The Implementation of Classification Algorithm C4.5 in Determining the Illness Risk Level for Health Insurance Company in Indonesia

Apriyudha Angkasa P. Universitas Mercu Buana Devi Fitrianah Universitas Mercu Buana

ABSTRACT

Fundamental thing on health insurance is how to manage all contributions fee from membership insurance, so it can use for finance health services. In this writer's case, the problem of health insurance is when registered membership insurance, there's no validation or adjustment about fee insurance with a history of illness from the applicant. That thing will be increasing financial cost if insurance does not use another approach from health services like promotive and preventive services for manage illness registered membership for health insurance, so that can be suppress financing of health services. Based on data on health insurance, they can do classification processing data and combined with algorithm C 4.5 for proses classification. Classification that has been used for mapping the level of risk illness membership in health insurance. Result from this research using a ten-fold cross-validation / confusion matrix with accuracy 99,87%.

Keywords

Algorithm C 4.5, Classification, Ten Fold Cross Validation, Confusion Matrix, The risk Illness.

1. INTRODUCTION

Health insurance is insurance which provides social guarantee for people and created by the government, manage relationship between insurance and all people in country. Based on social insurance, all membership from insurance at the beginning of the registration process, participants were not checked about their health from candidates membership insurance. Adverse selection occurs when the policyholder is better able to anticipate expenses than the insurer [1], Selection (adverse or advantageous) is the central problem that inhibits the smooth, efficient functioning of competitive health insurance markets [2]. Social insurance always have problem about membership who registered and have illness before, that things will be increasing financial cost if insurance not use another approach from health services like promotive, preventive and rehabilitative services for manage illness. so far on proses to do another approach like promotive and preventive doing by department healthy like sharing to membership and collaboration with another public government like Puskesmas and Hospital. The Regulation of the Minister of Health Number 75 of 2014 concerning the Community Health Centredefines Community Health Centre (Puskesmas) as a health care facility that performs public health efforts and primary individual health efforts, with a priority on promotive and preventive efforts [3]. Primary health care is a suitable setting for interventions to identify and reduce behavioural risks factors and recommend preventive activities (including immunization, screening for cardiovascular risk factors and cancer, and counselling) [4]. In every socialization are members of participants good healthy participants, and participants who had a history of illness (participants hospital) who get health checkups and questioning about the health of participants. Further analysis of participants who had a history of certain illness is not optimal do and it is making the costs for health being higher

Based on the issue, Health Insurance need to analysis to provide solutions to problems in analysis by determining the level of risk participants used the method, data-driven decision making to analyzed further data is based on participants who have registered. Data mining algorithms used for classified data is the algorithm c4.5. Information system and information technology are becoming critical role players for any organization to achieve its goals and become a winner in this globalization and competition era [5].

Algorithms c4.5 much used and implemented in a, decision making an excess of the algorithm c4.5 is able to process numerical data and discreet, can handle the value of attributes that empty, make the rules being easy to implementation and performance is one of the fastest compared to other algorithms [6], Can predict the class of objects of unknown [7]. The algorithm has several approaches, classification like based, decision trees induction, rule-based artificial neural network, genetic algorithms and the network Bayesian [8]. The algorithm c4.5 is most popular algorithms to classification in machine learning and data processing [9]. In the algorithms c4.5 is going on the process of classified data in the form of a different level, starting from the root to, leaves the process on a tree a decision was to turn the data format (table) become a model tree, changing models trees into rule and simplify rule [10].

2. LITERATURE REVIEW

Classification is one of the data mining techniques that is mainly used to analyze a given dataset and takes each instance of it and assigns this instance to a particular class such that classification error will be least [11]. C4.5 algorithm is one classification used algorithms and produced a decision on the developed by ross Quinlan in 1979 [12]. The basis of the c4.5 algorithm is making the decision, where it based on the election attribute that has prioritized commonly called with the highest gain based on the entropy those attributes as a pivot from the classification [13]. The form of the decision would form a condition in the form of if-then. c4.5 algorithms have basic work, of making the decision and the rules model [14]. Making tree decisions on algorithm c4.5 have step - step as follows :

a. Make the selection of attributes that referred to as the root is of the value of gain is the highest of attributes - attribute of being there

- b. Making branches for each of the value of, where is the manufacture of branches in accordance with the number of variable values gain is highest
- c. Divide each case in the branch based on calculation of the value of the gain highest and calculation done after calculation of the value of the gain highest earlier and later process was completed calculation the gain highest without including variable values the gain early
- d. repeating the process of in any branch of so that all the case in branches having the same class and repeating all the more so the highest level to mark the passing of time of each branch the case of until it is no can be conducted the process of and in addition to it

The algorithm c4.5 for implementing the and the use of can do a prediction and classification on the students that could potentially register or did not register at certain universities with how to make decisions based on data tree data already exist and perform a prediction to prospective students who will sign up to the university [15]. In addition, these algorithms used for the classification revenue from sales in the predicate recommendations partner company pt the atria artha persada [16]. In the other paper, Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease [17].

2.1 Entropy and Information Gain

Entropy is the size of the theory of information that confirmed impurity and homogeneity of an assemblage of objects or data. Based on value entropy obtained, we do the calculation of the value of information the gain of each attribute [13]. Calculation of the value of entropy using equation as follows:

$$Entropy = \sum_{i=0}^{n} -pi * \log 2 \text{ (pi) ...(1)}$$

Equation 1, Calculation Entropy

The Equation 1 is used for calculation entropy and determine how informative attributes of the data Information (Equation 1) is as follows:

s : The set of case

- n : Number of Partition
- pi : The number of cases in partition to i

Information gain is popular in the selection criteria. attributes information obtained from the alteration of entropy on an assemblage of objects well informed were obtained by means of observation or obtained from conclusion by conducting participation to a set of objects or data 10 calculation of the value of using Equation information gain as follows

$$GAIN(S,A) = Entropy(S) - \sum_{i=1}^{n} -\frac{Si}{S} * Entropy(S)$$

Equation 2, Calculation information gain

Following the completion of entropy calculation using Equation (1), then afterward used the Equation (2) as

calculation information the gain, the following information from formula are:

- s : The set of case
- n : The number of partition attribute a
- $|Si| \quad : The \ number \ of \ cases \ in \ partition \ i$
- |S| : The number of cases s

Based on the Equation (1) and Equation (2), the data is obtained can be inserted and done data processing using an algorithm c4.5 for the next process of making decision tree

2.2 Decision Tree

The tree of decision is flowchart, as structure tree, where every internal node showed a test on a spot an attribute, every branch of show the result of the test, and leaf nodes show class or class distribution (Han & Khamber, 2001). The Decision tree is one of the classification techniques which is done by the splitting criteria. The decision tree is a flow chart like a tree structure that classifies instances by sorting them based on the feature (attribute) value [18]. Decision tree is a technique for representation of data in hierarchical manner [19]. As a structured tree of various attributes to do testing in the purpose is just to foresee. Tree of decision is one of the most popular data mining techniques for his discovery of knowledge, systematically to analyze and extract the rules for the purpose of classifications / the prediction of [20]

3. METHODS

The methodology used in the application of the algorithm c4.5 to classified the level of risk of illness in the broad design using research on the next Figure :



Figure 1. Flowchart Research process design

Collection of Data

The process by which conducted to gather data-data that will be used in the process of an algorithm classifications c4.5

Selection Data

Selection of data needed to choose the data used in the algorithm c4.5 classifications selection of data intended to create data set the target , the data set, or focused on a subset of variable or sample data, where the discovery will be doing [21] in this research will be done, authors prepared data training, data testing that will be used for research purposes as follows

- a. Data training : 87.225 rows data
- b. Data testing : 8.726 rows data

Transformation Data

The transformation of data is a process data and transformation of data into a format or in the presence of, so that it can be processed by calculation c4.5 algorithm. The transformation of the data can be was said to be transforming or merge the data into a desired shape / in accordance to do the process of data mining through surgery summary or aggregation is [21]. Data transformation removes noise from data and also summarizes data [22].

• Calculation Entropy and Information Gain

Calculation all attribute/variable entropy using Equation (1) and information the gain is using Equation (2) to seek and to know information gain is highest that will be used as a knotted root at the decision-making process

Decision Tree

Decision trees are the result of the process of reckoning entropy and information gain, after the calculation repeat until all the attributes have a class and cannot be done process calculations[16] [10]. The learning and classification steps of decision tree induction are simple and fast [23].

Rule Model

The rule model is the discussion and explanation to represent a tree decision [16].

Validation and Testing

Validation and testing is a process or activities are conducted to determine all the functions as it should be working with. Validation does with validation ten-fold cross-validation is validation that done by means of divide a set of data being ten segments is as large as a way of doing 13. data encryption validation and testing done to determine the level of accuracy, precision and recall the predictions. classification accuracy is showing percent of the classified correctly in testing. dataset the percentages precision data classified as a model good is also fine. Recall is the measurement of the level of the introduction of showed signs of positive sentiments actually [24].

• Analysis of testing

Analysis of what can be done is to ensure that the results of the testing were done with right and in accordance with the discussion. An analysis was done by calculation back the results of validation and testing (accuracy and etc) manually. Do a calculation that done be produce the same value or not, will be carried out through a method of confusion the matrix Confusion. the matrix is a model will form the matrix consisting of true positive or tuple true negative or positive and negative tuple [21]. Confusion the matrix can visualize the performance of classification algorithm [9].

4. RESULT AND DISCUSSION

The section on the stage and the process of algorithms classifications c 4.5 can be delivered as follows :

4.1 Collection of Data

In terms of data collection, the author asks data to health insurance the data drawn from the database that is in health insurance. A period of the data drawn at random, both in the years of data and residence of participants Indonesia. scattered over The data obtained among them are: no, no identity, participants' weight, height, the score a illness, BMI and class risk. The count of the data collected as many as 87.225 data on participants' health insurance.

NO_ID	Weight	BMI	Height	Illness_A	Illness_B	Illness_C	Illness_D
Xxx	50	Normal	150	0.9	0.9	0.6	0
Xxx	61	Normal	158	2.05	4,4	1.95	1
Xxx	85	Overweight	170	0.6	0.6	0.7	0

Table 1. Structure Health Insurance Data

4.2 Selection Data

The data were drawn begins source data on health insurance, have many variables, based on the table data structure health insurance and necessaries data to be changed in c 4.5 classification, then got some variable that underlies the decision making, including: validity no _ id, weight validation, BMI, the score illness and class illness risk.

4.3 Transformation Data

At this stage, is the stage where he did change value variable/attribute who earlier data collected from the data, planning to value adjusted that can be done in the process of advanced classifications c4.5, including:

- a. The election of attribute that will be used for c4.5 classifications algorithm
- Make attribute the validity of no_id based on the number of no_id that is on the data. Value 1 = Valid, Value 2 = Not Valid
- c. Make attribute the validity of weight based on weight limits the agency which has been set a writer. Value 1 = Valid, Value 2 = Not Valid

- d. Make attribute BMI, where data BMI taken based on height and weight data participants. In terms of attribute the height restrictions, the writer does the conversion of the value of that was originally centimeters to meters that can be done in the calculation of the BMI during follow-up. On attribute BMI, done the conversion of value that can be incorporated into the process of algorithms c4.5 classifications
- e. Make attribute scoring, illness the authors originally do conversion scoring, illness to be normal middle and higher, we do to "1","2", conversion value and
- f. "3" that can be incorporated into the algorithm. c4.5 classifications made of scoring illness, 4 type illness that is illness_ a, illness_ b, illness_ c, and illness_ d
- g. Make attribute class illness risk, the risk of illness process based on BMI compared with scoring illness of participants

Validity_No_I D (1 = Valid, 2 = Not Valid)	Validity_W eight (1 = Valid, 2 = Not Valid)	BMI	Height (Cm)	Illness_A	Illness_B	Illness_C	Illness_D	Risk Level
1	1	2	150	3	1	2	1	High
1	2	1	143	1	1	1	1	NotValid
1	1	1	178	1	2	2	1	Medium
1	1	1	148	1	1	1	1	Low

Table 2. Final Result from Transformation Data

4.4 Calculation Entropy and Information Gain

The gain, entropy, and information performed calculations based on the Equation (1) and Equation (2). The first step in performing calculations entropy including

a. Calculate total, entropy total obtained by doing calculations all attribute classes in the high risk medium and low in comparison to the total data

Entropy (Total)
$$\left(-\frac{46309}{87225} * \log 2\left(\frac{46309}{87225}\right)\right)$$

+ $\left(-\frac{30540}{87225} * \log 2\left(\frac{30540}{87225}\right)\right)$
+ $\left(-\frac{10406}{87225} * \log 2\left(\frac{10406}{87225}\right)\right) = 0,4157$

Calculation Entropy and do calculation entropy from another variable...

b. Calculate gain information. information gains made in using Equation (2)

$$Gain (Total, Noka) = 0,4157 - \left(\left(\frac{85531}{87225} * 0,4168 \right) + \left(\frac{1724}{87225} * 0,3385 \right) \right) = 0,013661$$

alculation Information Gain and do Calculate information gain from another variable...

c. Entropy and gain information, the calculation is done the result of calculation included in Figure 4. Based on the calculation on the entropy and gain information the information can be taken that the highest gain in HT, attribute then attribute JK, DM and GK. Highest attribute Will be the roots of trees decision

Variabel	Value	Count of Case	Low	Medium	High	Entropy	Information Gain
All Risk		87.225	46.309	30.540	10.406	0,415725	
Noka							0,013661
	Valid	85.531	45.149	30.070	10.312	0,416851	
	Not Valid	1.724	1.160	470	94	0,338549	
Weight							0,005666
	Valid	86.619	46.091	30.164	10.364	0,415661	
	Not Valid	636	218	376	42	0,372277	
BMI		87.225					0,061715
	Not Obesity	80.094	46.309	27.269	6.516	0,385528	
	Obesity	7161	0	3271	3890	0,000000	
Illness_A		87.225	71.196	13.293	2.736	0,243654	0,261621
Illness_B		87.225	54472	27181	5572	0,361800	0,325636
	_						
Illness_C		87.225	68694	17768	793	0,241005	0,277206
Illness_D		87.225	72108	14971	146	0,204348	0,205032

Figure 2. Result from calculate Entropy and Information Gain

4.5 Decision Tree

Based on the calculation of entropy and gain information that has been conducted to 4.4, the next step is making the data into a decision tree. The decision the: are as follows

International Journal of Computer Applications (0975 – 8887) Volume 177 – No. 37, February 2020



Figure 3. Decision Tree

4.6 Rule Model

Based on the decision made in point 4.5, above now it is making the rules model for ascertaining the degree of risk illness of participants in health insurance rule of them are as follows:

- 1. IF Scoring Illness = 3, then Recommendation level of risk : High
- 2. IF Scoring Illness = 1 and BMI = Not Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : Low
- 3. IF Scoring Illness = 2 and BMI = Not Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : Medium
- 4. IF Scoring Illness = 3 and BMI = Not Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : High
- 5. IF Scoring Illness = 1 and BMI = Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : Medium
- 6. IF Scoring Illness = 2 and BMI = Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : High
- 7. IF Scoring Illness = 3 and BMI = Obesity and Validation Noka = Valid and Validation Weight = Valid, then The recommendation of the level of risk : High
- 8. IF Scoring Illness = 1 and BMI = Not Obesity / Obesity and Validation Noka = Not Valid, then The

recommendation of the level of risk : Data Not Valid

- 9. IF Scoring Illness = 2 and BMI = Not Obesity / Obesity and Validation Noka = Not Valid, then The recommendation of the level of risk : Data Not Valid
- 10. IF Scoring Illness = 3 and BMI = Not Obesity / Obesity and Validation Noka = Not Valid, then The recommendation of the level of risk : Data Not Valid
- 11. IF Scoring Illness = 1 and BMI = Not Obesity / Obesity and Validation Noka = Valid and Validation Weight = Not Valid, then The recommendation of the level of risk : Data Not Valid
- 12. IF Scoring Illness = 2 and BMI = Not Obesity / Obesity and Validation Noka = Valid and Validation Weight = Not Valid, then The recommendation of the level of risk : Data Not Valid
- 13. IF Scoring Illness = 3 and BMI = Not Obesity / Obesity and Validation Noka = Valid and Validation Weight = Not Valid, then The recommendation of the level of risk : Data Not Valid

Decision tree and rule / the model that has been reported as above, pertaining to items are:

- a. The rules of validation noka or validation weight should that not fulfilled, so recommendations to use the classification levels of risk can't be done. The result of the data if not valid is data is not valid
- b. Variable BMI is one of trigger that causes variable scoring illness will change to the process risk the end of classification. In this case, the level of risk

will go up 1 when BMI from participants is obesity, in this case as classifications scoring classification low risk be medium and classification Scoring medium be classifications of the high risk.

4.7 Validation and Testing

The process of validation and examining from the discussion in chapter 4, done testing using cross-validation, as ten-fold cross-validation. The following data were process:

Correctly Classified Instances	8676	99.427 %
Incorrectly Classified Instances	50	0.573 %
Kappa statistic	0.9905	
Mean absolute error	0.0057	
Root mean squared error	0.0532	
Relative absolute error	1.8748 %	
Root relative squared error	13.6929 %	

Figure 4 : Confusion Matrix

а	b	с	d	< Classified as
197	21	3	15	a = DatanotValid
0	4493	0	0	b = low
0	0	2970	0	c = medium
0	0	11	1016	d = high

Figure 5. Result Process Data and Confusion Matrix

Based on Figure 4 and 5, the following conclusions obtained:

a. classifications correctly is 99,427% or 8676 rows data

b. classifications wrong is 0,573% / about 50 rows data

c. the total data of testing about 8726 rows

4.8 Analysis of testing

The result of the process above data available for analysis, do testing again by calculation confusion the matrix manually. Confusion the matrix is one of the models used to a process of analysis of the results of testing. The following results, the conversion of among done based on the data point 4.7

		Low	Medium	High
		а	b	С
Nilai Prediksi	a (low)	4493	0	0
	b (medium)	0	2970	0
	c (high)	0	11	1016

Figure 6. Calculation from Confusion Matrix

Accuracy = $\left(\left(\frac{4493+2970+1016}{4493+2970+1016+11}\right) * 100\%\right) = 99,87\%$

Based on the calculation of the manual above, the results of confusion the matrix to the accuracy 99,87 %. Based on this data, thus it can be said the calculation accurate and in good category. Based on these results researchers concluded that the implementation of algorithms c4.5 on health insurance successfully implemented well, and can be used as / help health insurance in determining the risk participants, so as to be done by handling advanced health insurance

5. CONCLUSION

Based on the result, conclude that :

1. On health insurance, the application of algorithm classifications c4.5 be able to be implemented to find/know the level of risk illness from participants' health insurance. Performed validation and viewed based on the results of accuracy, where this stated that calculation has been done able to perform a

prediction or classify level risk illness in its good health insurance.

- 2. Rule model/rules has made, can be used as reference when will be developed further in the form of application
- 3. Based on the little research that has been done, the algorithm classifications c4.5 can be used to determine the level of any risk of the participants in health insurance

6. REFERENCES

- M. V. Pauly, P. Danzon, P. Feldstein, and J. Hoff, "A Plan For 'Responsible National Health Insurance," J. Med. Soc. N. J., vol. 77, no. 3, pp. 207–208, 1991.
- [2] M. Geruso and T. J. Layton, "Selection in health insurance markets and its policy remedies," J. Econ. Perspect., vol. 31, no. 4, pp. 23–50, 2017.
- [3] Ministry of Health RI, Indonesia Health Profile 2017 Ministry of Health of the Republic of Indonesia 2018. 2018.
- [4] C. Brotons, "Prevention and Health Promotion in Clinical Practice," no. 4, pp. 1–3, 2015.
- [5] M. Sadikin, I. Nurhaida, D. Fitrianah, A. R. Dwiyanto, Harwikarya, and M. M. Sarinanto, "IS Strategic Plan for Higher Education Based on COBIT Assessment: A Case Study," Int. J. Inf. Educ. Technol., vol. 5, no. 8, pp. 629– 633, 2015.
- [6] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," Int. J. Mod. Trends Eng. Res., vol. 4, no. 7, pp. 58–63, 2015.
- [7] Y. R. W. U. Joko Purnomo, Wawan Laksito, "Aplikasi Penunjang Keputusan Penerimaan," Implementasi Algoritm. C 4.5 dalam Pembuatan Apl. Penunjang Keputusan Penerimaaan Pegawai CV. Din. Ilmu, vol. 2, 2014.
- [8] M. Aryuni and E. D. Madyatmadja, "Feature selection in credit scoring model for credit card applicants in XYZ bank: A comparative study," Int. J. Multimed. Ubiquitous Eng., vol. 10, no. 5, pp. 17–24, 2015.
- [9] A. Wijaya and A. S. Girsang, "Use of Data Mining for Prediction of Costumer Loyalty," CommIT (Communication Inf. Technol. J., vol. 10, no. 1, p. 41, 2015.
- [10] D. W. T. Putra, "Algoritma C4.5 untuk Menentukan Tingkat Kelayakan Motor Bekas yang Akan Dijual," J. TEKNOIF, vol. 4, no. 1, pp. 16–22, 2016.
- [11] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," Int. J. Electr. Comput. Eng., vol. 8, no. 6, p. 4763, 2018.
- [12] S. A. Lusinia, "Algoritma C4.5 dalam Menganalisa Kelayakan Kredit (Studi kasus di Koperasi Pegawai Republik Indonesia (KP-RI) Lengayang Pesisir Selatan, Painan, Sumatera Barat)," Issn 2356-0010, vol. 1, no. 2, pp. 6–10, 2014.
- [13] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," J. KomTekInfo Fak. Ilmu Komput., vol. 2, no. 2, pp. 33– 38, 2016.

- [14] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," Issn 2548-8368, vol. Vol 1, No, no. June, p. 7, 2017.
- [15] Y. Andrian and M. R. Wahaydi, "Analisis Kinerja Data Mining Algoritma kampus ABC jurnal 12," 2012.
- [16] M. F. Arifin and D. Fitrianah, "Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada," no. January 2018, 2018.
- [17] W. Wiharto, H. Kusnanto, and H. Herianto, "Interpretation of clinical data based on C4.5 algorithm for the diagnosis of coronary heart disease," Healthc. Inform. Res., vol. 22, no. 3, pp. 186–195, 2016.
- [18] S. Sharma, J. Agrawal, and S. Sharma, "Classification Through Machine Learning Technique: C4. 5 Algorithm based on Various Entropies," Int. J. Comput. Appl., vol. 82, no. 16, pp. 28–32, 2013.
- [19] P. Gulati, A. Sharma, and M. Gupta, "Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors"

for Performance Improvement: A Review," Int. J. Comput. Appl., vol. 141, no. 14, pp. 19–25, 2016.

- [20] M. G. Mohiuddin and P. Premchand, "Performance Analysis of Decision Tree Classifiers," Int. J. Comput. Sci. Trends Technol., vol. 2, no. 2, pp. 62–70, 2014.
- [21] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C 4.5 untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta," J. Tek. Komput. AMIK BSI, vol. II, no. 1, pp. 132–137, 2016.
- [22] D. S. Kumar and K. Swati, "A Comparative Study of Various Data Transformation Techniques in Data Mining," Int. J. Sci. Eng. Technol., vol. 4, no. 3, pp. 146–148, 2015.
- [23] B. N. Patel, S. G. Prajapati, and D. K. Lakhtaria, "Efficient Classification of Data Using Decision Tree," Bonfring Int. J. Data Min., vol. 2, no. 1, pp. 06–12, 2012.
- [24] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, "Predicting Critical Courses Affecting Students Performance: A Case Study," Procedia Comput. Sci., vol. 82, no. March, pp. 65–71, 2016.