Enhancing Medical Insurance Pricing Prediction with SHAP-XGBoost for Informed Decision-Making

Danh Hong Le(✉)

Van Hien University, Ho Chi Minh City, Vietnam
danhlh@vhu.edu.vn

Abstract. This research explores the evolving field of predictive modeling in healthcare highlighting the continued interest of insurance companies in using Machine Learning (ML) techniques to improve operational efficiency. The author uses a set of regression based ML models incorporating different versions of Extreme Gradient Boosting (XGBoost) methods to predict medical insurance costs. Furthermore the study utilizes Explainable Artificial Intelligence (XAI) methods, Shapley Additive Explanations (SHAP) to identify and explain the key factors influencing medical insurance premium prices within the dataset. The dataset consists of 986 records from the KAGGLE repository and the models effectiveness is thoroughly assessed using various performance evaluation metrics such as R squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Additionally a comparison is made between the results generated by XGBoost and Random Forest (RF) models, in determining the features affecting Premium Prices. Despite requiring computational resources the XGBoost model stands out as the top performer overall. The authors aim to offer insights to help policymakers, insurance providers and individuals looking for medical coverage make informed decisions. This will assist them in choosing policies that best suit their needs and preferences.

Keywords: Medical insurance pricing · Predictive · SHAP · XGBoost

1 Introduction

In times there has been a growing emphasis on developing effective premium structures in the health insurance sector through actuarial modeling of insurance claims. This focus is crucial for attracting and retaining policyholders and managing existing plan members efficiently. However creating a model for medical insurance costs faces significant challenges due to the complex interplay of various influencing factors. Factors such as characteristics, health status, geographical accessibility, lifestyle preferences and provider attributes play a substantial role in determining medical insurance expenses [1]. Additionally key elements like coverage extent, plan type, deductible amounts and customer enrollment age greatly impact the costs associated with medical insurance coverage.

M. Pagac et al. (Eds.): SCFF 2024, LNNS 1062, pp. 317–328, 2024.
https://doi.org/10.1007/978-3-031-65656-9_32

Health insurance serves as a mechanism for pre-payment and risk sharing to cover expenses related to illnesses including hospitalizations, medications and consultations with healthcare providers. The introduction of national health insurance programs has improved equitable access, to healthcare services and helps individuals mitigate financial burdens from illnesses [2]. Currently most health insurance systems operate under multiple financing structures. In a single payer system there is a health insurance pool while the multiple payer model consists of several separate pools. Both funding methods have differences. The single payer system usually involves increased government supervision of healthcare delivery focusing on fairness concerns. Conversely, the multi-payer model affords consumers the freedom to select from available insurance providers, fostering innovation and competition within the industry [3].

In countries health insurance systems often blend elements from both payer and multi payer models creating various hybrid setups. These setups can generally be categorized into three types [4]: i). Health Maintenance Organization (HMO): This model lets insured individuals choose from a group of doctors affiliated with the HMO or contracted by it giving them the freedom to select their healthcare providers; ii). Preferred Provider Organization (PPO): PPO plans offer a list of contracted healthcare providers. Typically reimbursement follows a 60/40 split, where the insurer covers 60% of expenses and the insured pays the remaining 40%; iii) High Deductible Health Plan (HDHP) With a Health Savings Account (HSA): HDHPs involve policyholders setting up a health savings account from which treatment costs are deducted based on a percentage. This setup often leads to lower premiums. Additionally the text discusses two methods used by medical insurance companies to reimburse policyholders [5]: i). Cashless Treatment: In this system the insurance company directly handles payments, with hospitals removing the need for payment, by policyholders; ii). Reimbursement: Here policyholders typically pay for medical expenses upfront. Then request reimbursement from their insurance provider.

Machine learning (ML) models, often referred to as "black boxes," in the field of intelligence (AI) focusing on its ability to create systems that can learn independently from large datasets. These "black box" models pose a significant challenge, as their decisions are often difficult to comprehend, leading to concerns about trust, accountability, and ethical implications [6]. While AI includes technologies like expert systems, deep learning and robotics ML specifically concentrates on learning from data driven approaches as supported by research studies [7]. The widespread adoption of ML techniques across industries is credited to advancements in methodologies improved computing power of GPUs and the availability of diverse datasets according to experts [8,9]. Despite these progressions the text highlights that there is still more to explore in the potential of AI and ML leading to research in this area [10]. ML tools aid decision making by making predictions based on data and improving performance as they receive data, evident in medical applications as mentioned in a specific source [11]. In years health insurance companies have increasingly utilized AI and ML to better identify individuals who need protection and streamline their insurance processes. The text emphasizes that one key strength of ML in healthcare

management is its skill, in reasoning and quick trend analysis [12]. By utilizing intelligence to create accurate risk assessments and pinpoint clients requiring tailored attention insurance companies can allocate resources, towards policyholders instead of bureaucratic tasks. Systems that integrate data analysis enhance assessments and provide recommendations can significantly reduce the need, for human analysts and administrative costs. Moreover, the text highlights extensive research conducted on utilizing ML systems to forecast medical insurance costs, which employed various regression techniques and machine learning algorithms [13].

The authors in [14] introduced the SHapley Additive exPlanations (SHAP) method as a way to understand how different features contribute to an instance. This approach has been widely used to interpret social phenomena [15]. Using principles, from game theory SHAP assigns zero importance to features that don't impact the models prediction [16]. By analyzing SHAP values researchers can determine the relationship between each variable and the target variable, whether it's positive or negative. Moreover SHAP offers interpretability by assigning a value to all features [17]. Recent studies have shown an uptick in the use of eXplainable Artificial Intelligence (XAI) methods in fields, such as predicting medical insurance costs. In a study how Machine Learning models can predict costs for employer groups during health insurance renewals [18]. They focused on groups that could benefit from cost saving measures and developed models, at both patient and employer group levels. Based on a study involving data, from 14 million patients researchers found a 20% improvement in the effectiveness of insurance pricing models compared to existing ones. They successfully identified 84% of cost saving opportunities showcasing how machine learning systems can enhance the accuracy and fairness of health insurance pricing. Furthermore they used the SHAP XAI method with the LightGBM model to provide explanations for rate adjustments at the individual member level making the model more reliable [19]. The authors in [20] introduced a machine learning approach for predicting health insurance prices using regression models like linear regression and random forest regression. Researchers improved medical insurance pricing models by incorporating XAI techniques like Microsoft InterpretML, LIME, and SHAP, and demonstrated superior performance using machine learning methods such as Random Forest (RF), Gradient Boosting Machine (GBM), Artificial Neural Network (ANN), and Logistic Regression (LR) in predicting high-cost patients from healthcare claims data [21]. The study particularly emphasized the utilization of the SHAP XAI method exclusively for the RF model to assess variable influences on predicted class probabilities.

In a study, the author in [22] compared tree based classifiers to improve risk assessment for life insurance companies using predictive analytics. Their findings indicated that XGBoost outperformed other methods, underscoring the significance of model interpretability for stakeholders in the insurance sector. Using the SHAP method, they further analysed how dataset features influenced overall model performance. The comparison of various XAI methods underscored machine learning's potential in addressing healthcare challenges, particularly in

health insurance, by utilising Kaggle datasets to provide cost-effective solutions amidst growing demands and technological complexities. This research offers contributions:

- It forecasts medical insurance expenses using an ensemble learning approach with the XGBoost model.
- It evaluates the effectiveness of two XAI techniques, XGBoost and RF, in explaining how a black box model operates.

The papers structure is organized as follows; Sect. 2 covers medical insurance concepts and relevant literature reviews. Section 3 details the methodology used in the research. Section 4 presents outcomes and determinant analyses. Finally Sect. 5 summarizes the findings.

2 Methodology

2.1 Dataset and Preprocessing

The dataset utilized for medical insurance cost analysis was obtained from KAGGLE's repository in 2021 [23]. This dataset contains 986 entries, with 11 attributes. Following this data preprocessing steps were taken, which are essential in any data focused project seeking to uncover insights. During this phase checks were carried out to address missing values and ensure dataset uniqueness. Moreover the data was standardized using the Standard Scalar method.

Table 1. Statistical summary of the features

Features	Description	Mean	STD	Min.	25%	50%	75%	Max.
Age	Years old at the time of data collection	41.75	13.96	18.00	30.00	42.00	53.00	66.00
Diabetes	Whether the customer has been diagnosed with diabetes	0.42	0.49	0.00	0.00	0.00	1.00	1.00
BloodPressureProblems	Categorization of the customer's blood pressure based on pre-defined thresholds	0.47	0.50	0.00	0.00	0.00	1.00	1.00
AnyTransplants	Presence or absence of any major organ transplants in the customer's medical history	0.06	0.23	0.00	0.00	0.00	0.00	1.00
AnyChronicDiseases	Presence or absence of specific chronic illnesses (e.g., asthma, heart disease)	0.18	0.38	0.00	0.00	0.00	0.00	1.00
Height	Customer's height	168.18	10.10	145.00	161.00	168.00	176.00	188.0
Weight	Customer's weight	76.95	14.27	51.00	67.00	75.00	87.00	132.00
KnownAllergies	Presence or absence of any known allergies	0.22	0.41	0.00	0.00	0.00	0.00	1.00
HistoryOfCancerInFamily	Whether any blood relative of the customer has been diagnosed with cancer	0.12	0.32	0.00	0.00	0.00	0.00	1.00
NumberOfMajorSurgeries	Total number of major surgeries the customer has undergone	0.67	0.75	0.00	0.00	1.00	1.00	3.00
Premium Price	Cost of the insurance premium based on various risk factors, including the listed elements	24336.71	6248.18	15000.00	21000.00	23000.00	28000.00	40000.00

2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to explore the dataset aiming to uncover hidden patterns identify anomalies, assumptions and guide the selection of suitable ML techniques for solving the specific problem. The Statistical summary table provided insights into the distribution of features within the dataset. Additionally a Pearson correlation Heatmap (referenced as Fig. 1) was utilized to evaluate relationships among features and their correlations. The Heatmap indicated that aside from age there were no correlations, among variables. Such discoveries are quite usual; as, per the World Health Organization (WHO) getting older often leads to increased use of healthcare services and expenses.

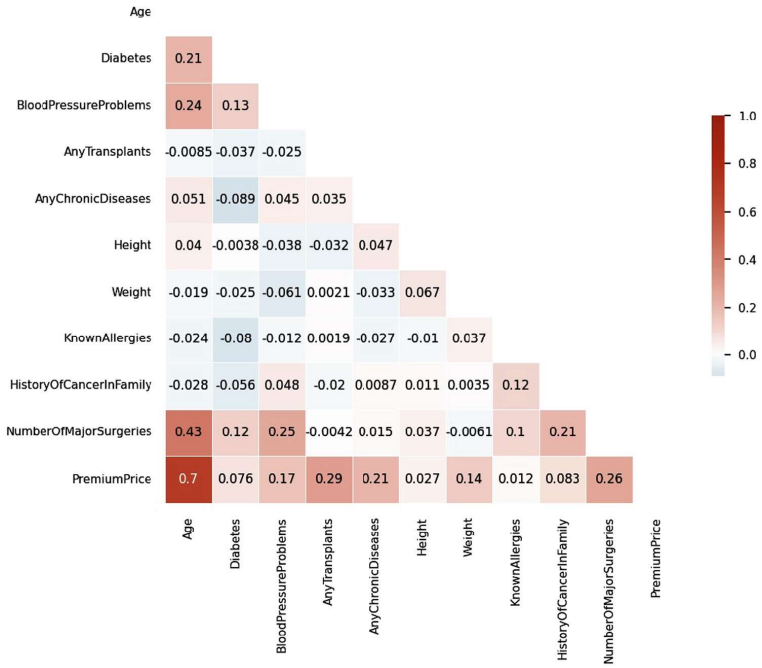

Fig. 1. Correlation heatmap.

2.3 eXtreme Gradient Boosting (XGBoost) Model

After the processing we selected all the features, for our study. Following that we split the dataset into two parts; one for training and the other for testing with 75% of the data used for training and 25% for testing purposes. During the training phase we created models to predict medical insurance costs while we used the testing dataset to evaluate how well the regression model performed. This approach demonstrates how ensemble models are effective in predictive healthcare modeling as discussed in [24]. In the section a brief overview of the XGBoost model that was implemented is presented.

The XGBoost model was designed to provide an scalable implementation of gradient boosting techniques first introduced in [25]. It is widely accepted as

a tool among gradient boosted trees algorithms due to its user nature as an open source platform and its effectiveness. Gradient boosting, a type of learning method aims to make predictions of a target variable by combining forecasts from simpler models that are less complex. As mentioned in reference [26], XGBoost generally delivers predictions that outperform those from RF model is comparable to results, from networks. One key feature of XGBoost is its use of threading strategies which optimize CPU core usage to improve overall performance and computational speed compared to traditional gradient boosting methods.

In regression analysis using gradient boosting, regression trees are utilized as fundamental learners, with each tree assigning input data points to leaf nodes holding continuous scores, while the model's objective function combines a convex loss function capturing prediction disparities with a penalty term addressing model complexity. Through iterative training, the algorithm integrates new trees to predict residuals or errors from prior trees. For a given dataset with n examples and m features, denoted $D = \{(x_i, y_i)\}\,(|D| = n, x_i \in R^m, y_i \in R)$, a tree ensemble model is constructed. Let $\hat{y}_i^{(t)}$ represent the prediction of the i-th instance at the t-th iteration, l is a differentiable convex loss function that quantifies the disparity between the prediction $\hat{y}_i$ and the target y_i, $\Omega()$ penalizes the complexity of the model, specifically targeting the regression tree functions. The objective function (comprising both the loss function and regularization) at iteration t that we aim to minimize is as follows:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i{}^{(t-1)} + f_t\left(\mathbf{x}_i\right)\right) + \Omega\left(f_t\right). \tag{1}$$

In our research we assessed the models using four metrics to evaluate their performance; R^2, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). R^2 also known as the coefficient of determination measures how well a model fits by looking at how much of the predicted price's explained by the features. The MAE provides a reflection of prediction errors while the RMSE indicates how well a regression model predicts the value of a response, by measuring the standard deviation of residuals. The RSME evaluate errors based on the value being predicted such as premium prices in our scenario. The MAPE calculates errors in percentage form showing the percentage difference between predictions and their intended targets in the dataset. It can be thought of as MAE presented as a percentage. As noted in reference [27], a MAPE below 10% suggests quality modeling. Moreover in a situation an R^2 score nearing 100% signals results and signifies a more precise and superior performing model. These metrics can be represented visually as shown below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2}, \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_i - y_i^p\right)\right|, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^p)^2}, \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{y_i - y_i^p}{y_i}\right)\right| * 100, \tag{5}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, n denotes the sample size, y_i^p represents the prediction for the i^{th} sample, and y_i denotes the actual value corresponding to the mean of the sample.

2.4 Shapley Additive Explanations (SHAP)

Shapley additive explanations (SHAP) serves the purpose of elucidating a model f at a specific individual point x^* through a value function denoted as e_S, $e_S = E\left[f(x) \mid x_S = x_S^*\right]$, where S represents a subset of $S \subseteq \{1, \ldots, p\}$. This approach, called SHAP allows for an examination of how each features impact varies across space and the linear connections, within the dataset on medical insurance costs. By assessing the importance of each factor through changes in variables SHAP offers insights, into feature significance making it easier to generate feature dependency graphs and conduct interaction analyses as mentioned in [28]

$$I_j = \sum_{i=1}^{n}\left|\phi_j^{(i)}\right| \tag{6}$$

where j is denoted by ϕ_j and calculated as the weighted average over all possible subsets S, $\phi_j^{(i)}$ represents the SHAP value of the j-th feature for instance i, $\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|s|!(p-|s|-1)!}{p!}\left(val\left(S \cup \{x_j\}\right) - val(S)\right)$. Here, p denotes the number of features, S refers to a subset of these features, x represents the feature values of a specific instance in the explained model, and $val(S)$ indicates the prediction for the feature values within set S.

3 Results and Discussions

3.1 The Overview of the Model Outcomes on the Test Dataset

The results, from the models shown in Table 2 demonstrate performance across the models that were evaluated. Specifically the XGBoost model utilized resources compared to the RF models. Notably the XGBoost model outperformed the RF model with a R^2 score of 87.290% and an RMSE of 2229.842 showcasing its superior predictive accuracy. While the RF model showed MAE and MAPE scores than the XGBoost models it's crucial to consider that RMSE penalizes deviations more harshly than MAE does. The high R^2 score achieved by the XGBoost model highlights its enhanced ability to explain variations in data compared to the RF models. With a Premium Price of 2421.753 based on Table 1 data all models displayed predictive capabilities.

Table 2. Overview performance outcomes on the test dataset

Model	MAE	RMSE	R^2 (%)	MAPE (%)	Elapsed time (s)	Memory used (MB)
XGBoost	1439.805	2229.842	87.290	4.975	5267.084	6.760
RF	1381.870	2421.753	83.863	5.951	54.978	0.892

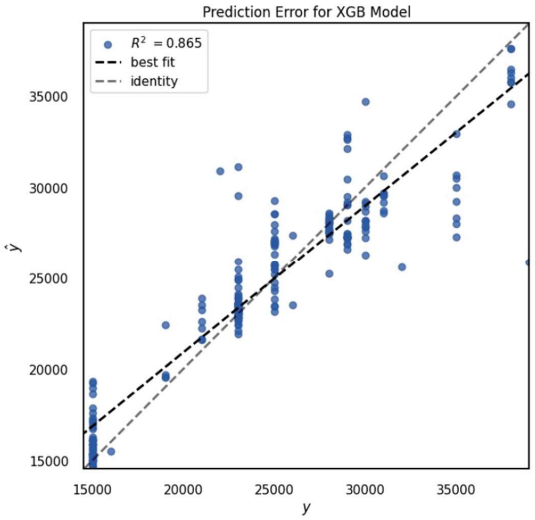

Fig. 2. Prediction error for XGBoost

Furthermore using Yellowbricks Prediction Error Visualizer we plotted targets from the dataset against predicted values. The plot shown in Figs. 2 and 3 illustrates how well each models predictions align with performance, along a 45° line.

3.2 The Feature Importance Analysis Conducted Using SHAP

The SHAP summary plot is a tool, for visualizing how various features impact the models predictions especially when it comes to predicting PremiumPrice. It combines SHAPley values. Feature importance to show the contributions of features, across the models output range. In Figs. 4 and 5, we start to see some insights, about these connections. The summary plot of the XGBoost model shows that when features like "Age", "BMI", "AnyTransplant", "AnyChronicDiseases", "HistoryOfCancerInFamily" and "BloodPressureProblems" have values the SHAP value also increases. This pattern suggests that higher values in these features are linked to premium prices. The SHAP analysis conducted on the RF model revealed that the impact of the "BloodPressureProblems" feature on premium prices was more pronounced compared to its influence in the XGBoost model. On the hand a high value for the "NumberOfMajorSurgeries" feature seems to have an effect on premium prices while a lower value has a positive impact. In addition the features "KnownAllergies" and "Diabetes" seem to have influence on premium prices based on the XGBoost models analysis.

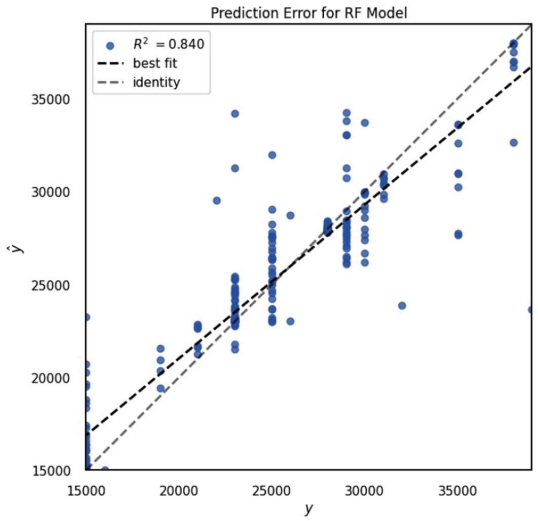

Fig. 3. Prediction error for RF

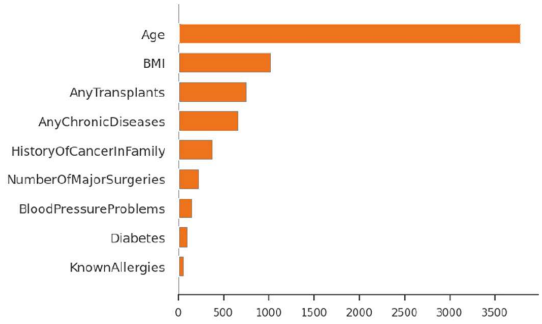

Fig. 4. SHAP summary plot (XGBoost)

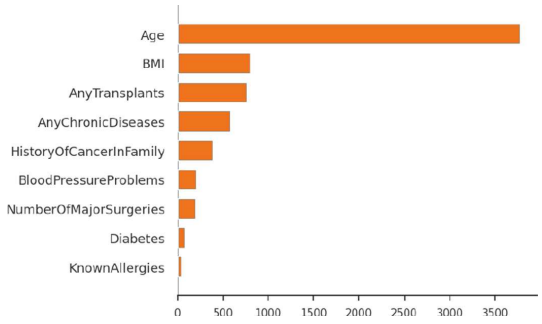

Fig. 5. SHAP summary plot (RF)

This study corresponds well with research by [19,21] which emphasize the effectiveness of XAI techniques in clarifying black box ML models used for predicting medical insurance expenses. Through XAI methods this research sheds light on the relationships between variables and their impact, on premium costs. Notably it shows that Age and BMI play roles in determining medical insurance expenses across all three models examined. The results align, with studies in Actuarial analysis as shown by research from [29,30]. This study emphasizes the importance of XAI techniques, in improving the understandability of machine learning models for predicting medical insurance costs through an investigation.

4 Conclusion

In this study the application of modeling, in the healthcare sector continues to be a focus of actuarial research, driven by insurance companies increasing interest in utilizing ML methods to enhance efficiency and productivity. By using the XGBoost model predictions were made on medical insurance costs based on a dataset from KAGGLEs database showing performance, across models. Notably the XGBoost model achieved a R^2 score of 86.470% and an RMSE of 2231.524 while the RF model excelled in terms of MAE and MAPE with values of 1379.960 and 5.831% respectively. Additionally the RF model demonstrated construction time and lower memory usage compared to XGBoost. To improve model interpretability the SHAP method of XAI was utilized to pinpoint factors influencing price predictions among the features analyzed. This research holds significance in offering decision support to overwhelmed prospective buyers in the medical insurance field by enabling insurers to streamline policy selection through a feature screening process that empowers buyers to find customized policies that suit their needs and financial circumstances.

References

1. Duncan, I., et al.: Testing alternative regression frameworks for predictive modeling of health care costs. N. Am. Actuar. J. **20**(1), 65–87 (2016). https://doi.org/10.1080/10920277.2015.1110491
2. Hartman, B., et al.: Predicting high-cost health insurance members through boosted trees and oversampling: an application using the HCCI database. N. Am. Actuar. J. **25**(1), 53–61 (2020). https://doi.org/10.1080/10920277.2020.1754242
3. Improving health insurance systems, coverage, and service quality. [Online]. https://ww1.issa.int/analysis/improving-health-insurance-systems-coverage-and-service-quality. Accessed 01/02/2023
4. Health Insurance: Definition, How It Works. Investopedia. [Online]. https://www.investopedia.com/terms/h/healthinsurance.asp. Accessed 01/02/2023
5. What Is Health Insurance: Meaning, Benefits & Types. Forbes Advisor INDIA. [Online]. https://www.forbes.com/advisor/in/health-insurance/what-is-health-insurance/. Accessed 01/02/2023
6. Carvalho, D.V., et al.: Machine learning interpretability: a survey on methods and metrics. Electronics **8**(8), 832 (2019). https://doi.org/10.3390/electronics8080832

7. Akter, S., et al.: Transforming business using digital innovations: the application of AI, blockchain, cloud and data analytics. Ann. Oper. Res. **308**(1–2), 7–39 (2020). https://doi.org/10.1007/s10479-020-03620-w
8. Nguyen, H.-S., et al.: Deep reinforcement learning autoencoder with RA-GAN and GAN. Int. J. Adv. Intell. Inform. **8**(3), 313 (2022). https://doi.org/10.26555/ijain.v8i3.896
9. Sánchez Fernández, I., Peters, J.M.: Machine learning and deep learning in medicine and neuroimaging. Ann. Child Neurol. Soc. **1**(2), 102–122 (2023). https://doi.org/10.1002/cns3.5
10. Nguyen, H.-S., et al.: Digital transformation for shipping container terminals using automated container code recognition. TELKOMNIKA (Telecommun. Comput. Electron. Control) **21**(3), 535 (2023). https://doi.org/10.12928/telkomnika.v21i3.24137
11. Ngiam, K.Y., Khor, I.W.: Big data and machine learning algorithms for health-care delivery. Lancet Oncol. **20**(5), e262–e273 (2019). https://doi.org/10.1016/s1470-2045(19)30149-4
12. Using AI and Machine Learning to Improve the Health Insurance Process. Forbes. [Online]. https://www.forbes.com/sites/forbesbusinesscouncil/2022/01/10/using-ai-and-machine-learning-to-improve-the-health-insurance-process/?sh=47ed47de42b1. Accessed 03/03/2023
13. ul Hassan, Ch.A., et al.: A computational intelligence approach for predicting medical insurance cost. Math. Probl. Eng. **2021**, 1–13 (2021). https://doi.org/10.1155/2021/1162553
14. Lundberg, S.M., et al.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp. 4768–4777. Curran Associates Inc., Red Hook, NY (2017)
15. Janizek, J.D., et al.: Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine, May 2018. https://doi.org/10.1101/331769
16. Stojić, A., et al.: Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. **653**, 140–147 (2019). https://doi.org/10.1016/j.scitotenv.2018.10.368
17. Stiglic, G., et al.: Interpretability of machine learning-based prediction models in healthcare. WIREs Data Min. Knowl. Discov. **10**(5) (2020). https://doi.org/10.1002/widm.1379
18. Kshirsagar, R.: Accurate and interpretable machine learning for transparent pricing of health insurance plans. Proc. AAAI Conf. Artif. Intell. **35**(17), 15127–15136 (2021)
19. Bora, A., et al.: Interpretation of machine learning models using XAI - a study on health insurance dataset. In: 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Oct 2022. https://doi.org/10.1109/icrito56286.2022.9964649
20. Gaurav, D., Tiwari, S.: Interpretability vs explainability: the black box of machine learning. In: 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), Feb 2023. https://doi.org/10.1109/iccosite57641.2023.10127717
21. Langenberger, B., et al.: The application of machine learning to predict high-cost patients: a performance-comparison of different models using healthcare claims data. PLoS ONE **18**(1), e0279540 (2023). https://doi.org/10.1371/journal.pone.0279540

22. Sahai, R., et al.: Insurance risk prediction using machine learning. In: Lecture Notes on Data Engineering and Communications Technologies, pp. 419–433 (2023). https://doi.org/10.1007/978-981-99-0741-0_30
23. Medical Insurance Premium Prediction. [Online]. https://www.kaggle.com/datasets/tejashvi14/medical-insurance-premium-prediction. Accessed 01/03/2023
24. Hartman, B., et al.: Predicting high-cost health insurance members through boosted trees and oversampling: an application using the HCCI database. N. Am. Actuar. J. **25**(1), 53–61 (2020). https://doi.org/10.1080/10920277.2020.1754242
25. Friedman, J., et al.: Additive logistic regression: a statistical view of boosting. Ann. Statis. **28**(2) (2000). https://doi.org/10.1214/aos/1016218223
26. Sheridan, R.P., et al.: Extreme gradient boosting as a method for quantitative structure-activity relationships. J. Chem. Inf. Model. **56**(12), 2353–2360 (2016). https://doi.org/10.1021/acs.jcim.6b00591
27. What is a good MAPE score and how do I calculate it? [Online]. https://stephenallwright.com/good-mape-score/. Accessed 01/02/2023
28. Teoh, E.Z., et al.: Explainable housing price prediction with determinant analysis. Int. J. Hous. Mark. Anal. **16**(5), 1021–1045 (2022). https://doi.org/10.1108/ijhma-02-2022-0025
29. Mendelson, D.N., et al.: The effects of aging and population growth on health care costs. Health Aff. **12**(1), 119–125 (1993). https://doi.org/10.1377/hlthaff.12.1.119
30. Kamble, P.S., et al.: Association of obesity with healthcare resource utilization and costs in a commercial population. Curr. Med. Res. Opin. **34**(7), 1335–1343 (2018). https://doi.org/10.1080/03007995.2018.1464435