International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 2Ver III || February 2018 || PP. 01-09

Comparative Analysis of the WEKA Classifiers Rules Conjunctiverule & Decisiontable on Indian News Dataset by Using Different Test Mode

Sushilkumar Rameshpant Kalmegh
Department of Computer Science, Sant Gadge Baba Amravati University, Amravati, (M.S.), India

Abstract: *Classification may refer to categorization, the process in which ideas and objects are recognized, differentiated, and understood. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. Classification is an important data mining technique with broad applications. It classifies data of various kinds. This paper has been carried out to make a performance evaluation of ConjunctiveRule and DecisionTable classification algorithm. The paper sets out to make comparative evaluation of classifiers rules ConjunctiveRule and DecisionTable in test mode (i) evaluate on training data, (ii) 5-fold cross-validation and (iii)10-fold cross-validation in the context of dataset of Indian news to maximize true positive rate and minimize false positive rate. The Weka workbench is an organized collection of state-of-the-art machine learning algorithms and data pre processing tools. The basic way of interacting with these methods is by invoking them from the command line. However, convenient interactive graphical user interfaces are provided for data exploration. These interfaces constitute an advanced environment for experimental data mining.*

Keywords – *Classification, Conjunctive Rule, Data Mining, Decision Table, WEKA*

Date of Submission: 09-02-2018 Date of acceptance: 24-02-2018

I. INTRODUCTION

Each of the past three centuries has been dominated by a single technology. The eighteenth century was the time of the great mechanical systems accompanying the Industrial Revolution. The nineteenth century was the age of the stream engine. During the twentieth century, the key technology has been information gathering, processing and distribution. Among other developments, we have seen the birth and unprecedented growth of the computer industry. Now as we have entered in the twenty-first century all the most of all manual services are replaced by machine operation i.e. complete computerization and hence released human intelligence is utilized in further developments.

The amount of data in the world and in our lives seems ever-increasing and there's no end to it. We are overwhelmed with data. Today Computers make it too easy to save things. Inexpensive disks and online storage make it too easy to postpone decisions about what to do with all this stuff, we simply get more memory and keep it all. The World Wide Web (WWW) overwhelms us with information; meanwhile, every choice we make is recorded. As the volume of data increases, inexorably, the proportion of it that people understand decreases alarmingly. Lying hidden in all this data is information.

In data mining, the data is stored electronically and the search is automated or at least augmented by computer. Even this is not particularly new. Economists, statisticians, and communication engineers have long worked with the idea that patterns in data can be sought automatically, identified, validated, and used for prediction. What is new is the staggering increase in opportunities for finding patterns in data.

Data mining is a topic that involves learning in a practical, non theoretical sense. We are interested in techniques for finding and describing structural patterns in data, as a tool for helping to explain that data and make predictions from it. Experience shows that in many applications of machine learning to data mining, the explicit knowledge structures that are acquired, the structural descriptions, are at least as important as the ability to perform well on new examples. People frequently use data mining to gain knowledge, not just predictions.

We are interested in techniques for finding and describing structural patterns in data, as a tool for helping to explain that data and make predictions from it. Experience shows that in many applications of machine learning to data mining, the explicit knowledge structures that are acquired, and the structural descriptions, are at least as important as the ability to perform well on new examples. This paper is organized into Seven parts. First part is the Introduction followed by Literature Survey which discusses the concept WEKA. Third one is Classification in which Classifier Conjunctive Rule & Decision Table has been discussed. Fourth is System Design followed by datasets used for analysis. Sixth one is the Performance Analysis and then Conclusions.

II. LITERATURE SURVEY

2.1 WEKA

Weka was developed at the University of Waikato in New Zealand; the name stands for Waikato Environment for Knowledge Analysis The system is written in Java and distributed under the terms of the GNU General Public License. It runs on almost any platform and has been tested under Linux, Windows, and Macintosh operating systems and even on a personal digital assistant. It provides a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset. Weka provides implementations of learning algorithms that can be easily apply to dataset. It also includes a variety of tools for transforming datasets, such as the algorithms.

The Weka workbench is a collection of state-of-the-art machine learning algorithms and data pre processing tools. It is designed so that we can quickly try out existing methods on new datasets in flexible ways. It provides extensive support for the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning. As well as a variety of learning algorithms, it includes a wide range of pre processing tools. This diverse and comprehensive toolkit is accessed through a common interface so that its users can compare different methods and identify those that are most appropriate for the problem at hand. All algorithms take their input in the form of a single relational table in the ARFF format. The easiest way to use Weka is through a graphical user interface called Explorer as shown in **Figure 1.** This gives access to all of its facilities using menu selection and form filling**.**

Fig. 1 : Weka GUI Explorer

The Weka contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to this functionality. Advantages of Weka include:

- Free availability under the GNU General Public License
- Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.
- A comprehensive collection of data pre-processing and modelling techniques.
- Ease of use due to its graphical user interfaces.

Weka supports several standard data mining tasks, more specifically, data pre-processing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. Weka's main user interface is the Explorer, but essentially the same functionality can be accessed through the component-based Knowledge Flow interface and from the command line. There is also the Experimenter, which allows the systematic comparison of the predictive performance of Weka's machine learning algorithms on a collection of datasets. The Explorer interface features several panels providing access to the main components of the workbench. **Figure 2** shows Opening of file *.arff by Weka Explorer and **Figure 3** shows processing of arff file for Conjunctive Rule Classifier using Test Mode: evaluate on training data. [1, 8]

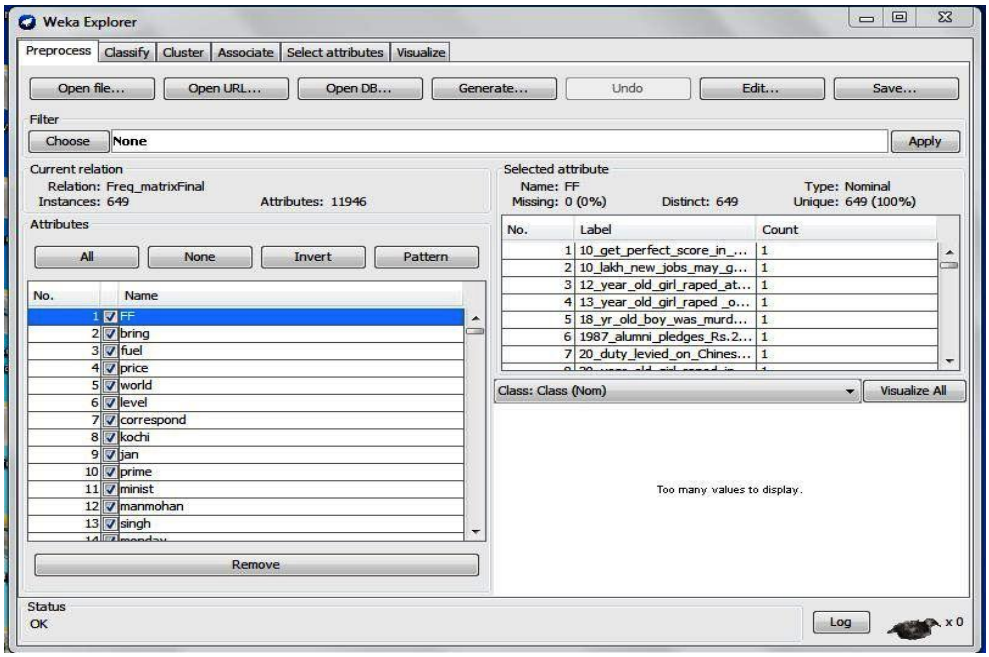

FIG. 2 : OPENING OF FILE *.ARFF BY WEKA EXPLORER

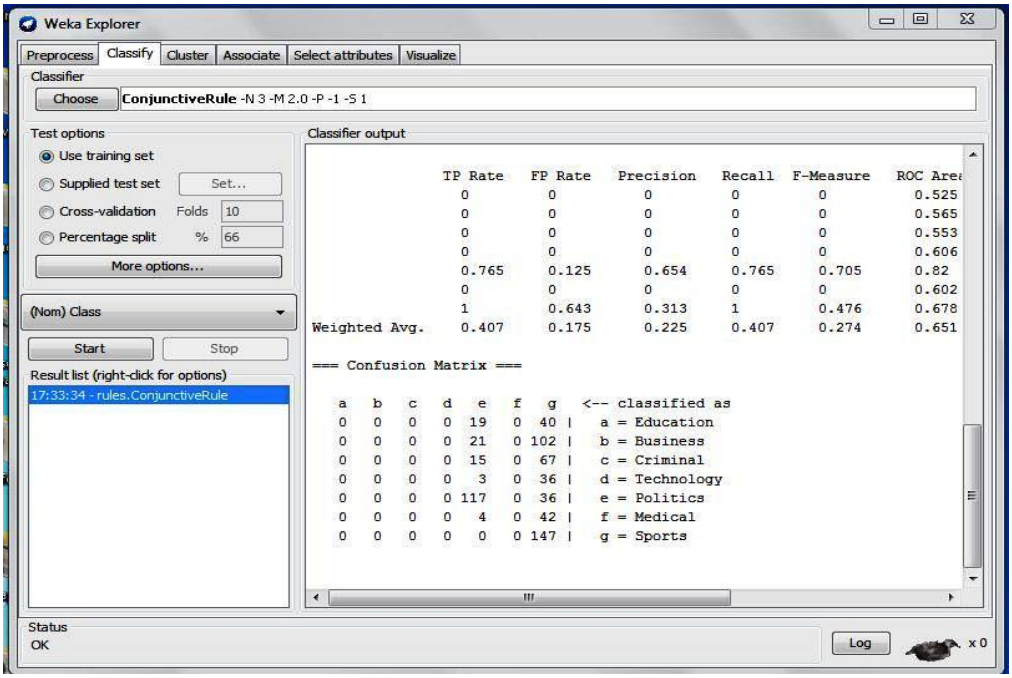

FIG. 3: PROCESSING OF ARFF FILE BY CONJUNCTIVE RULE CLASSIFIER, TEST MODE: EVALUATE ON TRAINING DATA

III. CLASSIFICATION

Classification may refer to categorization, the process in which ideas and objects are recognized, differentiated, and understood. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category.

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering or cluster analysis, and involves grouping data into categories based on some measure of inherent similarity.

Classification is a data mining algorithm that creates a step-by-step guide for how to determine the output of a new data instance. The tree it creates is exactly that: a tree whereby each node in the tree represents a

spot where a decision must be made based on the input, and to move to the next node and the next until one reach a leaf that tells the predicted output. Sounds confusing, but it's really quite straightforward.

There is also some argument over whether classification methods that do not involve a statistical model can be considered "statistical". Other fields may use different terminology: e.g. in community ecology, the term "classification" normally refers to cluster analysis, i.e. a type of unsupervised learning, rather than the supervised learning. [1, 8]

3.1. Conjunctive Rule Classifier

Conjunctive Rule algorithm implements a single conjunctive rule learner that can predict for numeric and nominal class values. Conjunctive rule uses the relation of logical AND to link stimulus attributes. The rule involves 'AND'ing the antecedents together and the consequent (class value) for the classification. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset. If this rule does not enclose the test instance, then the default class distributions/value of data that is not enclosed by the rule in the training data is used to predict it. An antecedent is selected by this learner by calculating the Information Gain of each antecedent and the generated rule is pruned using Reduced Error Pruning (REP) or simple pre-pruning depending on the number of antecedents. The weighted mean of the entropies of both the data covered and not covered by the rule is the Information of one antecedent used for classification. Single conjunctive rule learner is one of the machine learning algorithms and is normally known as inductive Learning.

The goal of rule induction is generally to induce a set of rules from data that captures all generalizable knowledge within that data, and at the same time being as small as possible. Classification in rule-induction classifiers is typically based on the firing of a rule on a test instance, triggered by matching feature values at the left-hand side of the rule. Rules can be of various normal forms, and are typically ordered; with ordered rules, the first rule that fires determines the classification outcome and halts the classification process.

Uncovered test instances are assigned the default class value (or distribution) of the uncovered training instances. The information gain (nominal class) or variance reduction (numeric class) of each antecedent is computed, and rules are pruned using reduced-error pruning.

The limitation of this classifier: if a concept does not have a single set of necessary and sufficient conditions conjunctive learning fails.[9, 2, 3, 4, 8].

3.2. Decision Table Classifiers

Decision Table is an accurate method for numeric prediction from decision trees and it is an ordered set of If-Then rules that have the potential to be more compact and therefore more understandable than the decision trees. Selection to explore decision tables because it is a simpler, less compute intensive algorithm than the decision-tree-based approach.

The algorithm, decision table, is found in the Weka classifiers under Rules. The simplest way of representing the output from machine learning is to put it in the same form as the input. It summarizes the dataset with a "decision table" which contains the same number of attributes as the original dataset. The use of the classifier rules decision table is described as building and using a simple decision table majority classifier. The output will show a decision on a number of attributes for each instance. The number and specific types of attributes can vary to suit the needs of the task. Decision Table classifier algorithm is used to summarize the dataset by using a decision table containing the same number of attributes as that of the original dataset. A new data item is allocated a category by searching the line in the decision table that is equivalent to the values contained in the non-class of the data item.

The entire problem of learning decision tables consists of selecting the right attributes to be included. Usually this is done by measuring the tables cross validation performance for different subsets of attributes and choosing the best performing subset. Fortunately, leave-one-out cross-validation is very cheap for this kind of classifier. Obtaining the cross-validation error from a decision table derived from the training data is just a matter of manipulating the class counts associated with each of the table's entries, because the table's structure doesn't change when instances are added or deleted. The attribute space is generally searched by best-first search because this strategy is less likely to get stuck in a local maximum than others, such as forward selection.

Decision Table are one of the simplest hypothesis spaces possible and usually they are easy to understand. Decision Table builds a decision table majority classifier. It evaluates feature subsets using best-first search and can use cross-validation for evaluation. An option uses the nearest-neighbour method to determine the class for each instance that is not covered by a decision table entry, instead of the table's global majority, based on the same set of features. [**5, 8, 6, 2, 3, 7]**

IV. SYSTEM DESIGN

In order to co-relate News with the categories, a model has been designed. Flow diagram of the model for news resources is shown below in **Figure 4.** As a input to the model, various news resources are considered which are available online like the news in Google news repository or online paper like Times of India, Hindustan Times etc. Around 649 news were collected on above repository. In order to extract context from the

news and co-relate it, the News was process with Stop words removal, stemming and tokenization on the news contents. The news then was converted into the term frequency matrix for further analysis purpose. The frequency matrix is having extension .csv, so it has to be converted in arff format for processing by WEKA. Based on this data, features (i.e. metadata) were extracted so that contextual assignment of the news to the appropriate content can be done. Title of the news also contains useful information in the abstract form, the title also can be considered as Metadata. The title of the news is processed using NLP libraries (Standford NLP Library) to extract various constituents of it

As shown in the figure, a news resource is processed to correlate with the Contents available. On the similar way, other text resources can be added directly in the repository, Image or Video resource can be processed for meta-data available.

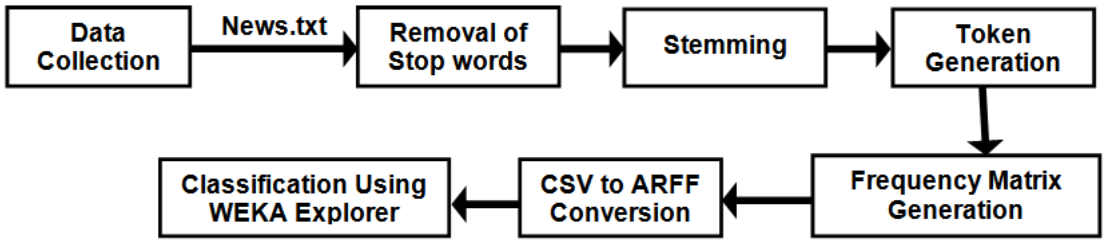

Fig. 4: Flow diagram of the model

V. DATA COLLECTION

Table 1: Categorization of News

News Category	Actual No. Of News
Business	123
Criminal	82
Education	59
Medical	46
Politics	153
Sports	147
Technology	39
Total	649

Hence it was proposed to generate indigenous data. Consequently the national resources were used for the research purpose. Data for the purpose of research has been collected from the various news which are available in various national and regional newspapers available on internet. They are downloaded and after reading the news they are manually classified into 7 (seven) categories. There were 649 news in total. The details are as shown in **Table 1.**

The attributes consider for this classification is the topic to which news are related; the statements made by different persons; the invention in Business, Education, Medical, Technology; the various trends in Business; various criminal acts e.g. IPC and Sports analysis. During classification some news cannot be classified easily e.g.

(1) Political leader arrested under some IPC code,

(2) Some invention made in medicine and launched in the market & business done per annum.

Hence, there will be drastic enhancement in the Contents when we refer to the latest material available in this regards. For example, if some news refers to the political situation of India, then the references needs to be dynamic as the situation may change depending on the result of election. [1, 5]

VI. PERFORMANCE ANALYSIS

The News so collected needed a processing. Hence as given in the design phase, all the news were processed for stop word removal, stemming, tokenization and ultimately generated the frequency matrix. Stemming is used as many times when news is printed, for a same there can be many variants depending on the tense used or whether it is singular or plural. Such words when processed for stemming, generates a unique word. Stop words needs to be removed as they do not contribute much in the decision making process. Frequency matrix thus generated can be processed for generating a model and the model so generated was used in further decision process.

With the model discussed above, classifier Conjunctive Rule & Decision Table from rules. was used on the data set of 649 news. For processing Weka GUI interface were used. The result after processing is given in

the form of Confusion Matrix and True Positive and False Positive Rate matrix for test mode i) evaluate on training data, ii) 5-fold cross-validation and iii) 10-fold cross-validation.

For Conjunctive Rule classifier Confusion Matrix for test mode i) evaluate on training data, ii) 5-fold cross-validation and iii) 10-fold cross-validation are given in the following **Table no. 2, 4 & 6** and True Positive and False Positive Rate matrix are given in **Table no. 3, 5 & 7.**

For Decision Table classifier Confusion Matrix for test mode i) evaluate on training data, ii) 5-fold cross-validation and iii) 10-fold cross-validation are given in the following **Table no. 8, 10 & 12** and True Positive and False Positive Rate matrix are given in **Table no. 9, 11 & 13.**

Performance of **Conjunctive Rule** algorithm is worst as compare to overall performance of **Decision Table** for all the three test mode i) evaluate on training data, ii) 5-fold cross-validation and iii) 10-fold cross-validation. It can been seen from Confusion Matrix **Table no. 2** for Test mode: evaluate on training data that all the **147** news from **Sports Category** & **117** news out of **153** from **Politics Category** are correctly classified, whereas news from all other category are classified into different category. And from the **Table 4 and 6** same situation appeared, whereas news from all other category are classified into different category for Test mode: 5-fold cross-validation and 10-fold cross-validation. This is because every category has some or other references of the other category. Further Conjunctive Rule consists of antecedents "AND"ed together and the consequent (class value) for the classification. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset. If the test instance is not covered by this rule, then it's predicted using the default class distributions/value of the data not covered by the rule in the training data. This learner selects an antecedent by computing the Information Gain of each antecedent and prunes the generated rule using Reduced Error Prunning (REP). For classification, the Information of one antecedent is the weighted average of the entropies of both the data covered and not covered by the rule. The limitation of Conjunctive Rule classifier is that if a concept does not have a single set of necessary and sufficient conditions conjunctive learning fails.

Overall Performance of **Decision Table** classifier is fairly good as compare to performance of **Conjunctive Rule** classifier for all the three test mode i) evaluate on training data, ii) 5-fold cross-validation and iii) 10-fold cross-validation. It can be seen from Confusion Matrix **Table no. 8, 10.12.** The performance of **Decision Table** classifier deteriorate in the test mode ii) 5-fold cross-validation and iii) 10-fold cross-validation as compare to test mode i) evaluate on training data. Hence as it can be seen in the **Table 10** and **Table 12** it has given less accuracy as compare to **Table no. 8**. The reason for this is that, in n-fold cross-validation, the original sample is randomly partitioned into n subsamples. Of the n subsamples, a single subsample is retained as the validation data for testing the model, and the remaining n – 1 sub samples are used as training data. The cross-validation process is then repeated n times (the folds), with each of the n subsamples used exactly once as the validation data. The n results from the folds then can be averaged (or otherwise combined) to produce a single estimation.

Table 2: Confusion Matrix for ConjunctiveRule for Test mode: evaluate on training data

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	0	0	0	0	19	0	40
Business	0	0	0	0	21	0	102
Criminal	0	0	0	0	15	0	67
Technology	0	0	0	0	3	0	36
Politics	0	0	0	0	117	0	36
Medical	0	0	0	0	4	0	42
Sports	0	0	0	0	0	0	147

Table 3: Table showing True Positive and False Positive Rate of ConjunctiveRule for Test mode: evaluate on training data

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	0%	0%	0%	0%	0%	52.5%
Business	0%	0%	0%	0%	0%	56.5%
Criminal	0%	0%	0%	0%	0%	55.3%
Technology	0%	0%	0%	0%	0%	60.6%
Politics	76.5%	12.5%	65.4%	76.5%	70.5%	82%
Medical	0%	0%	0%	0%	0%	60.2%
Sports	100%	64.3%	31.3%	100%	47.6%	67.8%
Weighted Avg. ➡	40.7%	17.5%	22.5%	40.7%	27.4%	65.1%

Table 4: Confusion Matrix for ConjunctiveRule for Test mode: 5-fold cross-validation

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	0	0	0	0	36	0	23
Business	0	0	0	0	64	0	59
Criminal	0	0	0	0	41	0	41
Technology	0	0	0	0	18	0	21
Politics	0	0	0	0	129	0	24
Medical	0	0	0	0	21	0	25
Sports	0	0	0	0	22	0	125

Table 5: Table showing True Positive and False Positive Rate of ConjunctiveRule for Test mode: 5-fold cross-validation

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	0%	0%	0%	0%	0%	55.6%
Business	0%	0%	0%	0%	0%	56.7%
Criminal	0%	0%	0%	0%	0%	54.2%
Technology	0%	0%	0%	0%	0%	56.7%
Politics	84.3%	40.7%	39%	84.3%	53.3%	76.8%
Medical	0%	0%	0%	0%	0%	56.2%
Sports	85%	38.4%	39.3%	85%	53.8%	79%
Weighted Avg. ➡	39.1%	18.3%	18.1%	39.1%	24.7%	66%

Table 6: Confusion Matrix for ConjunctiveRule for Test mode: 10-fold cross-validation

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	0	0	0	0	41	0	18
Business	0	0	0	0	74	0	49
Criminal	0	0	0	0	50	0	32
Technology	0	0	0	0	19	0	20
Politics	0	0	0	0	135	0	18
Medical	0	0	0	0	23	0	23
Sports	0	0	0	0	31	0	116

Table 7: Table showing True Positive and False Positive Rate of ConjunctiveRule for Test mode: 10-fold cross-validation

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	0%	0%	0%	0%	0%	58.2%
Business	0%	0%	0%	0%	0%	55.9%
Criminal	0%	0%	0%	0%	0%	53.1%
Technology	0%	0%	0%	0%	0%	61.4%
Politics	88.2%	48%	36.2%	88.2%	51.3%	76%
Medical	0%	0%	0%	0%	0%	60.2%
Sports	78.9%	31.9%	42%	78.9%	54.8%	79.5%
Weighted Avg. ➡	38.7%	18.5%	18.1%	38.7%	24.5%	66.5%

Table 8: Confusion Matrix For Decisiontable For Test Mode: Evaluate On Training Data

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	44	2	3	0	4	0	6
Business	4	79	2	0	9	0	29
Criminal	11	2	55	0	8	1	5
Technology	5	9	2	9	2	0	12
Politics	43	0	3	0	105	0	2
Medical	6	2	0	0	3	31	4
Sports	2	2	1	0	0	1	141

Table 9: Table showing True Positive and False Positive Rate of DecisionTable for Test mode: evaluate on training data

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	74.6%	12%	38.3%	74.6%	50.6%	86.3%
Business	64.2%	3.2%	82.3%	64.2%	72.1%	89.3%
Criminal	67.1%	1.9%	83.3%	67.1%	74.3%	89.8%
Technology	23.1%	0%	1%	23.1%	37.5%	66.7%
Politics	68.6%	5.2%	80.2%	68.6%	73.9%	94.8%
Medical	67.4%	0.3%	93.9%	67.4%	78.5%	91%
Sports	95.9%	11.6%	70.9%	95.9%	81.5%	92.1%
Weighted Avg. ➡	71.5%	5.8%	77.2%	71.5%	71.4%	89.8%

Table 10: Confusion Matrix for DecisionTable for Test mode: 5-fold cross-validation

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	24	6	3	0	18	2	6
Business	1	63	2	5	21	2	29
Criminal	7	4	49	0	16	0	6
Technology	2	8	2	8	8	1	10
Politics	19	5	8	1	117	0	3
Medical	5	5	0	2	13	15	6
Sports	2	12	1	0	4	1	127

Table 11: Table showing True Positive and False Positive Rate of DecisionTable for Test mode: 5-fold cross-validation

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	40.7%	6.1%	40%	40.7%	40.3%	75.6%
Business	51.2%	7.6%	61.2%	51.2%	55.8%	79.1%
Criminal	59.8%	2.8%	75.4%	59.8%	66.7%	85%
Technology	20.5%	1.3%	50%	20.5%	29.1%	59.1%
Politics	76.5%	16.1%	59.4%	76.5%	66.9%	88.8%
Medical	32.6%	1%	71.4%	32.6%	44.8%	75.2%
Sports	86.4%	12%	67.9%	86.4%	76%	87.2%
Weighted Avg. ➡	62.1%	9%	62.2%	62.1%	60.6%	82.2%

Table 12: Confusion Matrix for DecisionTable for Test mode: 10-fold cross-validation

Classified as ➡	Education	Business	Criminal	Technology	Politics	Medical	Sports
Education	30	5	2	0	16	0	6
Business	4	75	2	2	20	0	20
Criminal	4	3	48	0	25	0	2
Technology	2	18	2	4	6	0	7
Politics	14	6	3	0	128	0	2
Medical	3	8	0	1	7	15	12
Sports	4	12	1	0	4	1	125

Table 13: Table showing True Positive and False Positive Rate of DecisionTable for Test mode: 10-fold cross-validation

Class ⬇	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Education	50.8%	5.3%	49.2%	50.8%	50%	78.3%
Business	61%	9.9%	59.1%	61%	60%	81.1%
Criminal	58.5%	1.8%	82.8%	58.5%	68.6%	88.9%
Technology	10.3%	0.5%	57.1%	10.3%	17.4%	63.4%
Politics	83.7%	15.7%	62.1%	83.7%	71.3%	90.5%
Medical	32.6%	0.2%	93.8%	32.6%	48.4%	68.7%
Sports	85%	9.8%	71.8%	85%	77.9%	88.2%
Weighted Avg. ➡	65.5%	8.5%	67.1%	65.5%	63.5%	83.7%

VII. CONCLUSIONS

This paper has designed a model which will help to categorize the classifier Conjunctive Rule & Decision Table from rules from WEKA in different test mode (i) evaluate on training data, (ii) 5-fold cross-validation and (iii)10-fold cross-validation in the context of dataset of Indian news

As per the previous performance analysis discussion identification of news from dynamic resources can be done with the propose models. As a result it is found that Decision Table algorithm performs well in categorizing the News for all three Test modes but it is not giving 100% accuracy.

Conjunctive Rule algorithm perform well in categorizing in the news related to news category Sports & Politics. Hence Decision Table is good classifier as compare to Conjunctive Rule classifier.

REFERENCES

[1]. Sushilkumar Rameshpant Kalmegh, Analysis of Classification Method "VFI" from WEKA by Using Different Test Mode, *IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 4, April 2016*, 258-265.

[2]. Suril Shah, Sagar Vikmani, Sahil Modak, Prof. Kiran Bhowmick, Evaluating Effectiveness of Classification Algorithms on Personality Prediction Dataset, *International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 10, October-2015*, 401-405.

[3]. R. Srujana, Dr. G S N Murty, Evaluating the Effectiveness of Classification Algorithms Based on CCI, *International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 9, September 2014*, 15909-15916.

[4]. Sujata Joshi, S. R. Priyanka Shetty, Performance Analysis of Different Classification Methods in Data Mining for Diabetes Dataset Using WEKA Tool, *International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 3, March 2015*, 1168 – 1173.

[5]. S. R. Kalmegh, S. N. Deshmukh, Effective Evaluation of Classification of Indigenous News Using Decision Table and OneR Algorithm, International Journal of Advanced Information Science and Technology (IJAIST), Vol.26, No.26, June 2014, 6-11.

[6]. Geetali Banerji, Kanak Saxena, An Efficient Classification Algorithm for Real Estate domain, *International Journal of Modern Engineering Research (IJMER), Vol.2, Issue.4, July-Aug. 2012*, 2424-2430.

[7]. S. Vijayarani, M. Muthulakshmi, EVALUATING THE EFFICIENCY OF RULE TECHNIQUES FOR FILE CLASSIFICATION, *IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 10, Oct-2013*, 38-42

[8]. Ian H. Witten, Eibe Frank & Mark A. Hall., *Data Mining Practical Machine Learning Tools and Techniques, Third Edition* .(Morgan Kaufmann Publishers is an imprint of Elsevier, 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA, 2011).

[9]. Mohd Fauzi bin Othman,Thomas Moh Shan Yau, Comparison of Different Classification Techniques Using WEKA for Breast Cancer, *IFMBE Proceedings 15, Springer-Verlag* Berlin Heidelberg, 2007, 520-523.

International Journal of Engineering Science Invention (IJESI) is UGC approved Journal with Sl. No. 3822, Journal no. 43302.

Sushilkumar Rameshpant Kalmegh "Comparative Analysis of the WEKA Classifiers Rules Conjunctiverule & Decisiontable on Indian News Dataset by Using Different Test Mode" International Journal of Engineering Science Invention (IJESI), vol. 07, no. 02, 2018, pp. 01–09.