



# JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

## Comparative Survey of Random Forest and Decision Tree in Predictive Analytics

Dr. I. Carol

Assistant Professor, Department of Information Technology, St. Joseph's College, Trichy, TamilNadu, India

### Abstract:-

In today's world, humans are capable of producing a large amount of raw data which is furthermore considered as Big Data. Big Data requires high-performance hardware and software to process and to produce fruitful information which will be useful to create complex results. Big Data is productively used in the field of business analytics to create rightful insight. It is carried on with many programming languages and tools as well as various Big Data analytical techniques. Big Data analytics is classified into various analytics including predictive analytics which produces prediction based insights. Predictive analytics had been playing a vital role in every aspect of data processing. There are many algorithms such as Clustering, K- means, Classification, Random forest and decision tree which are capable of producing high precision of information. Additionally, this paper proposes a survey on predictive analytics and comparative literature between the Random Forest algorithm and the Decision Tree algorithm.

**Keywords:** *Big Data, Information, Business Analytics, Predictive Analytics, Classification, Random forest, and Decision Tree algorithm.*

### Introduction:-

Big data can be used to produce complex systems in order to construct a network of simulations. It is also used to create visualizations and constructive prediction based data. Furthermore, Big Data is used to predict results based on various queries like location, time, and customer review and so on. There are many types and varieties in Big Data processing based on Five V's such as

- Variety
- Velocity
- Volume
- Veracity
- Value

Big Data deals with a variety of data which is classified into classes and objects for furthermore predicted use. It deals with Velocity in which how to speed a result of prediction is obtained. Big Data deals with a huge volume of data and veracity of data. The value of data is considered where the most important thing to consider in Big Data is this *Value*.

In Big Data analytics, there are various part and types of analytics such as

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

These types of Big Data by each have its own mechanism to process the huge amount of data in order to produce useful insights.

In which this paper concentrates on Predictive analytics which is been an outbreak through various industries that try to predict insights to improve their business process. Predictive analytics is used in moreover many and all sectors of businesses. Predictive analytics is known for producing and providing way more useful insights along with the precision of predicted data regarding any sort of processed queries.

The predictive algorithm consists of many algorithms in order to produce valuable data with a high degree of precision.

- Clustering algorithm
- Classification algorithm
- K- means algorithm
- Random Forest algorithm
- Decision Tree Algorithm

Classification is used in supervised learning of data processing and also it used when we already have some predefined insights over data. In classification, data is been grouped together in a format that data in the same group will be under a certain scientific or some taxonomic name.

Random forest algorithm and decision tree algorithm is specifically taken into a comparative survey and discussed in detail as in this paper.

### Review of Literature:-

Amy Shi-Nash et al. explains the brief details of data analytics and predictive analytics. The author describes the brief history of KDD, data mining, machine learning, and artificial intelligence and so on. The author states the earlier method of data analysis is namely association rule learning which is also acknowledged as Basket Analysis. The types of data analytics were discussed. Here the author defined the concept of CRISP-DM, CRISP-DM abbreviated from Cross-Industry Standard Process for Data Mining which was used by the data mining projects. The execution of this method was done by the six stages. The characteristics of big data, the overview of data analytics tools have been discussed. In this research, the Current Status of the big data analytics, data science to the corporate level, and Acquiring talents and skills for the data scientist, Ecosystem and security of big data were discussed. [1]

This paper begins with a short review of the literature regarding what is predictive analytics and its types, the models and algorithms, an application used in different industries, management tasks used in predictive analytics. The author states that predictive analytics comes from business intelligence. The algorithms explained such as classification, clustering, association rules, and regression. The predictive modeling is defined as the kin of algorithms as a pictogram. The applications of predictive analytics in different industries are explained broadly, lots of management tasks in the predictive algorithm were explained deeply in the literature review. The tabular column depicts the usage of predictive analytics, where they used, which tools and algorithms have been used were clearly described. The drawbacks of Enterprise Resource Planning system, Service Level Agreements, Campaign Management Support Systems, and such many marketing system applications have been mended by using predictive analytics applications were discussed. [2]

Several studies have revealed the types, usage, and application of the decision tree. In this paper, the author Mr. Brijain R Patel et al. proposed the survey on the decision tree algorithm. The decision tree algorithm mainly focused on the classification model. The decision tree has commonly used in many diverse areas, industries for the classification. The author explained the brief definition of the decision tree. The types of decision tree algorithms namely, Chi-square Automatic Interaction Detector (CHAID), Classification and regression tree (CART), ID3 (Iterative Dichotomiser 3), C4.5, C5.0, Hunt's algorithm were discussed. Business, Image processing, Medicine, Energy modeling, E-commerce, Intrusion detection, Intelligent vehicles, Remote sensing, and web application are the most applications used in the decision tree. The comparison of decision tree algorithms has been explained by the neat tabular column. The Weka, GATree, Alice d'ISoft are such fewer tools to invoke the decision tree learning. Moreover, this paper also explicit the issues related to the decision tree algorithm was discussed. [3]

Mr. D. V. Patil et al. focused on the central theme of handling and predicting the targeted outcome value from the high volume of data. So here the author used the meld of genetic algorithms with a decision tree, which is also termed as the hybrid methodology. The author stated that the decision tree and genetic algorithm which give the classification accuracy and the best and fit solution to such particular problems respectively. Here the proposed system using global search techniques to get the optimum result and which is also used to reduce the congestion of memory. The k-fold cross-validation method has been used to attain the data set. To summarise this research which is used to get the best classification accuracy and lessening the size of the tree. [4]

In above research paper, the author proposed the decision tree algorithm along with the genetic algorithms. Data from several studies have identified the drawbacks of using a decision tree integrated with the k-means algorithm, Monte Carlo approach, Fuzzy learning, Fuzzy ID3, fuzzy C-means algorithm, C-fuzzy decision tree was discussed. To overcome those drawbacks, this approach has occurred. The new decision tree was introduced in this probe based on the genetic algorithm. The first the author designed the genetic algorithm for the new decision tree with such steps. The first step was initialization and the second one was an iterative generation which each has been divided into three phases. Reproduction, Crossover, and Mutation phase are those phases. The proposed system compared the performance of two concepts. The concepts namely as decision tree based on both the k-means algorithm and genetic algorithm. Finally, the author stated that the comparison of the two results reveals that decision tree based on a genetic algorithm that gives the optimal solution from the probe and which is swamped the drawback based on the k-means algorithm. [5]

Li Wenlong et al. proposed the parallel decision tree classification algorithm based on the combination which is different from the tree algorithms of ID3 (Iterative Dichotomiser 3), C4.5, CART and it is mainly focused on processing the enormous dataset with limited resources. This method provides the highest classification accuracy in the experiment. Here three strategies have been used in the parallel decision tree namely as dynamic data fragment, static data fragment which consists of vertical division and

horizontal division, and combination decision tree. Thus the author demonstrated that the high accuracy has occurred in the parallel decision tree is better than others. [6]

G. Niveditha et al. intended the paper about credit card fault detection. Here the author used the random forest algorithm which is the finest algorithm for classification and it has used the more decision tree to give a most accurate predictive solution. The purpose of this paper to detect the process of a transaction using a credit card whether fraud or non-fraud. True positive, False positive, True negative, False negative, Accuracy, Sensitivity, Specificity, and precision were the context of the current problem. The random forest is a classification algorithm while using a classification algorithm there are such steps to follow namely as, training, testing and usage phase. In this research the modules were divided into four categories, there are Pre-processing, Feature extraction, Trained data, and Secure transaction. There are a lot of algorithms has available to detect this particular problem, but the author depicts that the random forest was the good one to get the optimal solution. [7]

The study of find out the classification result on the vessel from space-based AIS data was carried out by Hanyang Zhong et al. In this research, the writer disclosures the method of vessel classification using the Random Forest method. Here it used the three types of vessels with dissimilar parameter strings. The author stated what is AIS data, how it can be important it is and AIS which gives the static and dynamic information about the vessels. The three types of vessels are Cargo, Tanker, Fishing has used for the classification. The Information entropy approach has been used to measure the feature. The *RandomForestClassifier* was the function modular used to construct the test. The number of decision trees and the maximum depth of each tree has considered as the two subjective factors. The best accuracy has been obtained by using a random forest algorithm for three sorts of vessels along with six geometricity features. [8]

Naili Suri Intizhami et al. proposed the paper to predict the battleship winner at the combat. Here the Random forest algorithm has been used to get the best result. Because it is one of the crucial ones in warfare simulation because it will one of the contributions of every commander. Each commander used to acquire the best battleship around the battle environment. The proposed system used the computer network system and the random forest algorithm as a methodology. The server, commander computer, spy computer, and hub has used in the computer network scheme. The accuracy of using random forest is better than the support vector machine and K-Nearest Neighbors. [9]

S. Karthick et al. proposed the concept of weather prediction using the random forest algorithm. To predict the weather is the difficulty in the department of meteorological. In this article, the related work discussed the existing algorithms used to predict the weather. This proposed system using the approach by following such steps namely, data collection and retrieval, data transformation, mining tool, data pre-processing, feature extraction, data mining. Each step follows specific functions to occur in the prediction. Here Weka tool is used as the mining tool. Random forest and C4.5 were used in the data mining step to analyzing the data set. Findings in the present study are consistent with the findings of which algorithm give the best accuracy. This concluded with the Random forest algorithm given the high accurate prediction than the C4.5 decision tree algorithm. The comparison between the random forest algorithm and decision tree algorithm to predict the weather was demonstrated here. [10]

## Comparative Survey

Whatever happened, the production of data doesn't stop at any time further. Data production had meeting its peak every day. In order to maintain and control the data production, every raw data needs to be processed to produce a connection of useful information which will be further used in many of today's business insights. Therefore, there is a separate field of study known as Business Intelligence which involves both Big Data concepts business managerial aspects. Big Data is a study that boomed in the last three decades which is now in its peak of giving birth to too many new and sub-technologies. Humans discovered that nothing such as innovations or inventions are possible in any further future. That is, it is very much needed in today's world that models and algorithms of data processing to evaluate and summarize a large amount of data so-called Big Data. Most important Big Data concepts are based upon the following factors of data processing,

The data production and data processing gave way into the prediction of data which is thereby known as Data analytics. Data analytics is a process by which we take up data for a predicted or the desired insight. When a data is passed through this process of data analytics, it undergoes various processes which are listed as follows

- *Data abstraction* – The process of collecting relevant data.
- *Data cleansing* – Data which are collected may contain many replications, hence it needs to be cleansed
- *Data analysis* – The process of analyzing the data regarding the obtained user queries.
- *Building model* – The process of building a unique model of analytics to obtain specific and desired results
- *Validating model* – The process of validating models using various Big Data techniques.
- *Deploying model* – The process of deploying the finalized model and winding up the process with successful insights.

Data analytics is therefore considered as one of the important processes among other processes of Big Data. Data analytics consists of many specifically varied analytics which can be used according to the purpose of the user. Therefore the classification of Data analytics are

- *Descriptive Analytics* – It is a type of analytics by which a complex data set is broken down into simple attributes for simpler processing.
- *Diagnostic Analytics* – The process of analytics involving algorithms specifically used to analyze a data set.

- *Prescriptive Analytics* – Analytics which is used to get data from the user and define a problem with a solution according to the mental schematic attributes of a user.
- *Predictive Analytics* – Predictive analytics is a process of a dataset by which it can predict the future.

Thus Predictive analytics is used in every sector of any existing business that every human being is expecting a predictive report of upcoming business decisions.

- *Classification*- It is part of the predictive analytical process under which this algorithm divides the data set into classes with relation. It is massively used under the supervised learning processes.
- *Regression*- Regression is another method of statistical analysis which is used in everyday to obtain predictive analytical result.

## Decision Tree

Decision Tree is the Supervised Machine Learning algorithm. It is mainly used for both classification and regression. The decision tree follows the tree-based structure to predict the progress of action by using past historical data. It is mostly focused to predict the non-linear data set which classifies the data sets into the number of branches to make the decision that's why it is called eager learners. There are two important key terms are must find out in the decision tree namely Entropy and Information Gain.

Entropy is the process of finding the measurement of randomness or unpredictability in the entire data set.

Information Gain is the process of reduction in the value of entropy which is measured by after splitting the data set.

The decision tree has several steps to make the solution to such a particular problem.

### Root node

The crucial part of the decision tree is the root node because it is the top of the tree which mostly consists of the best attribute in the data set.

### Decision node

It is splitting from the root node and also from the sub-node with two or more branches.

### Splitting

Splitting is the process of divide the dataset into the subsets.

### Internal node

An internal node is termed as the outcome of the node.

### Leaf node

The leaf node is also known as the terminal node which is the end node of the decision tree.

- Types of Decision tree

There are two types of decision trees depends on their type of data.

#### 1. Classification tree

It is also known as a Categorical Variable decision tree which is done by deciding a set of logical conditions to categorize the problems. This type of problem along with categorical solutions like yes or no, true or false, fit or unfit and so on.

#### 2. Regression tree

It is also known as a Continuous Variable decision tree which is used to when the target variable is a quantitative variable. This type of problem wherein continuous value needs to be predicted like weight, height, temperature, profit, and so on.

- **Types of decision tree algorithms**

There are lots of tree-based algorithms are available to use. Here the most used algorithms are mentioned as follows,

- Hunt Algorithm
- CHAID
- ID3
- CART
- C4.5
- C5.0

#### 3. Areas used

The decision tree is used in many areas such as industry, remote sensing, intelligent vehicles, web applications, E-commerce, image processing, health sector, business and so on. It is mainly focused on the area of Predictive analytics.

#### 4. Advantages

- Easy to implement and understand
- The decision tree does not require scaling, normalization of data.

#### 5. Dis- advantages

- Overfitting
- Less precision
- It consumes high more time to train the model.

- d. Problems on fragmentation, repetition, and replication.

**Random Forest**

The Random forest algorithm is a Supervised learning algorithm which has the main goal to accomplish the prediction with best-improved accuracy by without relying on a single learning model. It creates and merges the single decision tree into many decision trees that compose the one forest as the name implies. It is also can be used in classification and regression. The random forest also called Radom Decision Forest which is the ensemble learning model for both classification and regression. The problems are constructing by multiple decision trees during the training time and the final output came from the average prediction of the individual trees.

*Advantages*

The accuracy of prediction is higher than the decision tree.

Random forest algorithm overcomes the problem overfitting.

**The Real Time Analysis of Comparative Survey using the Weka tool:**

S.NO	TOPIC	DECISION TREE	RANDOM FOREST
		Split 70% train and 30% test data	
1	CONTACT LENSE	28	42
2	BREAST CANCER	63.95	69.7
3	CREDIT	73.6	77.3
4	DIABETES	76.5	76.5
5	IRIS	95.5	95.5
6	SOYBEAN	89.2	92.6
7	SUPERMARKET	62.6	62.6
8	UNBALANCED	98	98
9	VOTE	96.1	97.6
10	WEATHER	25	50

(Overall Analysis of data set using Weka Tool)

**Contact Lenses:**

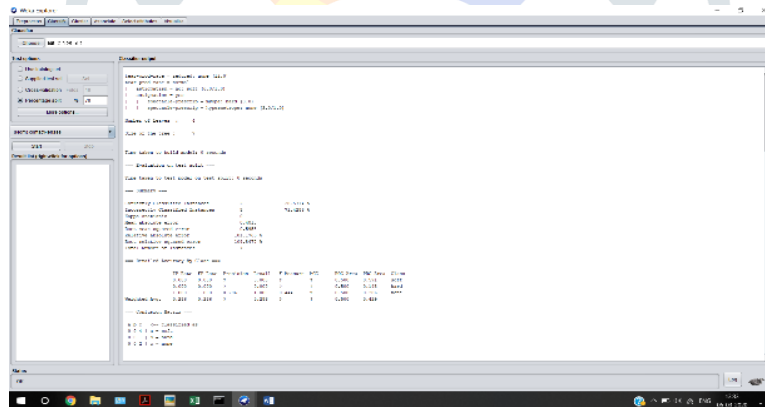
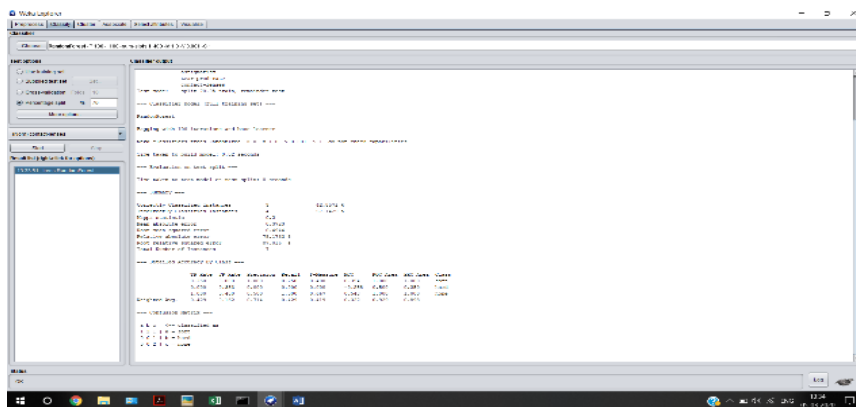


Fig. 1.

A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 04 which is the decisions underwent and Size of the tree is 07 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 28.5714% and incorrectly classified Instances are 71.4286% along with Kappa statistic as 0. Result also illustrates that Mean absolute error as 0.4818 as well as Root Mean squared error as 0.5958 and Relative absolute error as 101.1765% as well as Root Relative squared error 104.8477% with total no .of instance as 7.



Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 42.8571% and incorrectly classified Instances are 57.1429% along with Kappa statistic as 0.2. Result also illustrates that Mean absolute error as 0.3723 as well as Root Mean squared error as 0.4944 and Relative absolute error as 78.1733% as well as Root Relative squared error 87.016% with total no .of instance as 7.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Breast Cancer:**

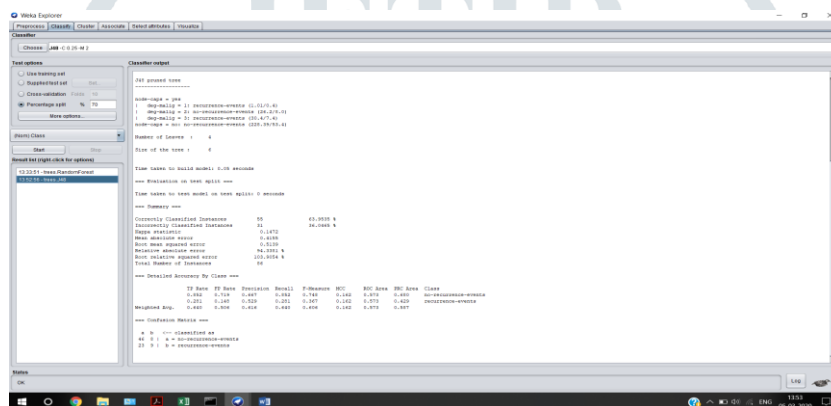


Fig 6.

A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 04 which is the decisions underwent and Size of the tree is 06 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 63.9535% and incorrectly classified Instances are 36.0465% along with Kappa statistic as 0.1472. Result also illustrates that Mean absolute error as 0.4155 as well as Root Mean squared error as 0.5139 and Relative absolute error as 94.3381% as well as Root Relative squared error 103.9054% with total no .of instance as 86.

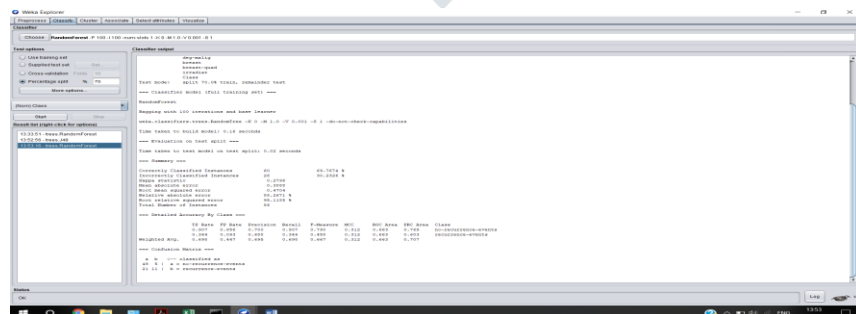


Fig 7.

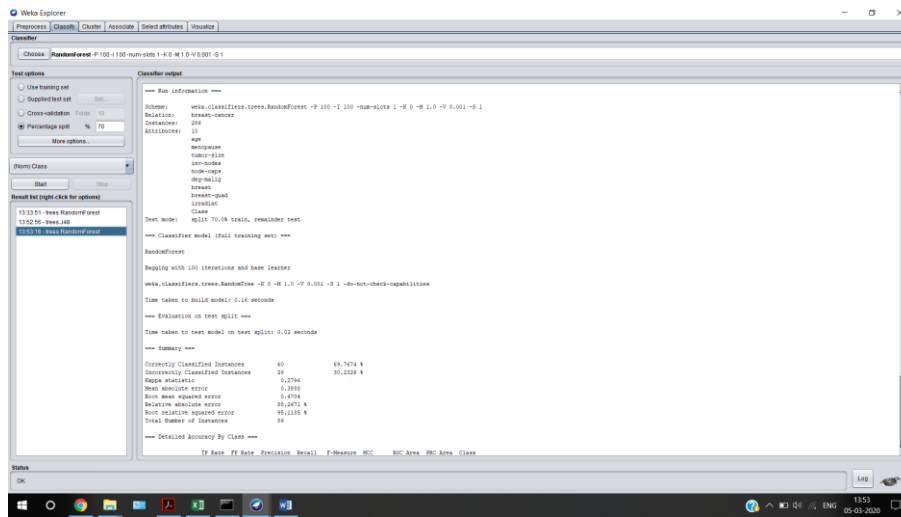


Fig 8 .

Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 69.7674% and incorrectly classified Instances are 30.2326% along with Kappa statistic as 0.2796. Result also illustrates that Mean absolute error as 0.3888 as well as Root Mean squared error as 0.4704 and Relative absolute error as 88.2671% as well as Root Relative squared error 95.1185% with total no .of instance as 86.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Credit:**

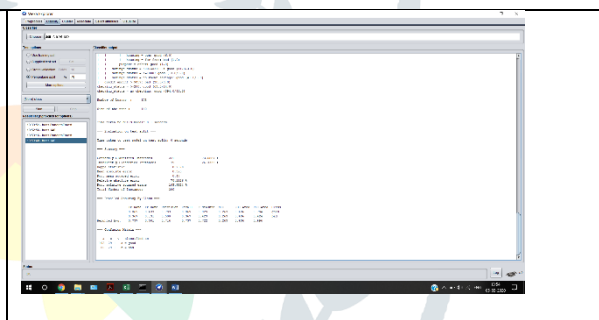
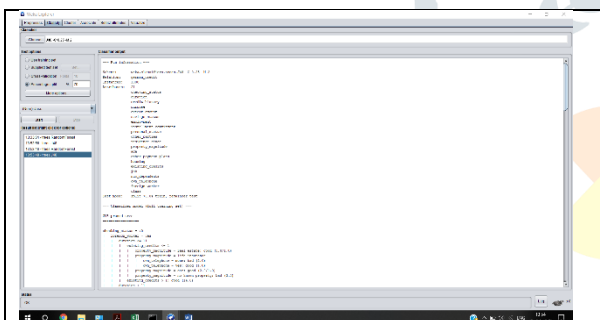


Fig 9.

Fig 10.

A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 103 which is the decisions underwent and Size of the tree is 140 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 73.6667% and incorrectly classified Instances are 26.3333% along with Kappa statistic as 0.2579. Result also illustrates that Mean absolute error as 0.323 as well as Root Mean squared error as 0.47 and Relative absolute error as 78.2126% as well as Root Relative squared error 105.9524% with total no .of instance as 300.

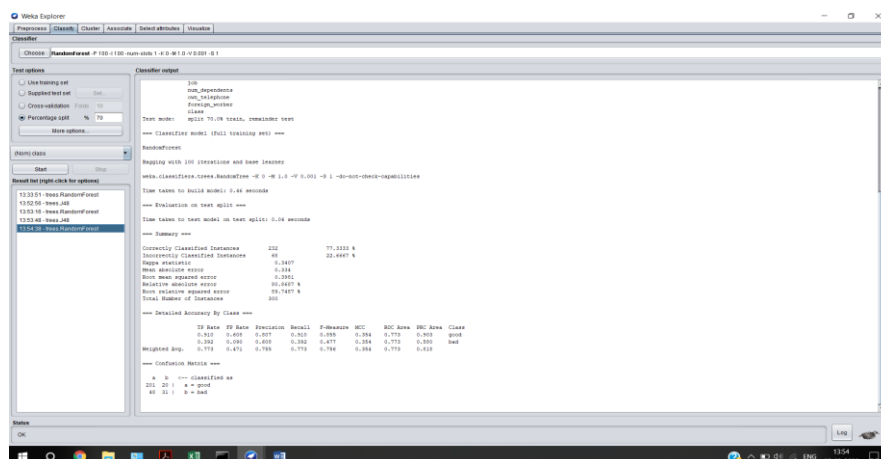


Fig.11

Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 77.3333% and incorrectly classified Instances are 22.6667% along with Kappa statistic as 0.3407. Result also illustrates that Mean absolute error as 0.334 as well as Root Mean squared error as 0.3981 and Relative absolute error as 80.8687% as well as Root Relative squared error 89.7487% with total no .of instance as 300.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Diabetes:**

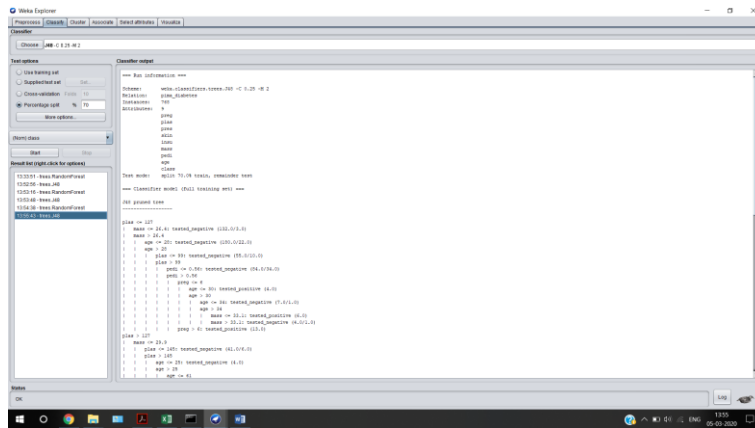
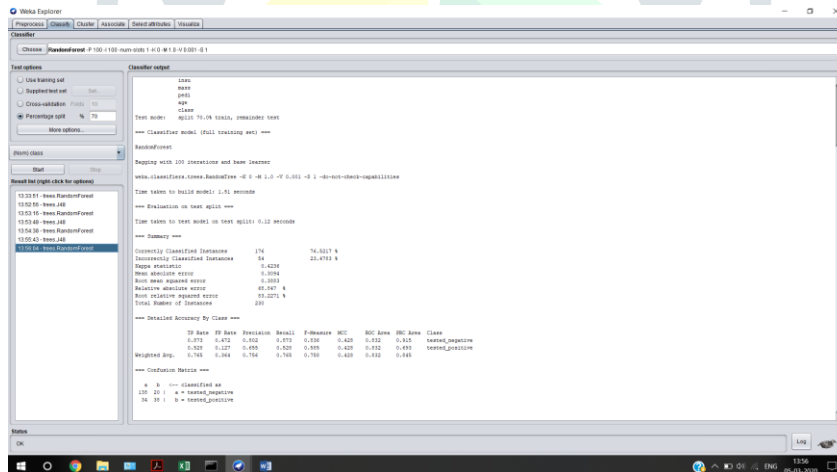


Fig 13.

A default dataset which is present in the Weka tool is present for data analysis and it has been taken into s\consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 20 which is the decisions underwent and Size of the tree is 39 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 76.5217% and incorrectly classified Instances are 23.4783% along with Kappa statistic as 0.4889. Result also illustrates that Mean absolute error as 0.3206 as well as Root Mean squared error as 0.4239 and Relative absolute error as 71.3381% as well as Root Relative squared error 90.8521% with total no .of instance as 230.

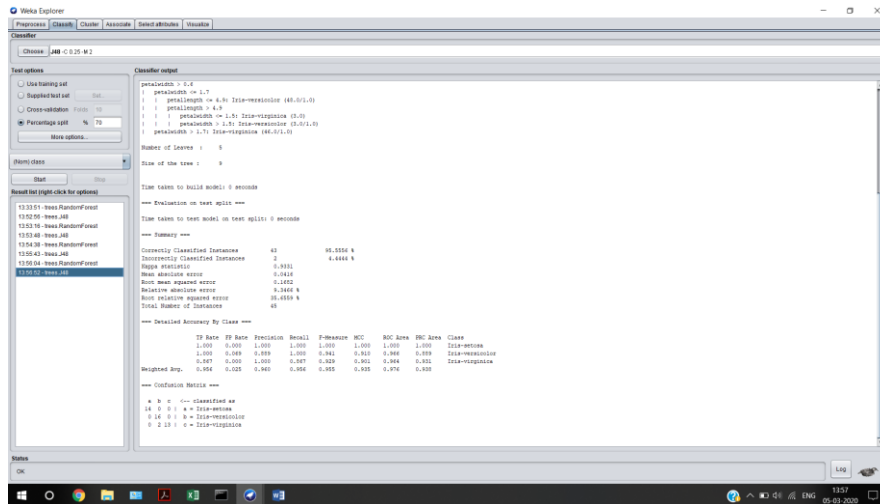


Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 76.5217% and incorrectly classified Instances are 23.4783% along with Kappa statistic as 0.4236. Result also illustrates that Mean absolute error as 0.3094 as well as Root Mean squared error as 0.3883 and Relative absolute error as 68.847% as well as Root Relative squared error 83.2271% with total no .of instance as 230.

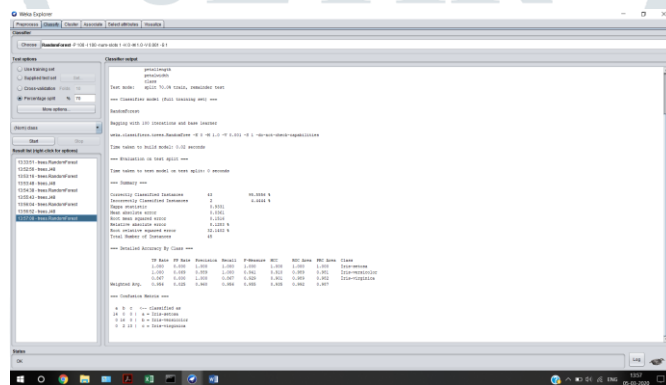
By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a same predictive CCI when compared to Decision Tree algorithm, Hence to prove random forest is the preferable, incorrectly classified instances are taken into consideration.



**Iris:**



A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 05 which is the decisions underwent and Size of the tree is 09 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 95.5556% and incorrectly classified Instances are 4.4444% along with Kappa statistic as 0.9331. Result also illustrates that Mean absolute error as 0.0416 as well as Root Mean squared error as 0.1682 and Relative absolute error as 9.3466% as well as Root Relative squared error 35.6559% with total no .of instance as 45.

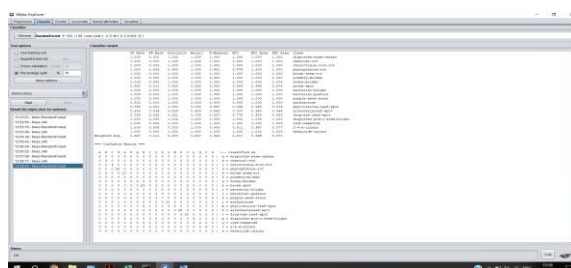


Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 95.5556% and incorrectly classified Instances are 4.4444% along with Kappa statistic as 0.9331. Result also illustrates that Mean absolute error as 0.361 as well as Root Mean squared error as 0.1516 and Relative absolute error as 8.1283% as well as Root Relative squared error 32.1402% with total no .of instance as 45.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Soybean:**

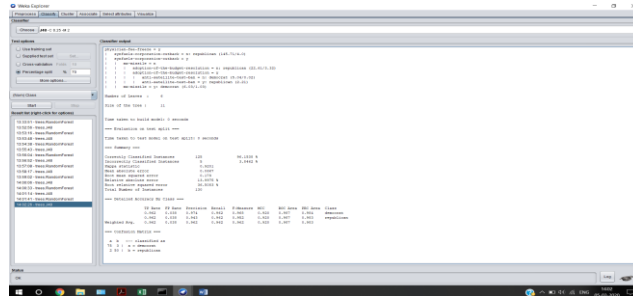
A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 04 which is the decisions underwent and Size of the tree is 07 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 89.2683% and incorrectly classified Instances are 10.7317% along with Kappa statistic as 0.8817. Result also illustrates that Mean absolute error as 0.0155 as well as Root Mean squared error as 0.093 and Relative absolute error as 16.1519% as well as Root Relative squared error 42.4571% with total no .of instance as 205.



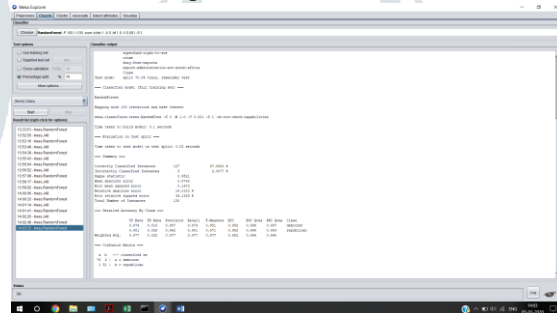
Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 92.6829% and incorrectly classified Instances are 7.3171% along with Kappa statistic as 0.9194. Result also illustrates that Mean absolute error as 0.0284 as well as Root Mean squared error as 0.0962 and Relative absolute error as 29.549% as well as Root Relative squared error 43.9172% with total no .of instance as 205.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Vote:**



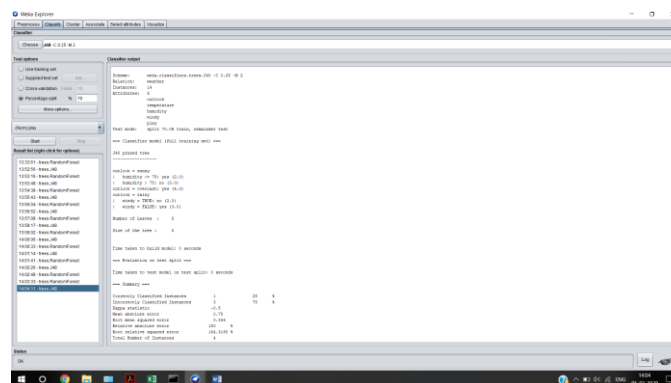
A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 06 which is the decisions underwent and Size of the tree is 11 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 96.1538% and incorrectly classified Instances are 3.8462% along with Kappa statistic as 0.9201. Result also illustrates that Mean absolute error as 0.0667 as well as Root Mean squared error as 0.179 and Relative absolute error as 13.9975% as well as Root Relative squared error 36.5083% with total no .of instance as 103.

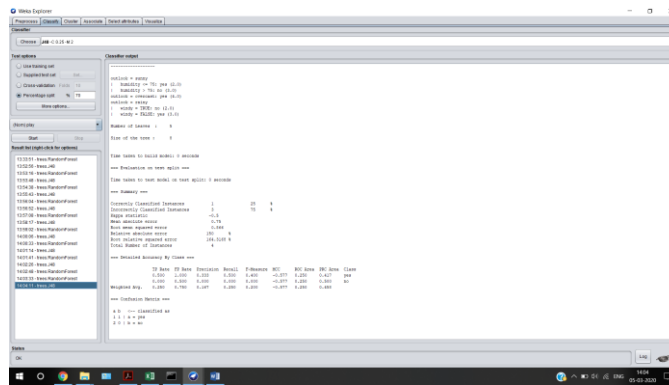


Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 97.6923% and incorrectly classified Instances are 2.3077% along with Kappa statistic as 0.9521. Result also illustrates that Mean absolute error as 0.0763 as well as Root Mean squared error as 0.1673 and Relative absolute error as 16.0183% as well as Root Relative squared error 34.1329% with total no .of instance as 130.

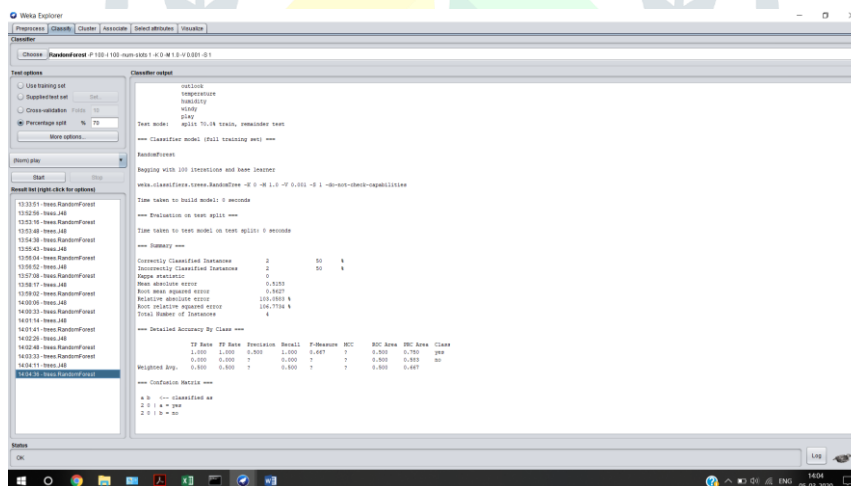
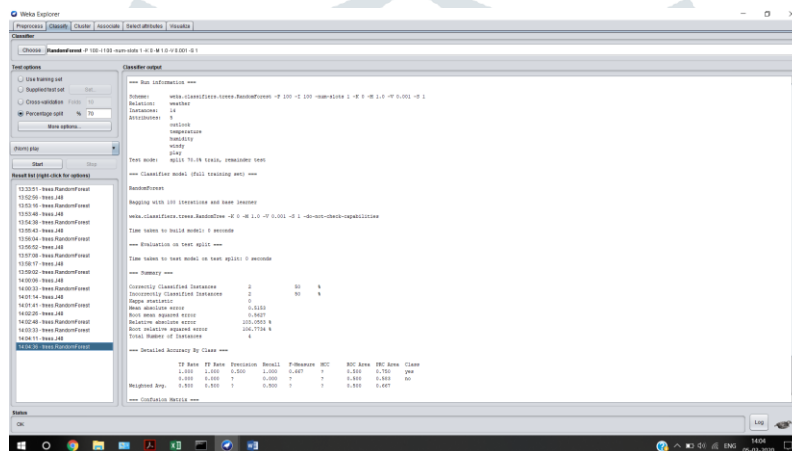
By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm.

**Weather:**





A default dataset which is present in the Weka tool is present for data analysis and it has been taken into consideration in the above data analysis using J48 or in other words known as decision tree algorithm. When the decision tree algorithm was applied over the Contact lenses data set, the result which had been obtained contains the following data such as No of Leaves is 05 which is the decisions underwent and Size of the tree is 08 which is the split of the data based on the process. Result also shows that the Correctly Classified Instances are 25% and incorrectly classified Instances are 75% along with Kappa statistic as -0.5. Result also illustrates that Mean absolute error as 0.75 as well as Root Mean squared error as 0.866 and Relative absolute error as 150% as well as Root Relative squared error 164.3168% with total no .of instance as 04.



Then the data set is passes on to analysis process where the Random Forest algorithm is applied and the result is gained. The result which had been obtained contains the following data such as Test mode of split 70.0% training data set. Result also shows that the Correctly Classified Instances (CCI) are 50% and incorrectly classified Instances are 50% along with Kappa statistic as 0. Result also illustrates that Mean absolute error as 0.5153 as well as Root Mean squared error as 0.5627 and Relative absolute error as 103.0583% as well as Root Relative squared error 106.7734% with total no .of instance as 04.

By the above comparative analysis between decision tree algorithm and Random forest algorithm that random forest had returned a good predictive CCI when compared to Decision Tree algorithm. For the case analysis, ten datasets which are present in WEKA as default. By the above analysis we came into the conclusion that random forest is better in giving up precision of data than decision tree. Since as we studied that random forest form multiple decision trees in order to obtain a particular result and thereby in other hand decision tree form a single structure when compared to the random forest which delivers us a better result. Hence is concluded that Random Forest algorithm is better than Decision Tree algorithm.

**Conclusion:-**

This paper thereby provides a detailed comparison between the Random forest algorithm and the Decision Tree algorithm along with a survey of how these algorithms play a vital role in predictive analytics. By the done surveyed study, this paper proposes the theory of how Random forest can be used to provide more precision in data than the Decision Tree algorithm in any of the predictive analytical processing.

**References:-**

- [1] Amy Shi-Nash and David r. Hardoon, "Data Analytics and Predictive Analytics in the era of Big Data" <https://doi.org/10.1002/9781119173601.ch19>.
- [2] Fatimetou Zahra Mohamed Mahmoud , "The Application of Predictive Analytics: Benefits, Challenges and how it can be improved " International Journal of Scientific and Research Publications, 2017.
- [3] Mr. Brijain R Patel, Mr. Kushik K Rana, "A Survey on Decision Tree Algorithm for Classification", <http://citeseerx.ist.psu.edu/doi=10.1.1.673.2797>.
- [4] . Mr. D. V. Patil Prof. Dr. R. S. Bichkar., "A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets" IEEE Xplore,2006.
- [5] Shiueng-Bien Yang,Shen-I Yang , "New Decision Tree Based on Genetic Algorithm" IEEE Xplore,2010.
- [6] Li Wenlong, Xing Changzheng , "Parallel Decision Tree Algorithm Based on Combination" IEEE Xplore,2019
- [7] G. Niveditha, K. Abarna, G. V. Akshaya, "Credit Card Fraud Detection Using Random Forest Algorithm"International Journal of Scientific Research in Computer Science Engineering and Information Technology, 2019
- [9] Naili Suri Intizhami, Ario Yudo Husodo, Wisnu Jatmiko, "Warfare Simulation: Predicting Battleship Winner Using Random Forest" IEEE Xplore,2019
- [10] Karthick, Malathi,,Arun , "WEATHER PREDICTION ANALYSIS USING RANDOM FOREST ALGORITHM " International Journal of Pure and Applied Mathematics, 2018

