Comparative study of Data Mining Approaches for Parkinson's Diseases

Dr.Hari Ganesh S[1], Gracy Annamary S[2]

[1]Asst. Professor, [2]M.Phil. Scholar,

Department of Computer Applications,

Bishop Heber College (Autonomous),

Trichirappalli-620 017

ABSTRACT

Data mining is the process of finding useful and relevant 3062information from the data base. The database store the various field of e-business, marketing, retail and medical. The data mining is used for medical and health areas of the most important factors in industrial societies. This paper analyzes the Parkinson's disease accuracy level of different classification algorithms. In this work taken for four different types of classifier names by J48, Naive bayes, Decision table and Random tree. The data set a used from the UCI repository. WEKA tool a used.

KEYWORDS: Parkinson's disease, Classification, Decision Table, Naive Bayesian, J48, Random Tree.

1. INTRODUCTION

Data mining is the technology provides user oriented approach to novel and hidden pattern in the data. The Parkinson's disease (PD) is a type of neurological disease. Parkinson's disease is a degenerative disease of the brain, which affects the nerve cells in brain called by Neurons. The neurons produce the dopamine, they control to brain of movements. Normally, this disease progresses slowly. Some people reside for many years with only minor symptoms. The people start to have symptoms between the ages of 50 and 60. But sometimes symptoms start previous. This paper aims at establish the best accuracy classifier algorithms by naive bayes, decision table, j48 and random tree evaluate Parkinson's data set.

1.1 CLASSIFICATION

Classification is perhaps the most familiar and most popular data mining techniques. There are different types of classification algorithms for Parkinson's disease. They are using the follows:

1. Naive bayes classifier
2. J48 algorithm
3. Decision table

4. Random tree

1.2 PARKINSON'S DISEASE

The Parkinson's disease is a neurodegenerative disease. The disease affect by brain cells (neurons) in human brain. They affected neurons by the brain cells by substantia nigra. The neurons put together an important chemical called dopamine. The dopamine sends singles to the part of the brain that controls movements. The little signals can help those parts of the brain work better. The decrease of dopamine of the brain the person unmorally movement.

They are four types of symptoms of Parkinson's disease are tremor, rigidity, bradykinesia and postural instability.

- Tremor: vibrate by hands, arms, legs or jaws.
- Rigidity: limbs and trunk inflexible.
- Bradykinesia: slow movements.
- Postural instability: depression and emotional changes.

The earliest indicators affects by 75-90% people with Parkinson's disease. This paper will analyze the different kinds of Parkinson's disease Techniques.

1.3 NAIVE BAYES CLASSIFIER

The Naive bayes classifier greatly simplifies learning by assuming those features are independent value of the attributes given class. The naive bayes classifier technique is base on the so-called Bayesian theorem. The dimensions of the input are high. Despite its simplicity, naive bayes can after outperform more sophisticated classification methods. An advantage of the naive bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables an assumed, only the variances of the variables for each class need for determined and not the entire covariance matrix.

1.4 J48 ALGORITHM

J48 builds decision trees from a set of labelled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to make a decision by splitting the data into smaller subsets. J48 examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created to the decision tree telling to choose that class. But it can also happen that none of the features give any information gain. In this case j48 creates a decision node higher up in the tree using expected value of the class. j48 can handle both continuous and discrete attributes; training data with missing attribute values and attributes with differing costs.

1.5. DECISION TABLE

The simplest, most rudimentary way of representing the output from machine learning is to make it just the same as the input a decision table. They creating a decision table might involve selecting some of the attributes. A decision table consists of a hierarchical table in which each entry in a higher level table gets broken down by the values of a pair of additional attributes to form another table. The structure is similar to dimensional stacking. Presented here is a visualization method that allows a model based on many attributes to be understood even by those unfamiliar with machine learning.

There are four main advantages of decision table.

1. A decision table provides a framework for a complete and accurate statement of processing or decision logic.
2. A decision table may be easier to construct than a flowchart. .
3. Complex tables can easily be split into simpler tables.
4. Table users are not required to possess computer knowledge.

1.6. RANDOM TREE

Random trees are a collection (ensemble) of tree predictors that are called forest further in this section. The classification works as follows: the random trees classifier takes the input feature vector, classifies it with every tree in the forest, and outputs the class label that received the majority of “votes”. In case of a regression, the classifier response is the average of the responses over all the trees in the forest. The entire tree is trained with the same parameters but on different training sets. These sets are generating from the original training set using the bootstrap procedure: for each training set, you randomly select the same number of vectors as in the original set. The vectors are chosen with replacement.

2. DATA DESCRIPTION

The data are collecting by UCI repository. The objective of this data set is to analyze the Parkinson’s disease based on the given attributes. The data set consists of 23 attributes are using to predict the Parkinson’s disease. The detail descriptions of the attributes are given as below

Table 1: The various attributes for Parkinson’s disease

Feature No	Feature Name	Description
1	MDVP: Fo(Hz)	Average vocal fundamental frequency
2	MDVP: Fhi(Hz)	Maximum vocal fundamental frequency
3	MDVP: Flo(Hz)	Minimum vocal fundamental frequency
4	MDVP: Jitter(%)	Kay pentax MDVP jitter as percentage
5	MDVP: Jitter(Abs)	Kay pentax MDVP absolute jitter in microseconds
6	MDVP: RAP	Kay pentax MDVP Relative Amplitude Perturbation
7	MDVP: PPQ	Kay pentax MDVP five-point period perturbationQuotient
8	Jitter:DDP	Average absolute difference of differences between cycles, divided by the average period.
9	MDVP: shimmer	Key pentax MDVP local shimmer
10	MDVP: shimmer(Db)	Key pentax MDVP local shimmer in decibels.
11	Shimmer: APQ3	3 Point Amplitude perturbation Quotient
12	Shimmer: APQ5	5 Point Amplitude perturbation Quotient
13	MDVP: APQ	Kay pentax MDVP eleven-point Amplitude perturbation Quotient
14	Shimmer: DDA	Average absolute difference between consecutive differences between the amplitude of consecutive periods.
15	NHR	Noise to Harmonic Ratio
16	HNR	Harmonics to Noise Ratio
17	RPDE	Recurrence Period Density Entropy
18	DFA	Detrended Fluctuation Analysis
19	Spread1	Non Linear measure of fundamental frequency
20	Spread2	Non Linear measure of fundamental frequency
21	D2	Correlation Dimension
22	PPE	Pitch Period Dimension
23	Status	Health Status 1-parkinson; 0-Healthy

The attributes taken on data set. The data sets are based on the numeric and nominal data type.

3. EXPERIMENTAL RESULTS

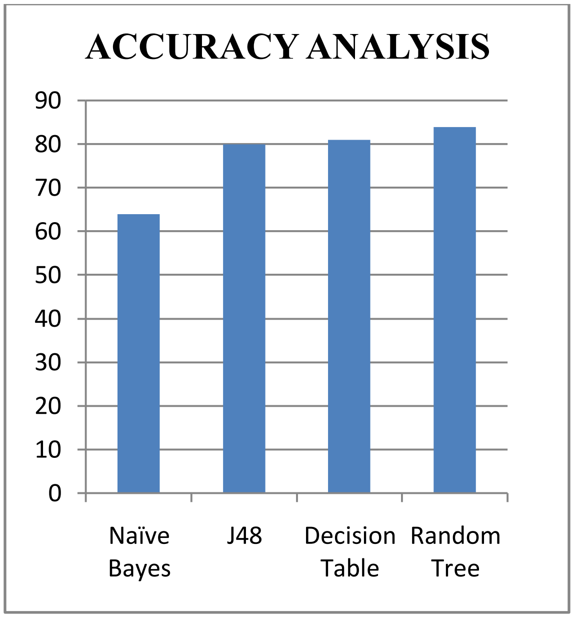
The given four types of algorithms like Naive bayes, Decision Table, J48 and Random tree an applied on the Parkinson's disease data set in WEKA. This performance of the algorithms are given based various factors. The factors can analyze on vibrate by hands, arms, legs or jaws and slow movements of depression and emotional changes.

3.1 Performance of the Algorithms based on the Accuracy level taken

The dataset consists of 195 instances and the applied as a test case in the classification algorithms. The performance can be obtained based on the accuracy taken to build the random tree correctly classified instances. The instances which are correctly classified using the WEKA tool.

Table 2 comparison table

S.No	Classification techniques	Accuracy (%)
1	Naive bayes classifier	69
2	J48	80
3	Decision table	81
4	Random tree	84



X-Axis: Classification of Algorithms

Y-Axis: percentage level

Fig 1 depicts the accuracy of various classification algorithms

4. DISCUSSION:

The above four algorithms, the final output will be patterns which used to find out whether the person is affected by Parkinson's disease or not. A Confusion Matrix is a useful visualization tool for analyzing the classifier accuracy. Structure of the confusion matrix can be given as below

Table 3 Structure of the Confusion Matrix

TP	**TN**
FP	**FN**

Where

- **TP** is True Positive: Parkinson's disease patients correctly identified as Disease.
- **FP** is False Positive: Healthy people incorrectly identified as Parkinson's disease.
- **TN** is True Negative: Healthy people correctly identified as healthy.
- **FN** is False Negative: Parkinson's disease patients incorrectly identified as healthy.

The Confusion Matrix for the classification algorithms such as Naive bayes, J48, Decision Table and Random tree canister given as follows based on execution of the algorithms using WEKA tool.

Table 3.1 Confusion Matrix for Naive Bayes

44	**4**
56	**91**

Table 3.2 Confusion Matrix for J48

28	**20**
18	**129**

Table 3.3 Confusion Matrix for Decision Table

28	**20**
17	**130**

Table 3.4 Confusion Matrix for Random Tree

33	**15**
15	**132**

5. CONCLUSION

Data mining plays a Parkinson's disease. They are use in different classification algorithms used in Parkinson's disease. Here we compared the classification algorithms to analyze which algorithm gives better accuracy. These models are built based as a test case on the UCI repository dataset. The experiment has been successfully performed with several data mining classification techniques.

The paper is verifying the various classifiers to the Parkinson's data set. The dataset comprises of 23 attributes with various range of values. Finally the paper, Random tree yields 84% accuracy.

REFERENCES

[1] Dr. R.Geetha Ramani, G.Sivagami, Shomona Gracia jacob " Feature Relevance Analysis and Classification of Parkinson's Disease Tele-Monitoring data Through Data Mining" , International Journal of Advanced Research in Computer Science and Software Engineering,vol-2,Issue 3, March 2012.

[2] Peyman Mohammadi, Abdolreza Hatamlou and Mohammed Msdaris "A Comparative Study on Remote Tracking of Parkinson's Disease Progression Using Data Mining Methods" , International Journal in Foundations of Computer Science and Technology(IJFCST),vol-3,No.6, Nov 2013.

[3] Dr. R.Geetha Ramani and G.Sivagami "Parkinson Disease Classification using Data Mining Algorithms", International Journal of Computer Applications (IJCA),Vol-32,No.9, October 2011

[4] Chandrashekharn Azad.,"Design and Analysis of Data Mining based prediction model for Parkinson's Disease", International Journal of Computer Science Engineering (IJCSE) ISSN: 2319-7323, Vol.3, No.03, May 2014

[5] Tawseef Ayoub Shaikh.,"A Prototype of Parkinson's and primary tumor disease prediction using data mining techniques", International Journal of Engineering Science Invention, vol 3,Issue:4, April 2014

[6] Jiawei Han, Micheline Kamber, Jian pei.," Data mining concepts and techniques", 3rd edition, page no:350-354

[7] Ianhwitten and Eibe frank.,"Data Mining Practical Machine Learning Tools and Techniques",2nd edition, page no: 62,365-368.

[8] Margaret H.Dunham.,"Data Mining Introductory and Advanced Topics".