kolkata. The data were pre-processed and classified as per earlier protocol (Witten et al., 2011; Talapatra et al., 2021). The predictive accuracy of data of satisfaction on health insurance policy through ML algorithms especially 4 tree algorithms viz. decision tree (DT) J48, Random forest (RF), Random tree (RT) and Fast decision tree learner tree (REPT) along with 9 attributes viz. Facilities, Claim_coverage, Tax_benefit, Unexceptional_risk, Trust_insurer, Policy_benefit_bonus, Policy_benefit_premium_amount, Amount_claim_offered_maturity, Policy_benefit_family_production and class (poor, moderate and good response) from dataset were determined. As per protocol of Bouckaert et al. (2020), the modelling summary of predictive results such as were separately retrieved from WEKA tool and the statistical parameters are F-value, Matthew’s correlation coefficient (MCC), receiver operating characteristic (ROC) and Precision-recall curve (PRC), respectively were obtained as per 10-fold cross validation (CV) test.

Results

In the present pre-processing step, graphical presentation of statistical data of different 9 attributes Facilities, Claim_coverage, Tax_benefit, Unexceptional_risk, Trust_insurer, Policy_benefit_bonus, Policy_benefit_premium_amount, Amount_claim_offered_maturity, Policy_benefit_family_production and 3 types of classes viz. poor, moderate and good response for health insurance satisfaction among respondents separately were obtained (Fig 1).

For facilities, a higher 179 instances of range (1.0-1.571) followed by 163 instances of range (2.714-3.286), 24 instances of range (4.429-5.0), 17 instances of range (1.571-2.143) and lower 8 instances of range (3.857-4.429) were obtained. For Claim_coverage, a higher 168 instances of range (1.0-1.571) followed by 155 instances of range (2.714-3.286), 26 instances of range (4.429-5.0), 22 instances of range (1.571-2.143) and lower 8 instances of range (3.857-4.429) were recorded. For Tax_benefit, a higher 179 instances of range (2.714-3.286) followed by 157 instances of range (1.0-1.571), 26 instances of range (4.429-5.0), 16 instances of range (1.571-2.143) and lower 7 instances of range (3.857-4.429) were noted. For Unexceptional_risk, a higher 191 instances of range (2.714-3.286) followed by 137 instances of range (1.0-1.571), 24 instances of range (4.429-5.0), 23 instances of range (1.571-2.143) and lower 10 instances of range (3.857-4.429) were observed. For Trust_insurer, a higher 166 instances of range (1.0-1.667) followed by 157 instances of range (2.333-3.0), 34 instances of range (4.333-5.0), 17 instances of range (1.667-2.333) and lower 11 instances of range (3.667-4.333) were recorded. For Policy_benefit_bonus, a higher 194 instances of range (2.714-3.286) followed by 143 instances of range (1.0-1.571), 21 instances of range (4.429-5.0), 15 instances of range (1.571-2.143) and lower 12 instances of range (3.857-4.429) were noted. For Policy_benefit_premium_amount, a higher 175

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

instances of range (2.714-3.286) followed by 164 instances of range (1.0-1.571), 23 instances of range (4.429-5.0), 16 instances of range (1.571-2.143) and lower 7 instances of range (3.857-4.429) were observed. For Amount_claim_offered_maturity, a higher 181 instances of range (2.714-3.286) followed by 152 instances of range (1.0-1.571), 28 instances of range (4.429-5.0), 16 instances of range (1.571-2.143) and lower 8 instances of range (3.857-4.429) were recorded. For Policy_benefit_family_production, a higher 179 instances of range (1.0-1.571) followed by 153 instances of range (2.714-3.286), 28 instances of range (4.429-5.0), 17 instances of range (1.571-2.143) and lower 8 instances of range (3.857-4.429) were noted.

Regarding classes, maximum of 326 instances of poor followed by 47 instances of good and 18 instances of moderate were obtained.

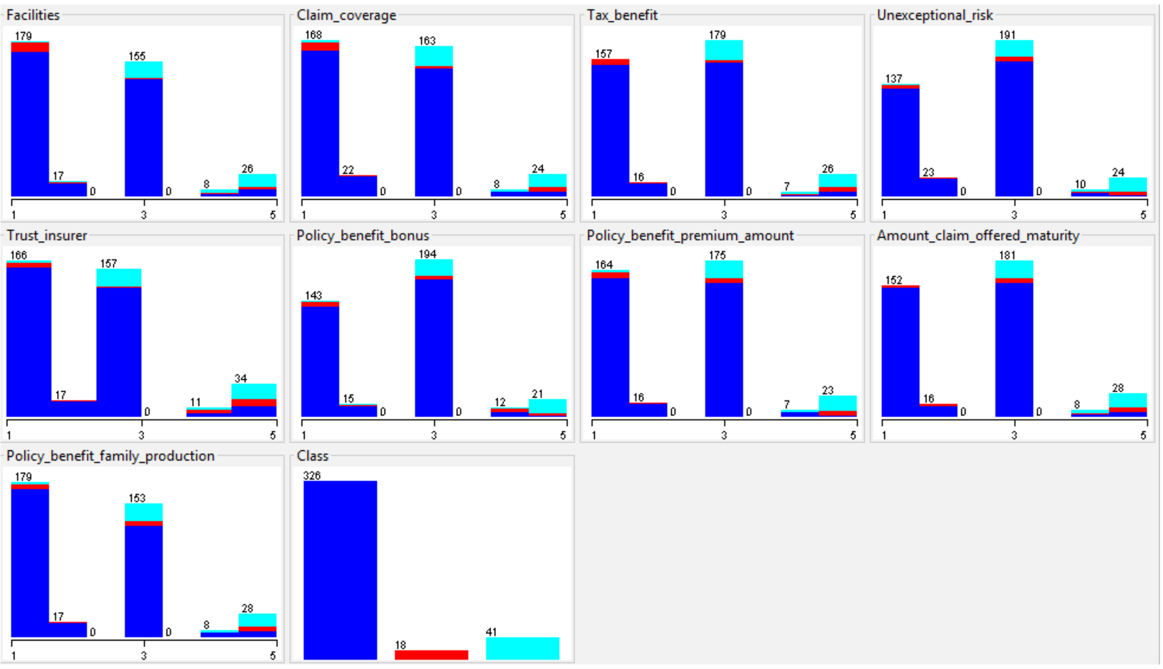

Figure 1: Graphical representation of different attributes after pre-processing

The prediction of model accuracy of studied ML algorithms as per correctly and incorrectly classified instances, KS, MAE and RMSE were studied as per 10-fold CV test. In the present study of algorithm model classification, the correctly classified instances were observed a higher value of about 88.052 for DT J48 followed by REPT (87.792) and RF (87.273) while lower value of about 85.714 for RT in the dataset (Table 1).

Table 1: Results on different classified instances and statistical values for different algorithm models on HIP

Classifier model	Correctly classified instances	Incorrectly classified instances	KS	MAE	RMSE
DT J48	88.052	11.95	0.46	0.12	0.27
RF	87.273	12.73	0.42	0.12	0.25
RT	85.714	14.28	0.40	0.12	0.30
REPT	87.792	12.21	0.40	0.13	0.27

DT J48 = Pruned and unpruned decision tree C4; RF = Random Forest; RT = Random tree; REPT = Fast decision tree learner; KS = Kappa Statistics; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error

Table 2 evaluates the detailed accuracy of studied tree models for the studied dataset. To evaluate the accuracy of a classifying values for F-measure, MCC, ROC and PRC, the better performances were studied for 3 classes. In our study, in the poor class a maximum value of PRC was obtained in the studied ML algorithms viz. RF (96%) and RT (93%) followed by REPT (91%) and DT J48 (90%).

Table 2: Statistical data for prediction accuracy of studied algorithms on HIP

Classifier model	**Effects**	**F-value**	**MCC**	**ROC area**	**PRC area**
DT J48	Poor	0.938	0.518	0.700	0.902
	Moderate	0.276	0.257	0.558	0.151
	Good	0.554	0.538	0.718	0.430
RF	Poor	0.934	0.477	0.824	0.957
	Moderate	0.133	0.102	0.845	0.196
	Good	0.571	0.568	0.860	0.593
RT	Poor	0.925	0.444	0.764	0.929
	Moderate	0.158	0.114	0.558	0.064
	Good	0.563	0.552	0.831	0.488
REPT	Poor	0.939	0.504	0.697	0.907
	Moderate	0.000	-0.011	0.556	0.081
	Good	0.507	0.478	0.738	0.422

DT J48 = Pruned and unpruned decision tree C4; RF = Random Forest; RT = Random tree; REPT = Fast decision tree learner; MCC = Matthew's correlation coefficient; ROC = Receiver operating characteristic; PRC = Precision-recall curve

Discussion

In this study, we used ML algorithms especially 4 tree models that identified the better performing classification models. All the studied models were run 10-fold CV method. Appiahene et al. (2020) observed the operational efficiency of Ghanaian bank's prediction was conducted by using three ML models such as "decision tree", "random forest", and "neural networks" and they found only "decision tree and its C5.0 algorithm" was the suitable predicted model, which is supported the present study. Our finding on the prediction accuracy was comparatively higher value as per the work by Hamid & Ahmed (2016) and Madaan et al. (2021) related to J48 (78.38%) and decision tree (73%) model accuracy. Bärtl & Krummaker (2020). evaluated 4 ML techniques such as Decision Trees, Random Forests, Neural Networks and Probabilistic Neural Networks on their ability to predict accurately for export credit insurance claims. They documented that random forest performed significantly better than decision tree, neural network and probabilistic neural network against all prediction jobs, and carried their validation performance most reliably forwarded to the test performance. Several studies on ML modelling based on chatbots, faster claim settlements, personalised health insurance policies, cost-effectiveness, fraud detection, faster underwriting related to HIP (Kaushik et al., 2022) but the prediction of accuracy to know satisfaction among participants related to HIP is a first-time endeavour.

Conclusion

It is concluded that the enriched information from studied dataset by using ML modelling especially tree classifiers are obtained effective performance accuracy of algorithms viz. RF and RT followed by REPT and DTJ48 as per CV test. From this study, it was predicted poor satisfaction on HIP among participants. It is suggested to validate the present predictive data.

Acknowledgment

Authors are thankful to the developers of WEKA tool used in this study.

Conflict of interest

None

References

[1] Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: Comparative study of decision tree, random forest, and neural networks. *Advances in Fuzzy Systems*, *2020*, 8581202.

[2] Bärtl, M., & Krummaker, S. (2020). Prediction of claims in export credit finance: A comparison of four machine learning techniques. *Risks*, *8*(1), 22.

[3] Doupe, P., Faghmous, J., & Basu, S. (2019). Machine learning for health services researchers. *Value in Health*, *22*(7), 808-815.

[4] Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management*, *42*, 78-89.

[5] Hamid, A. J., & Ahmed, T. M. (2016). Developing prediction model of loan risk in banks using data mining. *Machine Learning and Applications: An International Journal*, *3*(1), 1-9.

[6] Kakhki, F. D., Freeman, S. A., & Mosher, G. A. (2020). Applied machine learning in agro-manufacturing occupational incidents. *Procedia Manufacturing*, *48*, 24-30.

[7] Kar, A. K. (2016). Bio inspired computing - A review of algorithms and scope of applications. *Expert Systems with Applications*, *59*, 20-32.

[8] Kaushik, K., Bhardwaj, A., Dwivedi, A. D., & Singh, R. (2022). Machine Learning-Based Regression Framework to Predict Health Insurance Premiums. *International Journal of Environmental Research and Public Health*, *19*(13), 7898.

[9] Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. IOP Conf. Series: Materials Science and Engineering, 1022, 012042.

[10] Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing Journal*, *93*, 106384.

[11] Rawat, S., Rawat, A., Kumar, D., & Sai Sabitha, A. (2021). Application of machine learning and data visualization techniques for decision support in the insurance sector. International *Journal of Information Management Data Insights*, *1*(2), 100012.

[12] Sengupta, S., Basak, S., Saikia, P., Paul, P., Tsalavoutis, V., Atiah, F., Ravi, V., & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, 105596.