International Journal of Industrial Engineering & Production Research March 2022 Vol. 33, No. 1: 1-9
DOI: 10.22068/ijiepr.33.1.13

RESEARCH PAPER

Determining the Tariffs of Physicians Using A Combined Model of Data Mining Techniques and Fuzzy Logic In Health Insurance

Amir Akbarzadeh Janatabad[1], Ahmad Sadeghieh[2*], Mohammad Mehdi Lotfi[3] & Ali Mostafaeipour[4]

Received 24 January 2022; Revised 2 February 2022; Accepted 19 February 2022;

ABSTRACT

The health insurance system can play an effective role to control health expenditures. The purpose of this study is to provide a model for estimating the physician visit tariffs. To achieve this goal, a hybrid model was used. fuzzy logic is the most appropriate tool for controlling systems and deriving rules for the relationship between inputs and outputs. So, the output of the data mining techniques enter the fuzzy logic as an input variable. The data were collected from the Health Insurance Organization of Iran in two sections including the physicians' costs and physicians' deductions. Owing to the techniques used in this model, NN had the lowest rate of errors, as compared to the other data mining techniques (0.0034 and 0.0013, respectively). After the variables, the membership functions and fuzzy logic rules were defined, the accuracy of the whole control model was confirmed by random data. This research has dealt with the domains of health insurance , their connections and defining effective variables better and more extensively than the other studies in the field.

KEYWORDS: *Data mining technique; Fuzzy logic; Health insurance; Tariffs for physicians; Neural network.*

1. Introduction

Nowadays, governments often look for ways to control health expenditures and maximize the output. Rising costs are rooted in factors that are not necessarily in the control of health systems, such as demographic transitions, epidemiological transitions, changing health behaviors, and the medicalization of the society. On the other hand, certain factors such as access to services, their diversity and prices, introduction of new and expensive technologies and mechanisms of reimbursement to providers are to some extent in the control of the health sector. The proper management of these factors can help to adopt the best policies to control health costs [1]. Due to the limited resources and increased costs in the health system, an appropriate solution should be found to manage the costs.

Controlling the cost of treatment is one of the main goals of a health insurance system. The tariffs paid to the medical centers that have contracts with health insurance companies follow a single route; there is no model for calculating the tariff based on variables. As it seems, it is physicians who cause a lot of unnecessary costs in such areas as pharmacies, laboratories and radiology.

Based on what has been discussed, this research seeks to take an effective step to partially fill the gap in this field, which is due to the lack of a suitable model for the payment of tariffs to physicians. If the contracting institutions will be required to comply with the signed-up terms and avoid prescribing ineffective or impermissible items. They refuse to violate the rules of the organization.

The research questions can be expressed as follows:

1. How to determine the costs that a physician imposes on the insurance organization so as to calculate the tariff paid to doctors?

* Corresponding author: *Ahmad Sadeghieh*
sadegheih@yazd.ac.ir

1. *Department of Industrial Engineering, Yazd Universit, Yazd, Iran.*
2. *Department of Industrial Engineering, Yazd Universit, Yazd, Iran.*
3. *Department of Industrial Engineering, Yazd Universit, Yazd, Iran.*
4. *Department of Industrial Engineering, Yazd Universit, Yazd, Iran.*

2. What is the effect of a doctor's deductions on the payment tariff?
3. Given the heterogeneous data in the insurance industry, how should the required model be designed?

Obviously, the answers to these research questions have to do with controlling and managing costs in health insurance.

2. Literature Review

Currently, insurance agencies use different pricing strategies depending on their position, and can, accordingly, create a competitive advantage for themselves. These strategies are as follows:

1. Cost-based pricing: It includes pricing based on production, distribution, sales costs and addition of a fair return rate for the corresponding risks and efforts.
2. Segmented pricing: According to this strategy, offering different prices for different customers is based on a number of criteria such as the time of purchase, place of purchase, consumption of services and personal characteristics of customers.
3. Special event pricing: Sellers use this strategy to attract more customers to their stores on certain days or in right seasons. Manufacturers sometimes offer discounts for consumers who purchase the product directly from the dealer within a certain period of time. They send the discount directly to the customer in cash.
4. Discount and reward pricing: Many companies adjust prices to appreciate some of their customers' reactions, such as paying bills on time, bulk purchases, and end-of-season purchases. These price adjustments, known as discounts and sales, apply in the form of cash discounts, wholesale discounts, trade discounts, and bonuses.
5. Product set pricing: In this pricing strategy, sellers combine several products and sell at a cheaper price.
6. Optional product pricing: This means trying to get customers who are willing to spend a little more on the product by purchasing additional options or features with the product [2].

The present study focuses more on option 4.

There are different ways of pricing insurance services in the insurance industry. For example, [3] referred to an early model of life insurance. In that model, LR-fuzzy random variables were used to estimate the discount functions for general interest rates and future fuzzy random variable. Some annual models were thus created. Most studies in the insurance industry and the health sector have been done through data mining. This is because, in the field of health insurance, risks such as seasonal migration increase in certain populations and there is a rise in seasonal diseases, epidemics, number of departments and quantity of equipment in institutions, which cannot be ignored. In general, the insurance industry has always been associated with risks. Health insurance, for instance, is beset with the heterogeneity of data, which puts the future of insurance policies at risk. Decisions such as selling insurance, strategic service purchases, product design, variety and, most importantly, losses and costs are always associated with risks.

[4] made an innovation in the big data platform architecture by proposing the following platform for the insurance industry.

Fig. 1. Big data in the insurance industry

Based on this platform, the data in the insurance industry can be defined in six areas including the prices of services paid to institutions, the sale of insurance policies to the insured, the services that can be provided, costs and claims, prevention of fraud, and the design of diverse services to provide.Most studies have been conducted each on one of these six areas, and their method has been the comparison of data mining techniques, the combination of data mining techniques and the combination of data mining techniques with other algorithms such as fuzzy , Pso and neuro-fuzzy techniques.The research works done with data mining techniques in the insurance industry are reviewed below. Goodarzi and Janat [5] used common techniques to detect fraud in car insurance, including logistic regression, decision tree and simple Bayesian classification. It is to be noted that data mining tools typically serve to generalize models for detecting fraudulent claims and making predictions. The findings of the study indicated that the logistic regression model was more accurate in predicting all the claims

(fraudulent and non-fraudulent) than the tree decision and Bayesian methods. [6] pointed to a major problem in insurance companies.It regarded the plan to sell customer insurance coverage, and the customers were considered loyal. The study dealt with data cleansing in the first place. In the second step, several filtering and packaging methods were implemented to select the appropriate features. In the third step, the k-nearest neighbor algorithm was used to cluster the customers. A study was carried out by [7] with the aim of identifying data mining algorithms and using them to predict default risks, prevent potential payment problems and reduce potential problems so as to increase the insurance credit. The data of the study were obtained from a survey by the Turkish Statistical Institute in 2015 and contained information on the demographic and economic characteristics of individuals. Six classification algorithms including Naive Bayes, Bayesian networks, random tree, multilayer perceptron and logistic regression were implemented on the data set through the WEKA 3.9 data mining software. [8] examined customer services and automotive insurance claims through ML (machine learning) models for big-size insurance data. For this purpose, they used different ML models. Methods such as logistic regression, XGBoost, random forest, tree decision, simple Bayesian and K-NN were tested and compared for their performance. The results showed that random forest could serve the purpose better. For the first time, with a combination of data mining techniques and the support vector machine (SVM) technique, [9] predicted the customer turnover in an insurance organization. They first used the genetic algorithm to select the influential traits. After modeling the problem, they optimized the parameters of the backup vector machine model by network search and K-layer cross-validation. The performance of SVM method was predicted with the decision tree method, neural networks, logistic regression, random tree, Bayesian classification and k-nearest neighbor comparison. The parameters of each method were also optimized through network search. In that study, the backup vector machine method proved to have a better performance than the other methods. [10] predicted the stock prices of six pharmaceutical companies in the stock market of Bangladesh. In the first stage, K-means clustering was performed to identify the stocks with the most useful patterns. In the second stage, nonlinear self-regression was practiced with the neural network method to predict the stock price. The research showed that, with a combination of data mining and neural networking, decisions can be made for stock investment in the DSE pharmaceutical sector. According to [11], the production of a three-cluster algorithm can improve the decisions to manage risks in life insurance, although those researchers do not address data mining issues. In a conceptual paper, [12] discussed the role of artificial intelligence to detect fraud in the health insurance industry. He also pointed to the benefits and adverse consequences of different types of learning (deep, with supervision, without supervision) in the health insurance industry. Some other researchers have combined areas of insurance industry and data mining techniques with such solutions as fuzzy and artificial bee colonies algorithms. For example, [13] emphasized the issue of car insurance fraud, which is gradually expanding worldwide. They proposed the Kernel Ridge Regression (KRR) based on the artificial bee algorithm (ABC), coming up with the KRR-ABC algorithm to detect car insurance fraud. The performance of the KRR-ABC model was benchmarked in eight datasets, and it was compared with other methods. The test results showed that the model had a faster execution time and better performance. Another composite study was carried out by [14]. They used the adaptive genetic algorithm in a fuzzy logic model (AGAFL) to predict heart diseases, which would help physicians diagnose the diseases at an early stage. The tests were performed on heart disease data sets. A complete experimental analysis showed the better performance of this method than the existing ones. The proposed genetic algorithm was put in practice in the rules section of the fuzzy model. Similarly, [15] proposed a new model to predict type 2 diabetes. They used composite particle swarm optimization (PSO) and fuzzy clustering (FCM) tools to evaluate a set of medical data in the diagnosis of diabetes. Compared to other models, the proposed PSO-FCM method proved to have a better performance. [16] combined data conversion with cumulative training cases to improve predictions in an educational system. They concluded that the proposed hybrid fuzzy classification scheme could function with higher predictive accuracy. [17] referred to fuzzy expert systems. They stated that fuzzy logic is similar to the human reasoning system and can, therefore, control the uncertainty in medical diagnosis data. They examined the datasets on various data mining and fuzzy logic techniques and measured the accuracy of each of

those techniques in the diagnosis of diabetes. Almost all the studies conducted so far are on one of the three issues, namely insurance sales, customers or the insured, and claims of damage. The present study, however, deals with the relationships among damages, fraud; in fact, the two areas are taken into account. In addition, at each stage of analysis, the variables affecting the outputs are examined. It should be noted that the impacts of such variables have not been measured in other studies. Due to the heterogeneity of the data, a hybrid model is developed with a combination of non-linear data mining and fuzzy logic techniques. It is also to be mentioned that, regarding fraud and deductions in health insurance, there is a wide range of influential factors to address such as duplication of prescriptions, prescription distortion, prescription without visits, and non-commitment. Of course, there are many other factors that do not lie within the scope of this study; the paper is just dedicated to the effect of deductions on doctors' tariffs.

3. Methodology

3.1. Data mining

In short, to obtain the hidden information in a set of data, they are analyzed by certain data mining methods the most important of which are suggesting and recommending, Identifying incorrect data, Analysis of performance, risk management, customer segmentation, Prediction and targeted advertisement. Data mining is generally based on four tools and methods including classification, clustering, regression, and participation rules. Some of the data mining algorithms that typically use observer learning are Decision Tree, k-Nearest Neighbor, Naive Bayes, Neural Networks, Support Vector, Genetic Algorithm and Rugged Set Approach.

3.2. Research data

The reports related to the research were obtained from several software programs available at the Health Insurance Organization of Iran. The problem data were collected through separating, filtering and using the Excel and Access software programs. In order to obtain a payment tariff for physicians, separate data were collected about the average cost of each prescription, depending on the type of specialty and the amount of deductions for each physician during the year 2019. Deductions are the amounts deducted from the payments to doctors due to their non-compliance with the rules of the organization. In order to obtain the average cost of each prescription, the costs which individual physicians had imposed on the Health Insurance Organization of Iran in different provinces for pharmaceuticals, testing, imaging and physiotherapy. The input and output variables in this respect are shown in Tables 1 and 2. Since the data about the costs imposed by physicians and the deductions on their payments were separately collected from different provinces across the country, the corresponding data were fed into the model separately. Overall, 1433 and 763 data were collected and reported in the tables.

Tab. 1. Variables of the physicians (prescriptions)

Input variable		Output variable
Type of specialty	Rate of prescription without order drug and so on	Average cost per prescription

Tab. 2. Variables of the physicians (deductions)

Input variable	Input variable	Output variable
Type of specialty	Contract history	Amount of deductions

It should be noted that the general practice and specialty of the physicians were defined based on certain codes. The physician cost data were coded up to 17, and the deduction data were marked with codes up to 21. For example, orthopedic specialization code is 1 or internal medicine specialization code is 2 and so on. After the data collection, the preprocessing of the data was done before they were fed into the model and the data mining techniques were used.

3.3. Fuzzy logic in this study

The membership functions of the input and output variables are defined below.

A: Average medium costs imposed on the organization by the physician (based on the type of specialty and average prescription): As the data suggest, its function is triangular.
B: The average amount of deductions from the doctor (based on the type of specialty): Based on the data, its function is triangular.
C: Approved tariff rate: Since this study aims to make predictions for the next periods, the tariff approved in 2019 can be considered as a base. There are separate fees for general practitioners and specialists.
It is to be noted that, in items A to C, the membership functions are defined 1.

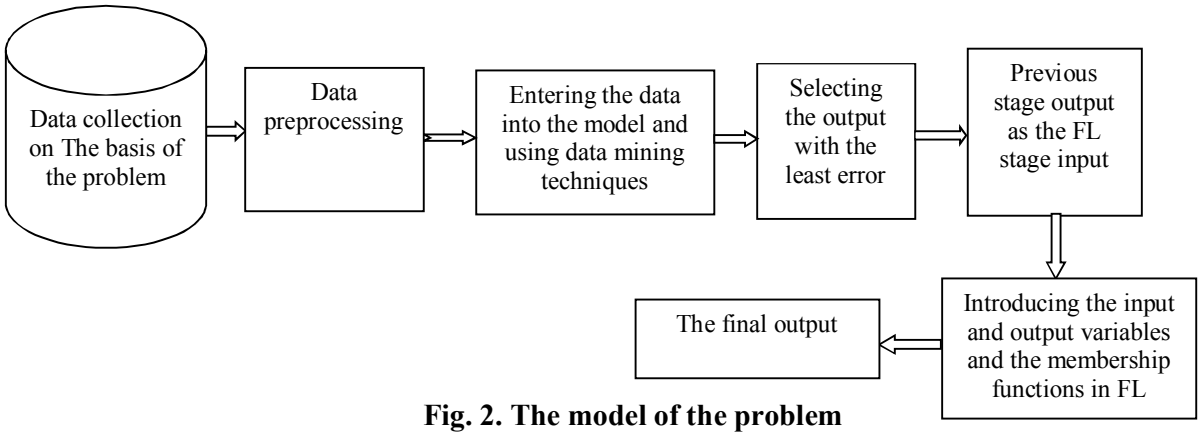

Fig. 2. The model of the problem

According to the above model, to solve the research problem, the problem data and the new influential variables were defined. Due to the heterogeneity of the data, non-linear data mining techniques were used more, and fuzzy logic was applied instead of other classification methods or mathematical relations. This was found as the most appropriate method to determine the doctors' tariffs for all the entries. Also, as the membership functions were defined, proper responses (appropriate output) were provided to all the inputs. The results suggest that the method used in this research has a special advantage over the ones used in health insurance. In general, new variables and a new method are selected and introduced here to reduce health insurance costs.

4. Results and Discussion

The examination of the research method above forms a basis for the further discussion of the proposed model and the research questions. As a matter of fact, the tariffs of the physicians were determined on the purpose of managing the costs of health insurance. To achieve this purpose, certain nonlinear data mining techniques including complex tree, simple tree, quadratic svm, cubic svm, fine Gaussian svm, exponential Gaussian process regression and NN were implemented on the problem data (Table 3).

Tab. 3. Evaluation of the data mining techniques

Technique	NN	Com.tree	Sim.tree	f.gu.svm	q.svm	c.svm	Gpr.exp
MSE value for type 1 data	0.0034	0.0049	0.0049	0.0049	0.0049	0.0049	0.0036
MSE value for type 2 data	0.0013	0.0036	0.0036	0.0036	0.0036	0.0036	0.0036

As it can be seen in Table 3, the NN technique is selected for the data 1 and 2. It is also possible to validate the model from Table 3. Seven techniques were examined and the MSE related techniques are close together. Now, more accurate predictions of the problem outputs can be made. The output variables of the first step are used as the input variables of fuzzy logic.After the variables and the membership functions are defined in the software, it is time to define the

rules. In the fuzzy model, however, the relationship between the membership functions is negative, meaning that the higher the cost imposed by the physician on the organization, the higher the physician's deductions, and the lower the tariff paid to the physician (Figure 2).

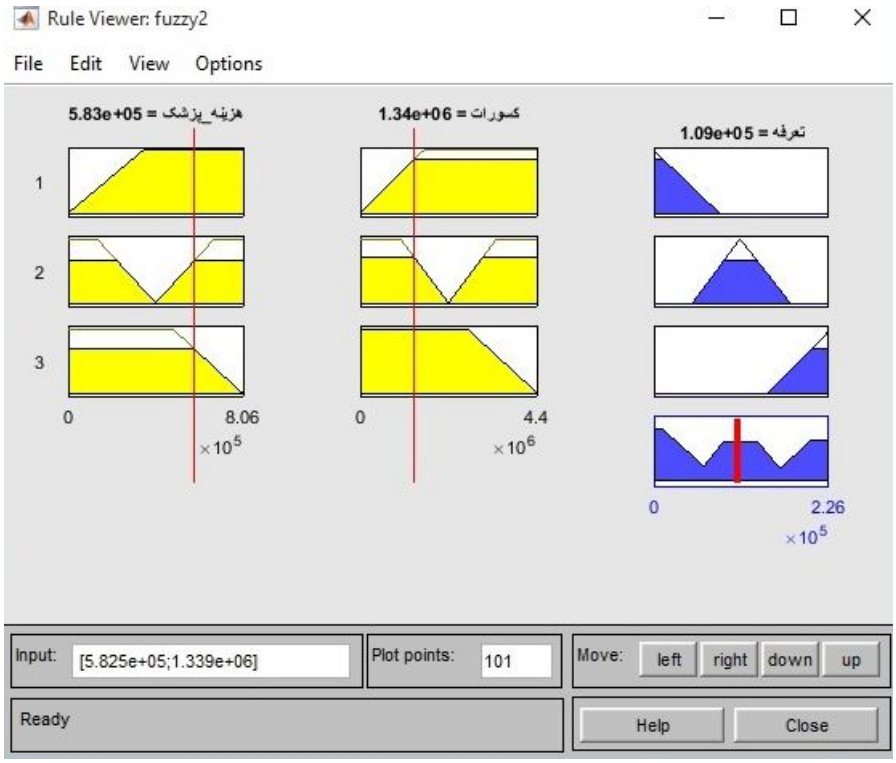

Fig. 3. Fuzzy model output for physician

In this research, the overall goal of the plan is to create and promote a culture of prevention among the organization's customers, namely the doctors, in order to reduce the organization's overall costs in the long run. Table 4 presents some numerical examples for obtaining a payment tariff for physicians. To achieve this goal, the membership functions of the fuzzy model are defined for type 1 specialists, such as orthopedics, and the fuzzy model is controlled.

Tab. 4. The final model format to determine physicians' tariffs

Row	Type of specialization	Rate of prescription without order drug and so on	First-stage output (average cost)	Type of specialization	Contract duration (years))	First-stage output (amount of deductions)	Combined model output (amount of tariff paid)
1	1	0.2	276514	1	3	2281248	1.12e+05
2	1	0.3	274086.4	1	5	3487968	1.05e+05
3	1	0.4	271460.7	1	9	23298798	7.38e+04
4	1	0.5	268627.4	1	2	1347868	1.13e+05
5	1	0.6	265578.7	1	4	3123289	1.11e+05

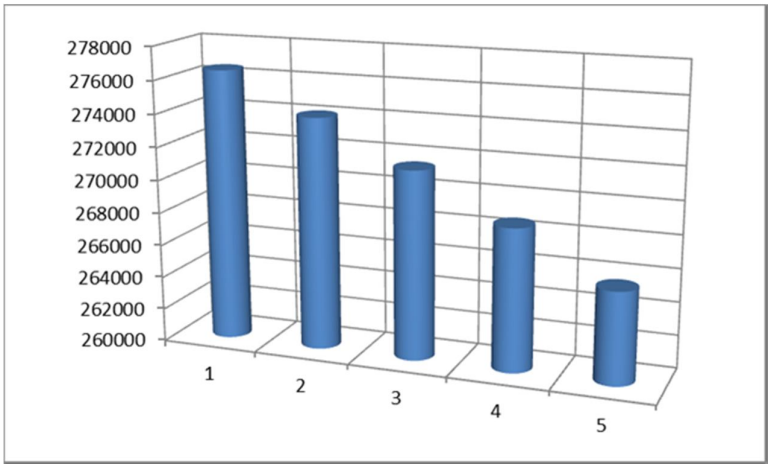

Fig. 4. First-stage output for physicians' tariffs (prescriptions)

As it has turned out, the designed model works well and the research questions are answered. That is, after the cost imposed by type 1 physicians and their deductions are estimated with artificial neural networks, it is possible to feed the data into the fuzzy logic model and to calculate the tariffs based on the defined variables. According to Table 4, with an increase in the deductions and average cost, the tariffs paid to doctors are decrease. According to Figure 4, Usually, the higher the rate of prescription without order, the lower the average cost. Although the duration of the doctors 'contract has no effect on the amount of deductions and factors such as the type of violation are more important, but the doctors' tariff is obtained in the above model. Physicians with longer contracts were expected to make fewer deductions. In general, compared to the other studies in the field of insurance industry especially health insurance, the present research clearly demonstrates the role of the costs imposed by physicians and their deductions in the payment of their tariffs.

5. Conclusion

The reduction of medical costs is a major goal in the public health system of every country. In this regard, health insurance has proved to be a good tool for this purpose. Due to heterogeneous data, risk management is an integral part of the insurance industry which, along with prediction, is considered to be one of the best applications of data mining. By combining data mining techniques with a fuzzy logic model, the present study has proposed a model for cost management in the Health Insurance Organization of Iran. The study can claim a degree of novelty, considering the fact that most studies on social health insuranceshave ignored the impacts of such variables as tariffs. There is also little attention paid to the costs imposed on the insurance organization by physicians and the amount of their deductions as influential factors. Any measure taken in this respect would promote the culture of prevention. Unlike the studies which focus on one area, this research addresses two issues including buying services and fraud as a factor which induces deductions. To deal with these issues, data mining techniques were first applied for prediction, and a technique with the least MSE error was selected. Neural networks with 0.0034 and 0.0013 were selected for data types 1 and 2 respectively. The output of this stage, however, did not serve the purpose of pricing the purchase of services. Therefore, a fuzzy logic model was used as the most appropriate tool to bring all the outputs of the first stage to the problem outputs. With the fuzzy logic method, for example, if orthopedic medicine is a specialty and rate of prescription without order drug and so on is 0.2, the neural network predicts 276514 as the average per-priscription cost that the physician imposes on the insurance company. If the doctor has specialized in orthopedics and worked with the Health Insurance Organization for three years, the neural network predicts 2281248 as the amount of his or her deductions. Then, 1.12 e + 05 is defined as the tariff for the physician. Apart from the variables examined in this paper, there are variables of possible effects to explore in future studies. A few recommended variables are physician's age, availability of services ranging from specialists to equipment.

References

[1] Alizadeh M, Gol khandan A, Article,Robust determinants of health sector costs in Iran: Bayesian model

averaging approach. Journal of healthcare management, Vol. 7, No. 2, (2017), pp. 47-61.

[2] Sabzi M, Rashidi A, Article, Pricing strategies in the insurance industry; With an approach to fair pricing challenges, International Conference on Insurance and Development, Vol. 23, (2015).

[3] Wang D, Article, A net premium model for life insurance under a sort of generalized uncertain interest rates, International Conference Series on Soft Methods in Probability and Statistics,SMPS 2018: Uncertainty Modelling in Data Science (2019), pp. 224-232.

[4] Liu Y, Peng J, Yu Z, Article, Big data platform architecture under the background of financial technology in the insurance industry as an example, 2018, BDET 2018: Proceedings of the 2018 International Conference on Big Data Engineering and Technology, (2018), pp. 31-35.
https://doi.org/10.1145/3297730.3297743

[5] Goodarzi A, Janat Babaei S, Article, Evaluation of Decision Tree Algorithms, Naive Bayes and Logistic Regression in Detection of Car Insurance Frauds, Insurance Research Quarterly, Vol. 1, No. 2, (2016), pp. 61-80.
https://doi.org/10.22054/IRISK.2017.9349

[6] Abdi F, Khalili K , Damghani SH, Article,Solving customer insurance coverage sales plan problem using a multi-stage data mining approach, emerald insight, Kybernetes, ISSN: 0368-492X, Vol. 47, No. 1, (2017).

[7] Cigsar B, Unal D, Article,Comparison of data mining classification algorithms determining the default risk, Scientific Programming, Vol. 2019, (2019).
https://doi.org/10.1155/2019/8706505

[8] Hanafy M, Ming R, Article, Machine Learning Approaches for Auto Insurance Big Data, Risks, Vol. 9, No. 2, 2021, p. 42.
https://doi.org/10.3390/risks9020042

[9] Rezaei Navaei S, Koosha H, Article, Applying Data Mining Techniques for Customer Churn Prediction in Insurance Industry, International Journal of Industrial Engineering & Production Management, Vol. 27, No. 4, (2016), pp. 635-653.
http://ijiepm.iust.ac.ir/article-1-1308-fa.html.

[10] Debashish D , Safa S. A ,Noraziah A, Article ,An Efficient Time Series Analysis for Pharmaceutical Sector Stock Prediction by Applying Hybridization of Data Mining and Neural Network Technique, Indian Journal of Science and Technology, Vol 9, No. 21, (2016).
DOI: 10.17485/ijst/2016/v9i21/95152, June 2016

[11] Yin Sh, Gan G, Valdez E. A, Vadiveloo J, Article,Applications of Clustering with Mixed Type Data inLife Insurance, ,arXiv:2101.10896v1 [stat.AP], Vol. 9, No. 3, (2021), p. 47.
https://doi.org/10.3390/risks9030047(registering DOI)

[12] Sharma K, Article, Artificial Intelligence for claim settlement and to analyze fraudulent practices in health insurance industries, Thapar University, India, (2021).

[13] Yan CH , Li Y , Liu W, Li M , Chen J , Wang L, Article, An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification, Neurocomputing, Vol. 393, (2020), pp. 115-125.

[14] Reddy G. T , Reddy M. P. K , LakshmaNNa K , Rajput D. S , Kaluri R , Srivastava G, Article, Hybrid genetic algorithm and a fuzzy logic classifier for heart disease Diagnosis, Evolutionary Intelligence, Vol. 13, (2020), pp. 185-196.

[15] Raja J. B, Pandian S. Ch, Article, PSO-FCM Based Data Mining Model to Predict Diabetic Disease,Computer Methods and Programs in Biomedicine, Vol. 196, (2020).
https://doi.org/10.1016/j.cmpb.2020.105659

[16] Zhao Q, Lai Sh.Ch, Wang J.L, Wang L.Y, Article,Hybrid Fuzzy Rule-Based Classification System for MOODLE LMS System,2021,Journal of Internet Technology, Vol. 22, No. 1 , (2021), pp. 81-90.

[17] Thakkar H, Shah V, Yagnik H, Shah M, Article,Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis, ClinicaleHealth, Vol. 4, (2021), pp. 12-23

Follow This Article at The Following Site:

Akbarzadeh janatabad A, Sadegheih A, lotfi M M, Mostafaeipour A. Determining the tariffs of physicians using a combined model of data mining techniques and fuzzy logic in health insurance. IJIEPR. 2022; 33 (1) :1-9
URL: http://ijiepr.iust.ac.ir/article-1-1427-en.html