International Journal of Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2022.8.10.2

Volume 8, Issue 10
October - 2022

Development of a Model for the Prediction of Diabetes Using Data Mining Technology

EBELOGU Christopher U[1], OKONKWO Ngozi Ukamaka[2], GARKI Farida Shehu[3]

Research Scholar[1-2], Lecturer[3]

[1-2]Department of Computer Science, University of Abuja, Abuja-Nigeria.

[3]Department of Computer Science, National Open University of Nigeria, FCT-Nigeria

ABSTRACT

One of the most powerful technologies which are of high interest in the computer world is data mining. However, it is a comparatively new field of research whose major objective is to acquire knowledge from huge amounts of data. The motive of this study is to design a model which can forecast the likelihood of diabetes in patients with maximum accuracy. Therefore two main machine learning classification algorithms namely Decision Tree and Naive Bayes are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from the UCI machine learning repository. The performances of the two algorithms are evaluated on various measures like Precision, Accuracy, F-Measure, and Recall. Accuracy is measured over correctly and incorrectly classified instances. The filtering is done by comparing the group value to the class value followed by applying the Naïve Bayes classification algorithm for predicting diabetes. Significant attribute selection was done via the principal component analysis method. Results obtained show Naive Bayes outperforms with the highest accuracy of 76.30% compared to the decision tree which has 73.82%.

Key Words: Pima Indians Diabetes Database, Machine Learning, Decision tree, Artificial Neural Networks, Naive Bayes, Support Vector Machine.

1. INTRODUCTION

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis [1]. Data mining tools allow enterprises to predict future trends. The Healthcare industry is among the most information-intensive industries. Medical information, knowledge, and data keep growing daily. It has been estimated that an acute care hospital may generate five terabytes of data a year. The ability to use these data to extract useful information for quality healthcare is vital [2]. Clinical Prediction is a rapidly growing field that is concerned with applying Computer Science and Information Technology to medical and health data. With the aging population on the rise in developed countries and the increasing cost of healthcare, governments and large health organizations are becoming very interested in the potential of Clinical Diagnosis to save time, money, and human lives.

Although human decision-making is often ideal, it is poor when there are gigantic amounts of data to be classified. Also, proficiency and accuracy of decisions will decrease when humans are put under stress and immense work. Computer-assisted information retrieval may help support quality decision-making and to avoid human error. Imagine in most cases in Africa where doctors are in high demand, especially in rural areas. If a doctor is supposed to examine 4 patient records a day; he or she will go through them with ease. But if the number of records increases from 4 to 35 with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only four records to be examined.

With the improvement of technology more smart systems are being designed and developed with better data mining technologies to give the most precise results that could be associated with the disease. Healthcare organizations produce and collect large volumes of information daily. The tendency for a data mining application in healthcare today is great, because the healthcare sector is rich with information, and data mining is becoming a necessity. The use of information technologies allows the automatization of processes for the extraction of data that help to get interesting knowledge and regularities, which means the elimination of manual tasks and easier extraction of data directly from electronic records, transferring onto a secure electronic system of medical records which will save lives and reduce the cost of the healthcare services, as well and early discovery of contagious diseases with the advanced collection of data.

Data mining can enable healthcare organizations to predict trends in patient conditions and their behaviors, which is accomplished by data analysis from different perspectives and discovering connections and relations from seemingly unrelated information. Raw data from healthcare organizations are voluminous and heterogeneous. They need to be collected and stored in organized forms, and their integration enables forming of a hospital information system. Healthcare data mining provides countless possibilities for hidden pattern investigation from these data sets. These patterns can be used by physicians to determine diagnoses, prognoses, and treatments for patients in healthcare organizations.

2. AIM AND OBJECTIVES

This research aims to show that data mining can be applied to medical databases, which will predict or classify the data with reasonable accuracy to assist both the physician and the patient. The objectives of this study are to design a model which can forecast the likelihood of diabetes in patients with maximum accuracy and also present a Decision Tree and Naïve Bayes model for diabetes prediction in different patients.

3. SCOPE OF THE RESEARCH

This research is intended to come up with a system that will be fed with various symptoms and the disease/illness associated with them. It then processes the user’s symptoms to check for various illnesses that could be associated with it. Using Alba clinic, Kaduna-Nigeria as a case study, we use some intelligent data mining techniques to guess the most accurate illness that could be associated with a patient’s symptoms.

4. LIMITATION OF THE RESEARCH

The limitation of this study is that a structured dataset has been selected but in the future, unstructured data will also be considered, and these methods will be applied to other medical domains for prediction, such as for different types of cancer, psoriasis, and Parkinson’s disease.

5. LITERATURE REVIEW

Today, data mining has grown so vast that it can be used in many applications; examples include predicting costs of corporate expense claims in risk management, financial analysis, insurance, and process control in manufacturing, health care, and other fields [3] Data mining is a technology that requires a class of database applications that looks for hidden patterns in a group of data that can be used to predict future requirements. For example, data mining software can help retail companies find customers with common interests [4]. The word data mining is generally misused to describe software that presents data in new ways. Data mining software not only changes the presentation but also actually discovers previously unknown relationships among the data.

Due to the availability of huge amounts of and the need to convert that data into useful knowledge Data mining techniques can be useful. In recent years, Data mining has found significance in almost every field including health care. The abundance of data and the need for powerful analysis tools for that data are described as “data-rich but information-poor” situations.

5.1 Data Mining and Knowledge Discovery (KDD)

Knowledge Discovery (KDD) is a process that allows the automatic scanning of high-volume data to find useful patterns that can be considered knowledge about the data**.** Once discovered knowledge is presented, evaluation methods can be improved, the data mining process can be further "refined", new data can be selected or subsequently processed, and new data sources can be integrated to get different results corresponding [5]. This is the process of converting low-level information into knowledge of high level. Therefore, KDD is a non-trivial extraction of implicit information, previously unknown, and potentially useful data in the database. Although data mining and KDD are often treated as equivalent, in essence, data mining is an important step in the KDD process. The knowledge discovery process involves the use of the database, along with any selection, pre-processing, sub-sampling, and transformation; application of data mining methods to enumerate the models; evaluation of the data mining product to identify subsets listed models representing knowledge. The data mining component knowledge discovery process refers to algorithmic means by which patterns are extracted and listed from the available data [6].

With the application of data mining tools in spreadsheets of the program that analyzes data to identify patterns and relations, user profiling and the development of business strategies can be started [7]. Most data mining software includes online analytical processing, traditional statistical methods, such as cluster analysis, discriminant analysis, and regression analysis, and non-

DOI: 10.31695/IJASRE.2022.8.10.2

traditional statistical analysis such as neural networks, decision trees, link analysis, and association analysis. This wide range of techniques is not surprising because data mining is derived from three different disciplines, database management, statistics, and computer science, including the use of artificial intelligence and machine learning [8].

Because of all this, the data mining process is inextricably linked to computers. With the help of special software, a big computer system analyzes data from different angles, finds a hypothesis, experiments with them, and learns from previous experiences. One should always bear in mind that the software is just a tool that is still required the presence of human experts to give the final decision in the fields in which data mining is being applied. But in the first stage of processing computer systems are indispensable for their speed and lack of prejudice. Unlike humans, which would let the obvious connection between the two data be missed because it is beyond their expectations, such an error cannot happen to a computer.

Also, a human can be a victim of the conditionality with previous experience, which can be both positive and negative, but in any case impossible to avoid. It can be argued that data mining represents finding out the legality of the information. The technology of data mining is closely associated with data warehousing and intertwined with the system for database management. Data mining involves the process of searching a large amount of previously unknown information, which is later used to make important business decisions. The key phrase here is "unknown data", which means that the information is cluttered with massive amounts of operational data that, when analyzed, provide relevant information to organizational decision-makers. Datasets are generally large, complex, heterogeneous, and hierarchical, and they vary in quality. Preprocessing and transformation of data are needed even before data mining and discovery can be applied. Sometimes data features are not optimal for data mining and analytical processing. The challenge here is to convert data into the appropriate form before learning and data mining can start [9].

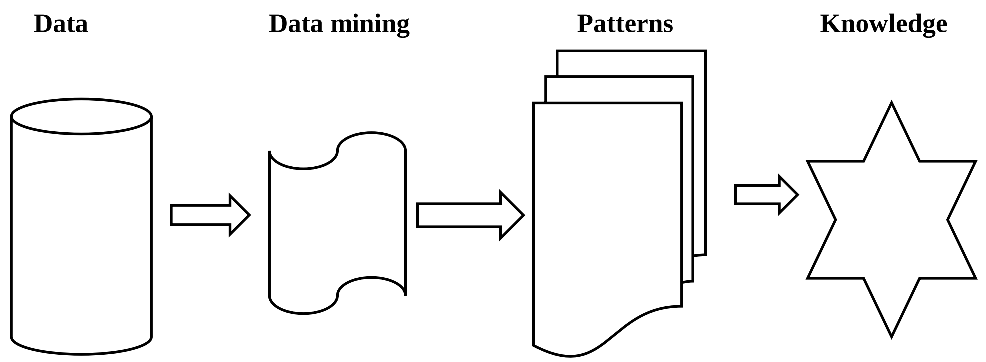

Figure 1: Process of applying data mining techniques to data

5.2 Integrating Data Mining Techniques into Telemedicine Systems

Today, in the healthcare industry huge volumes of data are collected and stored, data referring to patients, medical devices, hospital resources, diagnostics, treatments, etc. Data mining provides a set of techniques that can be applied to process and analyze the data to discover patterns that would enable healthcare personnel to make decisions based on the obtained knowledge. All healthcare institutions need expert analysis of their medical data, a project that is time-consuming and expensive.

There is great potential for data mining applications in healthcare. Healthcare institutions are very oriented toward the use of patient information. The ability to use data in databases to extract useful information for quality healthcare is a key to the success of healthcare institutions. Healthcare information systems contain large volumes of information that include information on patients, and data from laboratories that are constantly growing. With the use of data mining methods, useful patterns of information can be found in this data that will later be used for further research and report evaluation [7].

The key domains for which data mining can provide support in healthcare are:

- Diagnosis and treatment, by enabling doctors to identify diagnoses and provide effective treatment plans.
- Customer relationships, by making better decisions regarding customer services and offering patients improved services.
- Detecting fraud is used by insurers to identify fraud.

According to new data released by research firm InMedica, the American telehealth market is predicted to grow by 600 percent between 2012 and 2017. While there are currently 227,000 US telehealth patients, according to InMedica, that figure is forecast to reach up to 1.3 million patients in 2017. US telehealth revenues, meanwhile, will jump from $174.5 million last year to $707.9 million in 2017.

DOI: 10.31695/IJASRE.2022.8.10.2

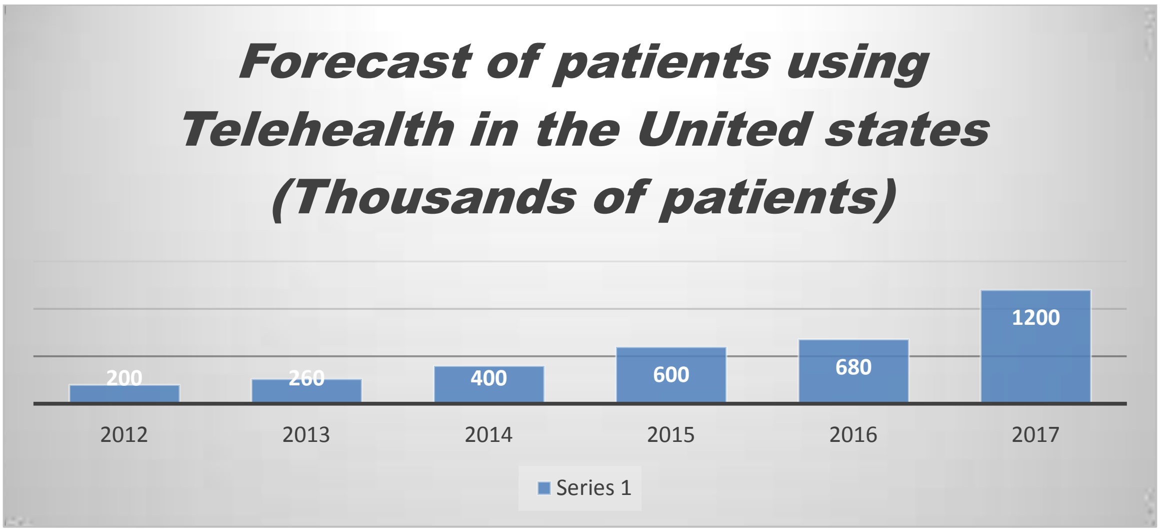

Figure 2: Source: HIS InMedica February 2013. Forecast of Patients using Telehealth in the United States (Thousands of Patients).

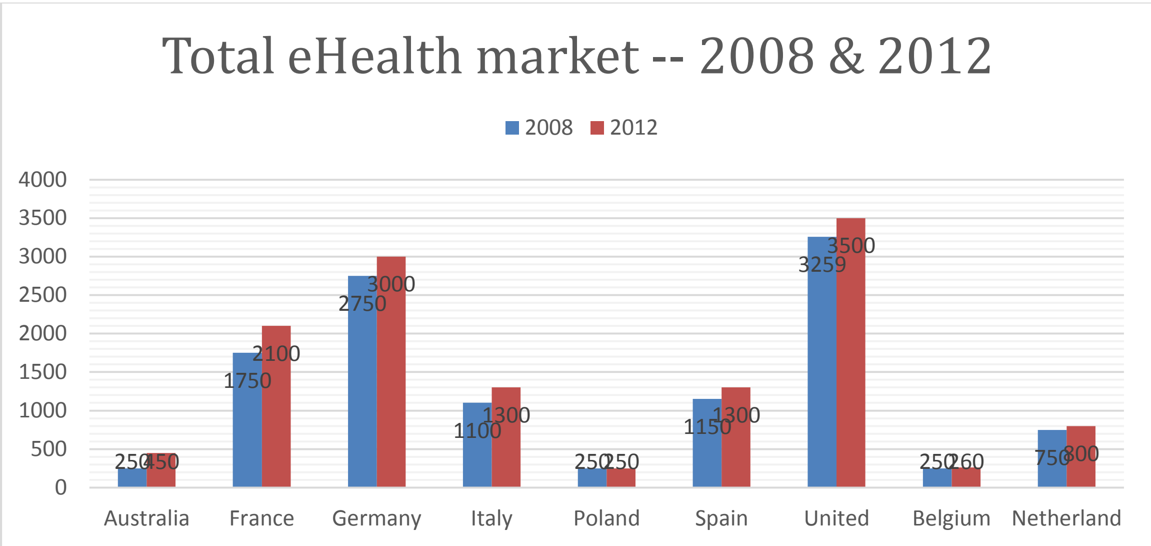

Figure 3: Telemedicine market growth in 2008 and 2012 period.

According to InMedica [10], the firm reported that United States leads telehealth growth worldwide, and has 75 percent of the world's telehealth patients. The next biggest countries in telehealth are the United Kingdom, China, and Germany. InMedica released a report last month that said there are about 308,000 telehealth patients worldwide, with an expected 1.8 million patients expected globally by the year 2017.

The urgency of curbing healthcare costs partly explains why the US is on the leading edge of the trend, InMedica suggested. "The cost of healthcare is a critical issue in the United States, with nearly one of every five dollars' worth of the country's gross domestic product (GDP) going to medical expenditures," Theo Ahadome, Senior Analyst at InMedica, said in a statement: "Telehealth can help mitigate these costs by reducing the number of patient readmissions and cutting down on in-home care visits. Because of this, the United States is the world's largest market for telehealth, driving the growth of the worldwide business."

6. SYSTEM ANALYSIS AND DESIGN

Classification strategies are broadly used in the medical field for classifying data into different classes according to some constrain comparatively an individual classifier. Diabetes is an illness that affects the ability of the body in producing the hormone insulin, which in turn makes the metabolism of carbohydrate abnormal and raise the levels of glucose.

This section captured the methodology for this research work which includes the collection of data, conversion of data, filtering of the data to remove any form of inconsistency, data training, and algorithm of the classifier model.

DOI: 10.31695/IJASRE.2022.8.10.2

6.1 Methodology Used

Model Diagram:

The proposed procedure is summarized in figure 4 in the form of a model diagram. The figure shows the flow of the research conducted in constructing the model.

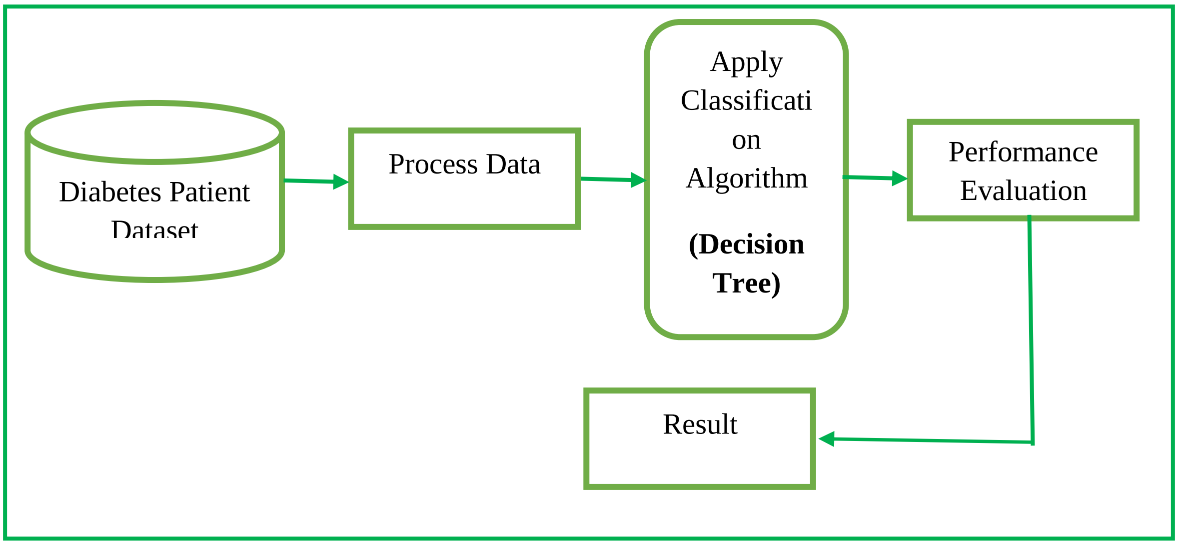

Figure 4: Proposed Model Diagram

6.2. Data Collection (Method)/Dataset Used

The data collection method used was by examining existing data in the form of databases. The data used for this research work is the Pima Indians Diabetes Dataset (PIDD) which is taken from the University of Californian Irvine (UCI) repository, Machine Learning Repository. The UCI machine learning repository is a collection of databases, domain theories, and data generators used for empirical analysis of machine learning algorithms by the machine learning community the UCI is used because the repository contains more than 350 datasets with labels like domain, the purpose of the problem and the dataset are diverse. This has over time been widely used and cited by many researchers and students as a source of machine learning data sets.

The proposed methodology is evaluated on Diabetes Dataset namely (PIDD). This dataset comprises medical detail of 100 instances which are female patients. The dataset also comprises numeric-valued 11 attributes where the value of one class '0' is treated as tested negative for diabetes and the value of another class '1' is treated as tested positive for diabetes.

Dataset description is defined by Table 3.1 and Table 3.2 represents Attributes descriptions.

Table 1: Data Set Description

S/N	Dataset	Number of Instances	Number of Attributes
1	Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney and Kidney Diseases	**768**	**9**

Table 2: Attribute Description

S/No	Attributes	Abbreviation of Attributes
1	Number of times pregnant	Preg
2	Plasma, glucose concentration	Plas
3	Diastolic blood pressure (mm Hg)	Pres
4	Skin fold thickness (mm)	Skin

DOI: 10.31695/IJASRE.2022.8.10.2

5	2-Hour serum insulin (mu u/ml	Insu
6	Blood Mass Index- BMI (weight in kg/(height in m)^m)	Mass
7	Diabetes pedigree function	Pedi
8	Age in years	Age
9	Class '0' 0r '1'	class

6.3 Data Training or Model Building

Learning or training algorithm in this case the Decision Tree classifier and Naive Bayes Classifier will be used to train the data after which a model is built; this model will be used to test and predict or cl accurately new instances. Often the major goal of separating data into a training set and test set is to help build a classifier with a minimum error rate.

6.4 Brief Description of Algorithm Used: (Decision Tree)

A Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using a Decision Tree in this research work is the prediction of the target class using decision rules taken from prior data. It uses nodes and internodes for the prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, the Decision tree chooses each node by evaluating the highest information gain among all the attributes.

Figure 5 shows a sample decision tree structure.

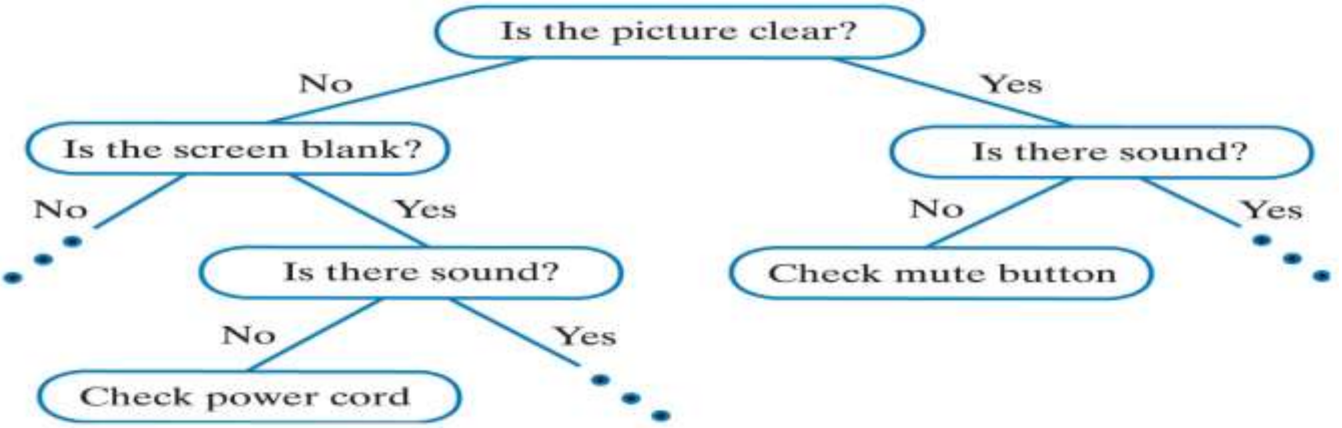

Figure 5: Sample Decision Tree Structure

A decision tree provides a powerful technique for the classification and prediction of Diabetes diagnosis problems. It creates a binary tree. The decision tree approach is most useful in the classification problem. With this technique, a tree is constructed to model the classification process. It consists of three types of nodes; root node, child node, and leaf node. The algorithm starts with defining a root node from the most relationship between every input and output variable. Next, the child node is selected by calculating Information Gain (IG). **IG (parent, child) = Entropy (parent) − [P(x1) × Entropyx1 + Px1 × Entropyx1 +…3 Entropy (Ci) = − P(xi) logP (xi) and P(xi)** is the probability of child node i. The node having the highest IG will become the parent for next the generation. This process is repeated until it gets a leaf node and a completed decision tree. The stopping criteria for the decision tree are that all the samples for a given node belong to the same class, there aren't remaining attributes for any further partitioning and there aren't any leftover samples. It requires little data preparation. While different techniques typically require data normalization, the creation of dummy variables, and, blank values to be removed.

DOI: 10.31695/IJASRE.2022.8.10.2

Table 3: Training Set

preg	Plas	pres	skin	insu	mass	pedi	age	class	0/1
6	148	72	35	0	33.6	0.627	50	Tested positive	
1	85	66	29	0	26.6	0.351	31	tested negative	
5	116	74	0	0	25.6	0.201	30	tested negative	
2	197	70	45	543	30.5	0.158	53	Tested positive	
10	139	80	0	0	27.1	1.441	57	tested negative	
1	189	60	23	846	30.1	0.398	59	Tested positive	
5	166	72	19	175	25.8	0.587	51	Tested positive	

7. EXPERIMENT AND RESULTS

7.1. J48 Decision Tree

Decision tree J48 implements Quinlan's C4.5 algorithm for generating pruned tree. The tree generated by J48 can be used for the classification of whether a patient has tested positive or negative for diabetes. The data mining technique uses the concept of Information Gain (IG). Each attribute of the data is used to decide by splitting the data into smaller modules.

It examines normalized information gain (IG) (difference in entropy) that results from choosing an attribute as a split point. The highest normalized IG is used at the root of the tree. The procedure is repeated until the leaf node is created for the tree specifying the class attribute that is chosen. Figure 4.1 shows the J48 pruned tree that was generated by WEKA.

7.2 Classifier Output

Based on Cross-Validation Technique

The J48 algorithm gives the following correct results for the given dataset.

Table 4: Performance Results from J48 Decision Tree Classification Algorithm – Cross-Validation.

	A-tested positive	B-tested positive
A - tested positive	149 (i)	119 (ii)
B - tested negative	74 (iii)	426 (iv)

In table 4, the values represent the following:

i. Number of correct forecasts that the instance tested positive
ii. Number of incorrect forecasts that the instance tested negative
iii. Number of incorrect forecasts that the instance tested positive
iv. Number of correct forecasts that the instance tested negative

Based on Percentage Split (70:30) Technique

Since a 70:30 percentage split was applied to the dataset 230 of the instances were used as the test dataset while the rest were used for training the model. The decision tree algorithm gives the following correct results for the given dataset.

Table 5: Performance Results from J48 Classification Algorithm – Percentage Split

	No. of Instances	Percentage
Correctly Classified Instances	567	73.8281 %
Incorrectly Classified Instances	201	26.1719 %

DOI: 10.31695/IJASRE.2022.8.10.2

7.3 Confusion Matrix for Decision Tree Classifier

A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions is summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows how your classification model is confused when it makes predictions [11]. It gives your insight not only into the errors being made by your classifier but more importantly the types of errors that are being made. It is this breakdown that overcomes the limitation of using classification accuracy alone.

Table 6: Confusion matrix of Decision Tree

S/N	A (0)	B (1)
A - Tested Negative (0)	407	93
B - Tested positive (1)	108	160

7.3.1 True Positive (TN)

We predicted positive and it's true. We correctly predicted that 160 people have diabetes and they have (positive).

7.3.2 True Negative (TN)

We predicted negative and it's true. We correctly predicted that 407 people don't have diabetes and they don't have (negative).

7.3.3 False Positive (FP)

We predicted positive and it's false. We predicted that 93 have diabetes but they don't have (negative). It is also known as a Type 1 error.

7.3.4 False Negative (FN)

We predicted negative but it's false. We predicted that 108 people don't have diabetes but they have (positive). It is also known as a type 2 error.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         567               73.8281 %
Incorrectly Classified Instances       201               26.1719 %
Kappa statistic                          0.4164
Mean absolute error                      0.3158
Root mean squared error                  0.4463
Relative absolute error                 69.4841 %
Root relative squared error             93.6293 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.814    0.403    0.790      0.814    0.802      0.417    0.751     0.811     tested_negative
                 0.597    0.186    0.632      0.597    0.614      0.417    0.751     0.572     tested_positive
Weighted Avg.    0.738    0.327    0.735      0.738    0.736      0.417    0.751     0.727

=== Confusion Matrix ===

   a   b   <-- classified as
 407  93 |   a = tested_negative
 108 160 |   b = tested_positive
```

Figure 6: Results from Weka on the Diabetes Dataset using decision tree (J48) classifier

DOI: 10.31695/IJASRE.2022.8.10.2

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         586               76.3021 %
Incorrectly Classified Instances       182               23.6979 %
Kappa statistic                          0.4664
Mean absolute error                      0.2841
Root mean squared error                  0.4168
Relative absolute error                 62.5028 %
Root relative squared error             87.4349 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.844    0.388    0.802      0.844    0.823      0.468    0.819     0.892     tested_negative
                 0.612    0.156    0.678      0.612    0.643      0.468    0.819     0.671     tested_positive
Weighted Avg.    0.763    0.307    0.759      0.763    0.760      0.468    0.819     0.815

=== Confusion Matrix ===

   a   b   <-- classified as
 422  78 |   a = tested_negative
 104 164 |   b = tested_positive
```

Figure 7: Results from Weka on the Diabetes Dataset using Naïve Bayes classifier

7.4 Results

Table 7: Classifier's Performance based on Classified Instance

Total number of Instances	Classification Algorithm Used	Correctly classified Instances	Incorrectly classified Instances
768	Decision Tree (J48)	567	201
	Naïve Bayes	586	182

Table 8: Comparative Performance of Classification Algorithms

Classification Algorith	Precision	Recall	F- Measure	Accuracy %	ROC
Decision Tree	0.735	0.738	0.736	73.82	0.751
Naïve Bayes	0.759	0.763	0.76	76.3	0.819

Accuracy measures

- **Precision:** Classifiers' correctness/accuracy is measured by precision**.**
- **Recall:** To measure the classifiers' completeness or sensitivity**.**
- **F-Measure:** the weighted average of precision and recall.
- **Accuracy %:** Determines the accuracy of the algorithm in predicting instances.
- **ROC:** (Receiver operating curve) is used to compare the usefulness of the test.

Table 8 shows the performance values of all classification algorithms calculated, it is examined that Naïve Bayes shows the maximum accuracy. So the Naive Bayes machine learning classifier can predict the chances of diabetes with more accuracy compared to the decision tree classifier.

8. SUMMARY

Data mining is important to medicine, and it represents a comprehensive process that demands a thorough understanding of the needs of healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve the success of healthcare organizations and the health of the patients. The results in this work show

DOI: 10.31695/IJASRE.2022.8.10.2

that among two of the best classification algorithm used in data mining Naïve Bayes has proven and shown more accuracy than the Decision Tree.

9. CONCLUSION

In this research, an attempt has been made to use the Decision Tree algorithm and Naïve Bayes algorithm for diabetes prediction. However, data mining technology provides an oriented approach towards new and hidden patterns in data, from which the knowledge is being generated, the knowledge that can help in providing medical and other services to both physicians and patients. Healthcare institutions that use data mining applications can predict future requests, needs, desires, and conditions of the patients and make adequate and optimal decisions about their treatments.

One of the important real-world medical problems is the detection of early symptoms of diseases at an early stage. In this study, systematic efforts are made in using two major classification algorithms in data mining for the prediction of diabetes. During this work, two machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on Pima Indians Diabetes Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 76.30 % using the Naive Bayes classification algorithm. With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in medical data.

10 RECOMMENDATION

Based on the examination and conclusions presented in this work, it is suggested that the designed healthcare prediction system to be used in the future should have more work done by using more data sets related to diabetes.

REFERENCES

1. Rouse, P. and Putterill, M. (2003), "An integral framework for performance measurement", Management Decision, Vol. 41 No. 8, pp. 791-805.
2. Arun G. E, (2004) “Application of Data mining in Medical Applications”, A thesis presented to the University of Waterloo in fulfillment of the thesis requirement for the degree of Master of Applied Science in Systems Design Engineering, (2004).
3. Mahboob K. “Smart Health Prediction Using Data Mining”, [Online]. Available:file:///C:/Users/hpuser/Downloads/smarthealthpredictionusingdatamining-160430174838.pdf
4. Nawani, Aditya & Gupta, Himanshu & Thakur, Dr. Narina. (2013). Prediction of Market Capital for Trading Firms through Data Mining Techniques. International Journal of Computer Applications. 70. 7-12. 10.5120/12165-8005.
5. Zaiane. O. R, Hagen. E, and Han. J, " Word Taxonomy for On-Line Visual Asset Management and Mining ", Proc. Int. Workshop on Application of Natural Language to Information Systems (NLDB'99), Klagenfurt, Austria, June 1999 (PDF).
6. Fayyad. U, Piatetsky-Shapiro. G and Smyth. P, "From Data Mining to Knowledge Discovery: An Overview," in Advances in Knowledge Discovery and Data Mining, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, eds., MIT Press, Cambridge, Mass., 1996, pp. 1-36.
7. Milovic, B. and Milovic, M. (2012) Prediction and Decision Making in Health Care Using Data Mining. International Journal of Public Health Science, 1, 69-78.
8. Koh, Eric. (2015). Koh, E.H.Y., Avvari, M.V. and Tan, K.H. (2015), "An integrated framework for competency development: perspectives of risk managers in banks", Service Business, http://link.springer.com/article/10.1007/s11628-015-0283-0. Service Business. 10. 10.1007/s11628-015-0283-0.
9. Hosseinkhah, F., Ashktorab, H., Veen, R., & Owrang O., M. M. (2009). Challenges in Data Mining on Medical Databases. In J. Erickson (Ed.), Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1393-1404). IGI Global. https://doi.org/10.4018/978-1-60566-058-5.ch083.
10. Kwankam, s. (2013). The way forward for telemedicine in health care delivery. World hospitals and health services: the official journal of the International Hospital Federation. 49. 18-21.
11. The machine learning in software project management: A journey. Part I - Scientific Figure on ResearchGate. Available from:https://www.researchgate.net/figure/ML-Classifiers-tree-as-per-the-Jason-Brownlees-classification_fig18_284716145

DOI: 10.31695/IJASRE.2022.8.10.2