Using classification data mining methods to predict the level of efficiency of services in dental clinics during the COVID-19 pandemic

Anhar K.ALDeen Mohammed[1*] , Reem Ali AL-Jarah[2]
[1,2]College Of Dentistis, Mosul University,IRAQ
Rivrskm2023@yahoo.com[1]
reemaljarah1@yahoo.com[2]

Using classification data mining methods to predict the level of efficiency of services in dental clinics during the COVID-19 pandemic

Anhar K.ALDeen Mohammed[1*] , Reem Ali AL-Jarah[2]
[1,2]College Of Dentistis, Mosul University,IRAQ
Rivrskm2023@yahoo.com[1]
reemaljarah1@yahoo.com[2]

Abstract

The Covid-19 pandemic has always affected all life facilities, dental clinics, like other institutions. research goal is to reach the patient's evaluation of the competency of the service offered to him in the clinic during the pandemic.

Three classification- data mining algorithms -decision tree, logistic regression and cluster analysis- were used to rank clinic reviewer opinions.

Using the programming languages (HTML, PHP, My-SQL) an electronic system has been created that provides services and facilitates the procedures for organizing reservations and making appointments.....Etc., according to the necessary, safety instructions during the pandemic. The System Development Lifecycle (SDLS) methodology is used to determine the level of service efficiency, and ODBC is used to send data from the database to SPSS-V26.

The study variables, like the possibility of returning to the clinic, which has the greatest potential to classify observations and is contributing the most to differentiating for each of the two clusters, have a statistically significant relationship with the likelihood that you will recommend this clinic to others, A list of findings were included in the research's conclusion..

Keywords: Database, Relational Database, SQL, ODBC, Multivariable Analyses , CHIAID algorithm, Logistic Regression algorithm, Two-Step Clustering algorithm, Dental Clinics.

1. Introduction

Health institutions seek to prove their existence, by providing the best possible services, especially providing the best is the key to achieving progress and excellence to reach the highest levels of satisfaction, as the patient has become the focus of attention.

many clinics have stopped working due to the risks faced by health workers, at a time when demand has decreased due to patients' "fear" of resorting to services. And while the dentists are taking more precautions at the moment, many of these procedures are just an extension of their regular routine. A spokesperson for the American Dental Association stated: **"In dentists'** offices, we used to take comprehensive preventive measures, such as disinfection and personal protective equipment (PPE), before they became common**."**

The patient's condition, "may be unprepared, worried about his health, or looking after an elderly relative, or afraid to see a doctor, so he acts cautiously and does not want to take risks", Thus, the person responsible for fixing the reservation in the clinic is constantly working to deal with the many postponed appointments, and communicates with patients who are scheduled to undergo treatment to organize alternative appointments. In our research, we highlight the level of services in government and private dental clinics during the COVID-19 pandemic and their opinion of the preventive and curative services provided. In this research , Data mining techniques were used, such as algorithm of (CHAID tree, logistic regression and clustering), to reach the goal of the search.

There are some studies were done in the same filed but in another place ,such as a study got in a "Valencia primary dental care facility- SPINE" to assess parents' contentment with the dental care

their kids, by using CHAID algorithm and decision tree , the researchers conclude the parents who participated in study were contented with the of their children received[1]. Also, by using the statistical modeling and CHAID analysis ,. et.al. [2] . conclude that there is an urgent need to create inclusive public policies because the state's epidemic is getting worse. Another study for [3] . It uses conventional psychometric measures to assess healthcare personnel's reported psychological well-being by using the decision tree model

Not only in the health field, the CHAID algorithm was applied. A case study: the National Opinion Research Center of the U. S. chose to assess the job satisfaction in the U. S. and and the factors affecting it.[4].

In order to better the education institution, researchers took refuge to use the CHAID algorithm and decision tree to examine students' perceptions of the academic program and student services offered by the university to discuss how to upgrade their establishment.[5].

2. Theoretical Part

There are several techniques in data mining, the selection of the appropriate ones depends on the nature and size of the data under study[6], Data mining can be implement in comparison with the data market and data store. Figure 1 showing the data mining technique which can use .

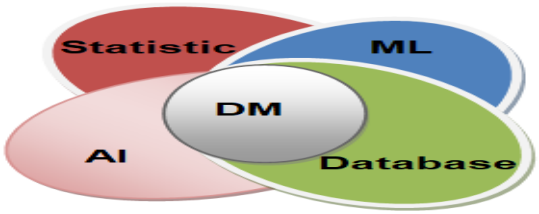


Figure 1. Data Mining Technique
Sourc:[6]

There is no specific theory based on to choose the appropriate data minig technique, Usually the choice is made based on the experience in this field and the actual techniques effective, the superlative between traditional techniques and modern techniques to the extent that the appropriate tools are available, However, with the increase of experience, we can evaluate the options and select the appropriate ones for the application. [7] .

2.1 Data Mining Models

As Figure 2 showe, the Predictive Model : It works on the future prediction of the mechanism of data work, based on a model whose data has been previously trained.[8][9] .

The Descriptive Model : It analyzes the data in order to extract the relationships and correlations between them, and to reveal the characteristics of the data to build a model that did not exist previously. For instance, Clustering Analysis،Association Rules Analysis ،..etc

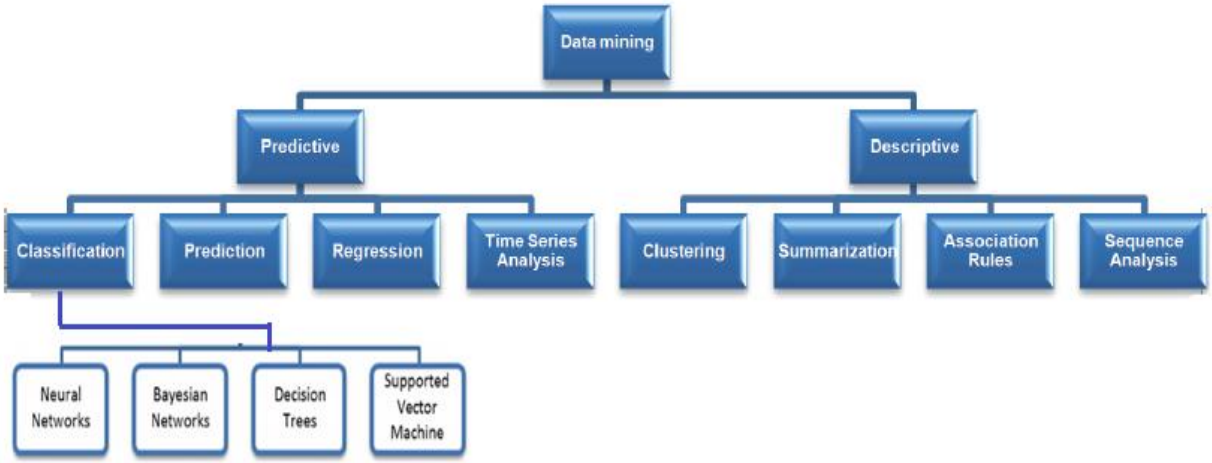


Figure 2. Dataminig
Sourc:[9]

2.2 Data Mining Technique

Research progress in data mining and knowledge discovery, has led to the many of data mining techniques coming to lights , which are built on three strong technical pillars: database, mathematical statistics, and artificial intelligence, our interest in this researchare

2.2.1 Classification

The major goals of a Classification algorithms are to make the predictive accuracy obtained as large as possible by the classification model[6] [10] .

2.2.1.1 Decision Tree

"uses as a predictive model which maps observations about an item to conclusions about the item's target value".[11] [12], The algorithm which used in this research is CHAID ("Chi-squared Automatic Interaction Detection") [13], it's a technique depends on the χ2_test - eq.1- and employs a step-by-step procedure

The creating of nodes and sub-nodes among the predictor variables is significant.[14] [15].

$$\chi^2 = \sum \frac{(E_{ij} - O_{ij})^2}{E_{ij}} \dots\dots\dots (1)$$

2.2.2 Logistic regression

is "a method in which the Y response variable has a binomial distribution. To determine probabilities of success P(Y = 1|X), and defeat P(Y = 0|X).[16] [17] The odds are defined as in eq.2" .[18] .

$$(\Theta) = \frac{P(Y = 1|X)|}{P(Y = 0|X)} = \frac{P(Y = 1|X)}{1 - P(Y = 1|X)} \dots\dots\dots (2)$$

"The linear dependencies between log-odds-logit- and input variables are analyzed as" in eq.3

$$\ln(\Theta) = \ln\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right) = X\beta \dots \rightarrow P(X, \beta) = \frac{e^{X\beta}}{1 - e^{X\beta}} \dots\dots\dots (3)$$

where $\beta = (\beta 1, \dots, \beta m) \in R^m$. From Equation (3) [19]

2.2.3 Two-Step Cluster algorithm

"is a statistical method[9] for assigning data into subgroups with related characteristics" [7].

The procedure is performed by constructing a cluster feature (CF) tree, which contains the cluster centers.

"To automatically determine the number of groups, the method uses BIC (Schwartz Bayesian Information Criterion) or AIC (Akaike Information Criterion) which is calculated for each number of clusters from a given range; This indicator is then used to find a preliminary estimate of the number of clusters."[20] For groups J, the two indicators are calculated according to equations (4) and (5) below:

$$BIC(J) = -2\sum_{j=1}^{J} \xi_j + m_j \log N \dots\dots (4)$$

$$AIC(J) = -2\sum_{j=1}^{J} \xi_j + 2m_j \dots\dots\dots (5)$$

. *where* .

$$m_J = J\left\{2K^A + \sum_{k=1}^{K^B}(L_k - 1\right\} \dots (6)$$

$$t = \frac{\hat{\mu}_k - \hat{\mu}_{sk}}{\hat{\sigma}_{sk}}\sqrt{N_k} \dots\dots\dots\dots (7)$$

$$.\chi^2 = \sum_{l=1}^{L_k}\left(\frac{N_{skl}}{N_{kl}} - 1\right)^2 \dots\dots\dots (8)$$

3.Materials & Methods

In order to obtain the necessary data, a questionnaire - with three point Likert Scale (Agree, Neutral, Disagree)- was designed that includes factors of interest to dental clinic references during the Covid-19 pandemic, and distributed to reviewers in different dental clinics.

X1:Gendar, **X2**:Age(20-,30-,40-,50-59), **X3**: Kind's clinic (Private, general)**,** **X4**: Help in booking an appointment at the clinic**,** **X5:** The receptionist pays attention to you and obliges you to follow prevention methods (wearing a mask and gloves), **X6**: The time between your request for an appointment and your presence at the dentist, **X7**: Adequate working hours, **X8**: Cleanliness of the clinic and the provision of means of dusting, sterilization and appropriate ventilation, **X9**: Conducting a preliminary examination for those who enter the clinic as an indicator of temperature or something else, **X10**: Not allowing the patient to accompany him, except in critical cases, **X11**: Availability of a waiting room with social distance, **X12**: acceptable waiting period for the patient to see the dentist, **X13**: Availability of a worker responsible for wiping surfaces and door knobs periodically during working hours, **X14**: Measures taken to prevent infection in the patient's examination room, **X15**: The measures taken by the clinic staff to protect you from infectious diseases and radiation, **X16**: The period of stay for the dentist with you, **X17:** The interest shown by the dentist about your questions or concerns, **X18**: Explain your treatment options in an understandable manner, **X19**: The competence of the dental assistant and his interest in guiding and educating you to maintain oral and dental health, **X20**: Information provided on ways to avoid dental problems in the future, **X21**: Comprehensive examinations and treatment, **X22**: The interest shown by the dentist in general, **X23**: Overall evaluation of the care provided in the dental clinic, **X24**: Collaboration and teamwork demonstrated by the dentist's team, **X25**: General assessment of the skill of the dentist, **X26**: Your trust in the dentist, **X27**: Possibility to review the clinic again when needed, **Y**: Chances that you will recommend this clinic to others. Reliability was calculated using Cronbach's alpha.

3.1 Databases System

Databases are designed to suit the movement and flow of data in addition to organizing the operations of saving, recovery and modification, in consideration the availability of access and protection, and the possibility of expanding these databases, whether in terms of the increase in the number of users or data or tables and the resulting complexity in query sentences, Figure 3 shows – respectively- the database table and relation database of our system

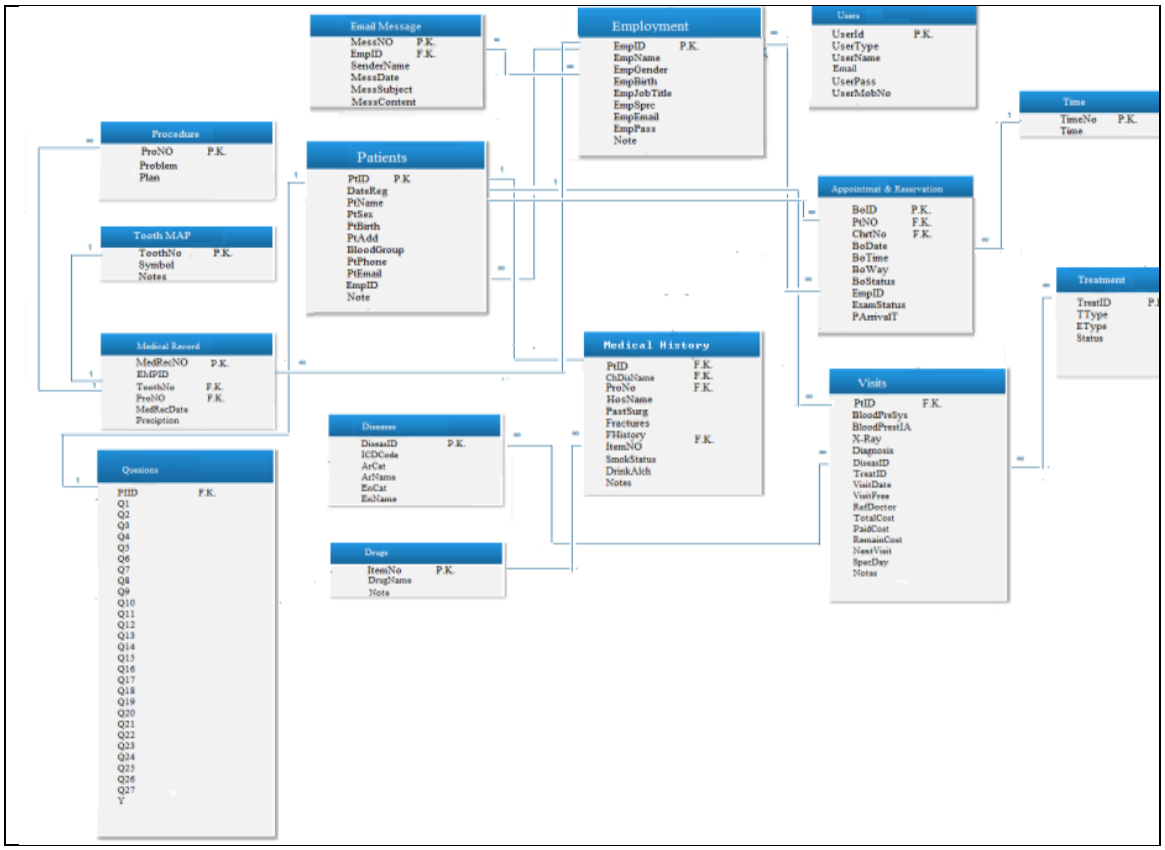


Figure3. Relational Database

To take advantage of the volume of available data and the above-mentioned databases, an has been created, in line with modern technology in the use of databases, which can be linked to other systems.

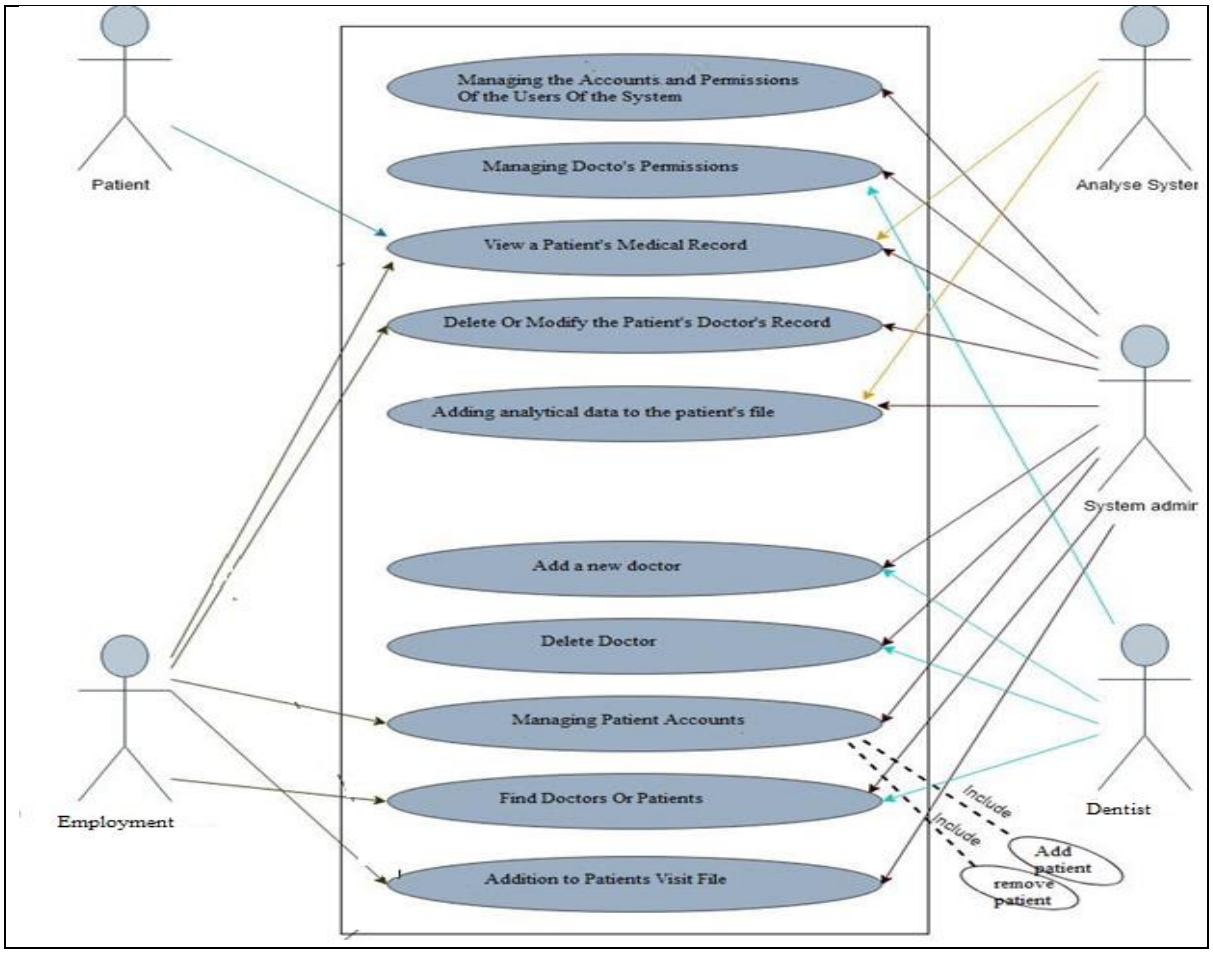


Figure4. analyze system(User Case)

When logging in, and in order to be ensure about the protection of information, the user is allowed to log into the system with his name, password and doctor ID number.

The system, not only doesn't allow its users to access the account of others and see their private information, but also restricts the content of the screens according to the user permission.

The (PHP) language ensures protection by rejecting repeated attempts to log in, NOSQL Injection, and activating all authentication functions to ensure the correct entry.

Some features of the system are Adding, Viewing, Modifying, Deleting, And Searching.

The doctor or his assistant are allowed to add everything new to the patient's visit profile, starting from the status of his booking - new, confirmed, postponed, canceled - and the required data on the visit schedule, new patient mean new file open, View the patient's medical record by entering his number , modify or delete the patient's medical record , Allow users to inquire about doctors or patients by number, full name, or username of the doctor or patient.

3.2 CHAID Algorithm

Table 1. Model Summary Of CHAID methode

Specifications	Dependent Variable	Y
	Independent Variables	$X_1 - X_{27}$
	Validation	none
	Maximum Tree Depth	3
	Minimum Cases in Parent Node Minimum Cases in Child Node	100
	Minimum Cases in Child Node	50
	ndependent Variables Included	X_1 , X_7 , X_{27}

Results	Number of Nodes	7
	Number of Terminal Nodes	4
	Depth	3

Tree diagram in Figure 5 show, the most significant independent variable is **X_{27}**: "Possibility of revisiting the clinic",

X_{27} at α=0.001: ($\varkappa^2 = 165.221, df = 2, p_{value} = 0.000$). As the first steep, it splits the sample of (407) responses, into 2 groups containing different groups of X_{27} presented as node 1 & 2. The first node, included (215) respondents from the variable X_{27} which coded (3: disagree), while the second one is terminal node included (92) respondents from the variable X_{27}, which coded (1: agree, and 2: neutral) within tree's second level , one statistically significant variable Adequate working hours (X_7) is identified at α=0.05, ($\varkappa^2 = 14.853, df = 2, p_{value} = 0.002$), which in turn leads us to the third and fourth node. Node3, included (150) respondents from the variable X_7 which coded (3: disagree), while Node4, included (50) respondents from the variable X_7 which coded (1: agree, & 2: neutral).

Within tree's third level(final) :

Gender: X_1, it is significant for splitting of Nodes5 and Nodes6 at α=0.05, $\varkappa^2 = 6.253, df = 1, p_{value} = 0.012$

Generally, the leaf tree frame, (fixed as 4, 5 & 6) pertain to (3: disagree with recommend this dental clinic to others), and one (marked as 2) pertain to (2: indifferent to recommend this dental clinic to others).

In fact, the track from the root to the leaf nodes produce some rules for classifying reviewers' opinions (in one of the variable-specific categories, for example, the rule in Node6: Who does not want to visit the dental clinic again, and the working hours are not suitable for him , 89.6% would not advise anyone else to visit this clinic, and the other rules derived can be explained by a similar tactic.

To finish the process of classification, all its performance have been assessed. Our findings suggest that if the reviewers' opinions in terms of the 3 indep.var. are known, the hazard of inaccurately classified -based on entire sample- is 17.9%, with std. Error is 0.02.

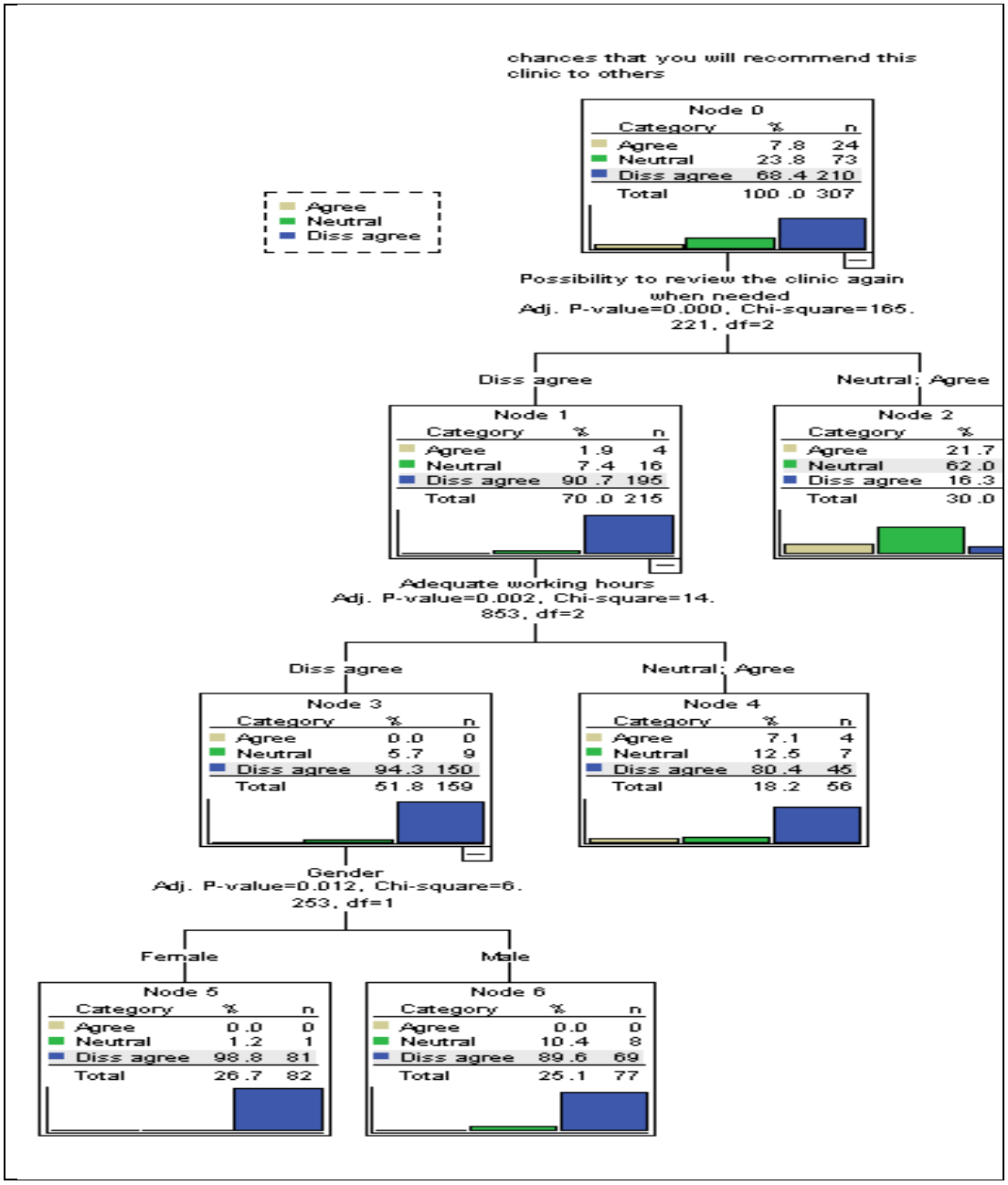


Figure5. Decision tree analysis with CHAID model.

Table 2. presents the classification CHAID

Reviewers' Opinions		Predicted			Correctly Classified
		Agree	Neutral	Diss. agree	
Observed	1:Agree	0	20	4	0.0 %
	2:Neutral	0	57	16	78.1 %
	3:Diss-agree	0	15	195	92.9 %
	% Over all	0.0	30.0	70.0	82.1 %

In accordance with the table 2 , It can be said that in general the accuracy of the model is 82.1%, the model accurately ranked (252) out of (297) in the sample under discuss. Significant differences can be observed in the classification accuracy, according to the categories of the variable.

3.3 Multinomial Logistic Regression (MLR) Algorithm

The(MLR) model's estimation and testing steeps is "The chances that you will recommend this clinic to others".

Independent categorical variables (27) were selected for inclusion in the multivariable logistic regression model. As shown in Table 3, the χ^2 test for the last model, the 2 - Log Likelihood for the full model (86.835) less than it is for the null model (491.546), the likelihood ratio represents the unexplained variance in the outcome variable.

The Likelihood Ratio chi-square test which is a test of goodness-of-fit, for model predicts, χ^2 (106) = 404.710, p < 0.001,

Table 3. Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	491.546			
Final	86.835	404.710	106	.000

the resp. var. explained by the predictors indep.var., 73.2% of the variance in dep. variables (the proportion of variance) according to "Cox and Snell R^2 " value, 91.7 % according to " Nagelkerke R^2 " value, and 82.3% according to "McFadden R^2" value.

Table 4 illustrates the likelihood ratio test, As the results shown, it can be seen that relationship between some of the independent variables (X_{27},X_{26},X_{24},X_{21} respectively) and the dependent variable is a statistically significant

Table 4. Evaluate the relationship between variables and the dependent variable

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	2_Log Likelihood of Reduced Model	Chi-Square	DF	Sig.
Intercept	86.835	.000	0	.
X_1	88.918	2.082	2	.353
X_2	90.194	3.358	6	.763
X_3	90.223	3.387	2	.184
X_4	94.973	8.138	4	.087
X_5	92.836	6.001	4	.199
X_6	88.825	1.990	4	.738
X_7	88.648	1.813	4	.770
X_8	87.794	.959	4	.916
X_9	87.979	1.143	4	.887
X_{10}	86.845	.009	4	1.000
X_{11}	88.213	1.378	4	.848
X_{12}	89.333	2.497	4	.645
X_{13}	94.873	8.038	4	.090
X_{14}	90.868	4.033	4	.402
X_{15}	88.836	2.001	4	.736
X_{16}	88.023	1.188	4	.880
X_{17}	91.994	5.159	4	.271
X_{18}	88.067	1.232	4	.873
X_{19}	88.015	1.179	4	.881
X_{20}	89.694	2.859	4	.582
X_{21}	96.904	10.068	4	.039
X_{22}	92.380	5.545	4	.236
X_{23}	88.697	1.862	4	.761
X_{24}	96.790	9.955	4	.041
X_{25}	86.933	.098	4	.999
X_{26}	101.653	14.817	4	.005
X_{27}	394.129	307.294	4	.000

Correct classification rates for the MLR model as shown in Table 5 , It indicate that the accuracy of the model is 95.4, and the expected values according to the dep. variable (7.8, 22.5, 69.7) % isn't significantly different from that in the main data (100.0, 87.7, 97.6) %.

Table5. classification rates for the MLR model

Observed	Predicted	Correctly Classified

	Agree	Neutral	Disagree	
1:Agree	24	0	0	100.0%
2:Neutral	0	64	9	87.7%
3:Diss-agree	0	5	205	97.6%
% Over all	7.8%	22.5%	69.7%	95.4%

3.4 Cluster Analysis

As shown in table 6, although the lowest "BIC coefficient" is for 7 clusters, according , the optimal No. of clusters is two, because the lowest value of the Schwarz's Bayesian Criterion as well as the largest proportion of space is for two clusters

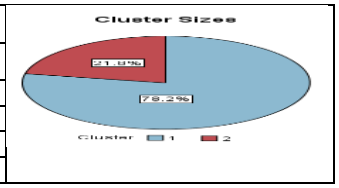
Table 6.Cluster Distribution

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change	Ratio of BIC Change	Ratio of Distance Measures
1	13879.447			
2	12679.278	-1200.170	1.000	2.198
3	12304.983	-374.295	.312	1.579
4	12183.328	-121.654	.101	1.117
5	12107.391	-75.937	.063	1.239
6	12106.766	-.625	.001	1.382
7	12193.355	86.589	-.072	1.052
8	12291.273	97.917	-.082	1.070
9	12403.394	112.122	-.093	1.130
10	12538.898	135.504	-.113	1.065

The clusters distribution is shown in Table 7

Table 7. Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	240	78.2%	78.2%
	2	67	21.8%	21.8%
	Combined	307	100.0%	100.0%
Total		307		100.0%



From cluster quality in Figure 6 show the value of average silhouette is 0.358

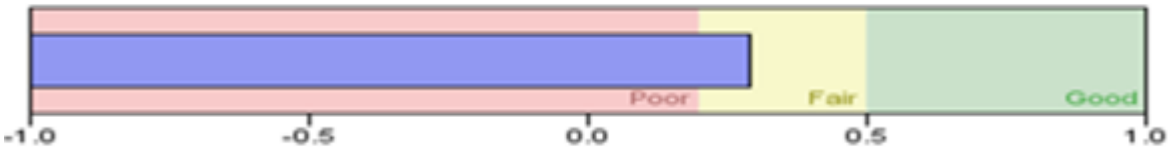


Figure6. silhott measure of cohesion separation

As our interest is to know the important variables that affect the opinion of reviewers to visit the dental clinic during the Covid-19 pandemic, Figure 7 show that X_{27} , X_{21} then X_{26},X_7,X_{24}, X_{13},X_{17} are the most contribute to differentiating the first cluster and X_{27} , X_{24} , X_{21}, then X_{13}, X_{17} , X_1 differentiate the second

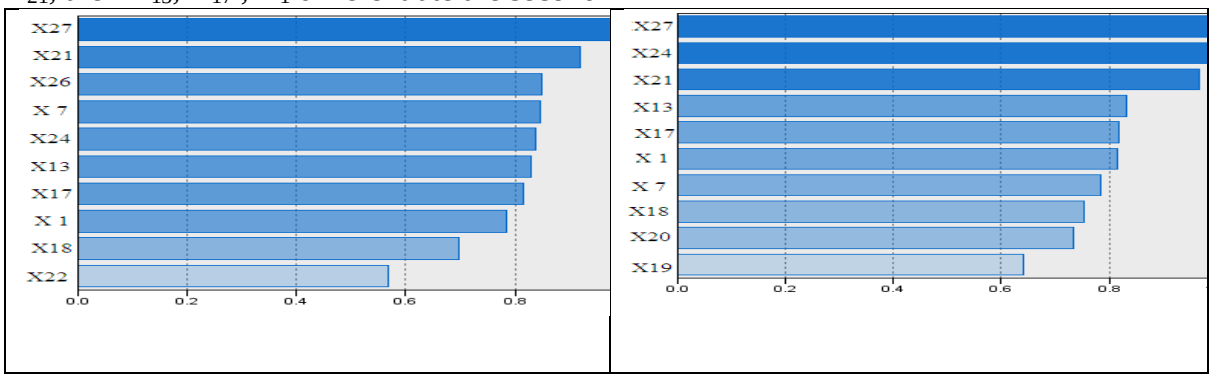


Figure7. Variables Importance for Clusters one & cluster 2

4. Conclusion

In accordance with CHAID Algorithm, In general the model accuracy is very good. Although the differences can be observed in the classification accuracy, the our data the variable X_{27}: Possibility of revisiting the clinic has the biggest potential to classify observations

In accordance with Multinomial Logistic Regression Algorithm (MLR), The change value of Model Fitting Criteria for the final model and the null model indicate better fit, and the test of goodness-of-fit indicates significant model predicts. From Cox and Snell R2 value, Nagelkerke R^2 value, McFadden R^2 value, the predicted model was able to explain the variance in the conviction of individuals well. There is a statistically significant relationship between: (X_{21}: Comprehensive examinations and treatment), (X_{24}: Collaboration and teamwork demonstrated by the dentist's team), (X_{26}: Your trust in the dentist), (X_{27}: Possibility to review the clinic again when needed) and the dependent variable (y: Chances that you will recommend this clinic to others). The correct classification rates for the MLR model are highly valued, on other hand, there is no significant different between the expected values of the dep. variables and the main data.

In accordance with 2-steeps cluster algorithm indicate that The optimal number of clusters is two, with size 78.2% and 21.8% respectively. From silhouette measure, the cluster quality is fair. The categorical variables which contribute the most of differentiating for each two cluster are: (X_{27}: Possibility to review the clinic again when needed) ,(X_{24}: Collaboration and teamwork demonstrated by the dentist's team) and (X_{21}: Comprehensive examinations and treatment).

Reference:

[1] Carmen Llena; Gonzalo Clemente and Leopoldo Forner,2010," Parental Satisfaction with Children's Primary Dental Care in Valencia, Spain", Primary Dental Care 2011 ; 17(4):25-30, DOI:10.1308/135576111794065739

[2] Fabrício P. A. ; Fabrício Pelizer Almeida ; Moisés Keniel Guilherme de Lima ; Demóstenes Coutinho Gomes ; Esther Ferreira de Souza,2020, "Logistic Regression Models and Classification Tree for Deaths and Recovered Patients Records of Covid-19 in the State of Minas Gerais, Brazil", IJAERS, Vol-7, Issue-10,OCT.(2020), ISSN: 2349-6495(P) | 2456-1908(O), https://dx.doi.org/10.22161/ijaers.710.16

[3] Anirban Ray ; Debjani Sen Gupta ; Jayasree Sar Choudhury ; Pramit Ghosh , "Resilience of Healthcare Providers during COVID-19 Pandemic: A Rapid Assessment using Digital Platform", Journal of Clinical and Diagnostic Research.2022 Dec, Vol-16(12): VC01-VC07
DOI: 10.7860/JCDR/2022/53497.17198

[4] Farhad Sheybani; 2019, "Predicting the Individuals' job satisfaction and determining the factors affecting it using the CHAID Decision Tree Data Mining Algorithm, Case Study: the National Opinion Research Center of the United States.", EJERS, European Journal of Engineering Research and Science, Vol. 4, No. 3, March 2019, DOI: http://dx.doi.org/10.24018/ejers.2019.4.3.1169

[5] Maryli F. Rosas; Shaneth C. Ambat; Alejandro D. Magnaye; John Renbert F. Rosas, 2019, "Classification and Regression Decision Tree: A Mining Technique for Students' Insights on the University Services with Text Analysis", 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) , DOI: 10.1109/UEMCON44053.2018

[6] Dimitrios Papakyriakou & Loannis S. Barbounakis, " Data Mining Methods: A Review", International Journal of Computer Application, Volume 183, No.48, 2022,5-19. DOI: 10.5120/ijca2022921884

[7] Kun Gao; Hassan Ali Khan; Wenwen Qu, 2022, "Clustering with Missing Features: A Density-Based Approach", Song-Kyoo (Amang) Kim and Chan Yeob Yeun, 14(1),60, 2022, https://doi.org/10.3390/sym14010060

[8] Gaurav Gosavi; Gajanan Gawde, 2022, "Prediction Techniques for Data mining", National Conference on Emerging Trends in Computer Engineering and Technology.

[9] Kwame Boakye Agyapong; Dr. J.B Hayfron-Acquah; Dr. Michael Asante,2016, "An Overview of Data Mining Models(Descriptive and Predictive)",: International Journal of Software & Hardware Research in Engineering, Vol. 4 ,Issu. 5, ISSN-2347-4890.

[10] Sudhir M. Gorade; Ankit Deo; Preetesh Purohit ,2017, "A Study of Some Data Mining Classification Techniques", International Research Journal of Engineering and Technology (IRJET), , e-ISSN: 2395 -0056 , p-ISSN: 2395-0072 Vol. 04 Iss. 04,.

[11] Himani Sharma ; Sunil Kumar,2013, "A Survey on Decision Tree Algorithms of Classification in Data Mining", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391.

[12] Bahzas Taha Jijo; Adnan Mohsin Abdulazeez; 2021,Classification Based on Decision Tree Algorithm for Machine Learning, Journal of Applied Science and Technology Trends 2(01):20-28, DOI:10.38094/jastt20165

[13] Naim Monjezi, 2021,The Application Of the CART and CHAID Algorithms insugar Beet Yield Prediction ,Basrah J.Agric.Sci.,34(1):1-13

[14] Murat Gunduz ; Ibrahim AL-Ajji,2021, "Employment of CHAID and CRT decision tree algorithms to develop bid/no-bid decision-making models for contractors" , Engineering, Construction and Architectural Management, ISSN: 0969-9988, https://www.emerald.com/insight/0969-9988

[15] Etik Zukhronah; Yuliana Susanti; Hasih Pratiwi; Respatiwulan and Sri sulistijowati H. ,2021, "Decision tree technique for classifying cassava production", AIP Conference Proceedings : Vol. 2021, Iss. 1 ,10.1063/1.5062777, https://doi.org/10.1063/1.5062777.

[16] Andrew Gelman; Jennifer Hill; Aki Vehtar; 2022, "Regression and Other Stories" ,Andrew Gelman, Jennifer Hill, and Aki Vehtari, ch. 13,pp217-237

[17] Tomasz Rymarczyk; Edward Kozłowski; Grzegorz Kłosowski; Konrad Niderla, 2019: "Logistic Regression for Machine Learning in Process Tomography", Sensors *19*(15), 3400, https://doi.org/10.3390/s19153400.
[18] Andrés Villanueva Manjarres ; Luis Gabriel Moreno Sandoval ; Martha Salinas , "Data Mining techniques applied in educational environments: Literature review"; Digital Education ; Number 33; 2023
[19] Sixiang Jia ; Haochen Mou ; Yiteng Wu ; Wenting Lin ; Yajing Zeng ; Yiwen Chen ; Yayu Chen ; Qi Zhang ; Wei Wang ; Chao Feng and Shudong Xia,2022: "A Simple Logistic Regression Model for Predicting the Likelihood of Recurrence of Atrial Fibrillation Within 1 Year After Initial Radio-Frequency Catheter Ablation Therapy", Frontiers in Cardiovascular Medicine , https://doi.org/10.3389/fcvm.2021.819341

[20] Daniela Şchiopu; "Applying TwoStep Cluster Analysis for Identifying Bank Customers' Profile", Seria Ştiinte Economic, Vol. LXII, No. 3, 2010, pp66-75.