

Prediction Of Consumer Review Analysis Using Naive Bayes And Bayes Net Algorithms

¹R. P. Kannan , ²A. S. Arunachalam

¹Research Scholar, Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamilnadu India. Email: rpkannan1@gmail.com,

²Associate Professor, Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamilnadu India. Email: arunachalam1976@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: Data mining field that aims to bring out knowledge from three different form of Structured and unstructured, Semi structured forms. Classification techniques can be used to classify the large Volume of data and Variety of data. Classification supervised learning approach to data processing and WEKA tool is an effective and efficient tool with many inherent methods for extracting useful information. In this paper using the weka tool to analyze consumer data in the weka tool having the number of algorithm. In that algorithm we are used classification algorithm Bayes Network algorithm, Naive Bayes algorithm. Consumer behavior analysis is important of making decisions in the supermarket, consumer behavior prediction different data include in consumer behaviour analysis explains the all data and also use to identify the hidden relationships of data.

Index Terms: Consumer behaviour, Data Mining, Weka tool, Classification, Naive Bayes, Bayes net.

1.Introduction

Data Mining is a group of process that the applies huge, complex databases. This is to eliminate random data and find the hidden data. These data mining methods are always computational concept. Data mining concepts are used to extract the hidden information or useful most informative data. Data mining concepts is most powerful technique in these techniques to implement the different data to analysis and get different result. In this result helps to taking the decision making process. We use data mining, procedures and concepts to reveal patterns in data. There is many of driving power at present. This is the reason was data mining processing has become such an important study.

WEKA is machine learning tool. In this developed by Java. It is having the set of instructions to implement the DM algorithm and machine learning algorithm. Directly put in to the data or java code using WEKA tool. WEKA is a group of tools for: Regression, Clustering, Associations, Data pre-processing, Classifications and Visualizations.

1.1 Shopping Data set

Consumer actions is different based on action data are stored and use to analysis in this data information are categories in several criteria's age, income, budget, needs, products this are all in the dataset .in this paper all dataset how it is used and also how to implement the algorithm, which algorithm is give the best result .

Attributes	Description
Personal detail	Consumer Name, Age, Gender, City
Consumer income details	Categories to low level (income 10000), middle level (income 50000), high level (income 100000 above)
Budget	Categories to different budget levels
Price	Describes the price level
Brand	Categories to different brand like the consumer
Items	What the items purchased consumer
Payment mode	Payment mode on cash, card, net bank, or other payment mode
Number of super market	Consumer like which super market
Consumer satisfaction	Based on consumer satisfaction

Table 1: List of Various Dataset Descriptions

Based on the data set classification to predict the consumer behaviours in the all attributes give the current scenario of the market. Research of analysis different category one is who buy the goods, why, what, where in this all category analysis used to market prediction, Most important process in analysis of consumer patterns estimates the market data[1].

1.2 Classification

First introduce the sample data model to obtain the best limit conditions use the stimulate every selected group of data. Only one limited Constance has been determining, the coming up assignment is to measurement the selected groups of data. Total progress is called classification techniques. Two types of learning one is supervised

learning and second is unsupervised learning .Classification techniques one of the supervised learning. The classification technique used to classify the data by implementing tested data.

Selected group of data examples:

- Predicating the consumer data either he will purchase food item (selected group of data: Yes or No)
- Classify the food item from characteristic like colour, taste, size, weight (selected group of data: dhal, oil, milk, and coffee)
- categorize from gender (selected group of data : Male or Female)

The selected group of data classification involves following process:

- Implementing the testeddata file.
- Associate the categories labels and groups.
- Classified the functional elements to the located.
- Sample data and Tested data used by learning method.
- Used to identify the model and to categorize the hidden data set[2].

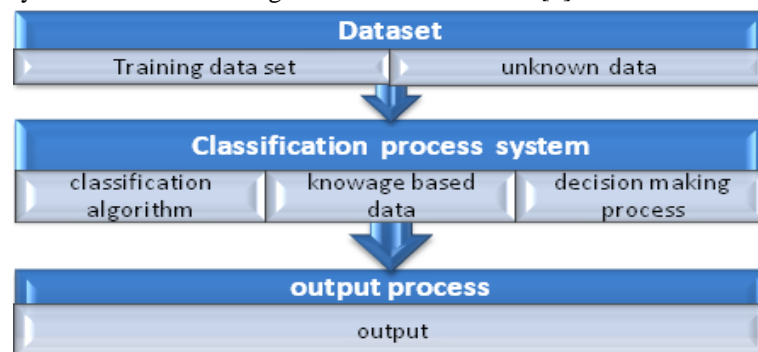


Fig 1: Classification process system

This paper summarizes the Consumer based analysis has different approach. In this survey section 2 specifies various related works. Section 3 gives the information for type of classification algorithms. Section 4 implies the Execution methods section 5 deal results and discussions Consumer behaviors and section 6 explain the conclusions.

2.Related works

Shaffy goyal,Namisha modi using different classification algorithm and compare the Naïve Base algorithm identified probability of two class label[3].Dr.anil sharma and Balrajpreet kaur to compare the different classification algorithm like KNN,Naïve Base, decision tree in that three algorithm give the best result naïve base only.[4] Nagaraju Orsu Gopala Krishna Murthy Nookala , Suresh B. Mudunuri ,Bharath Kumar Pottumuthu, using different classification algorithm and different data all classification algorithm and data is give the different result[5].Abdul Hamid M. Ragab, Amin Y. Noaman, Abdullah S. AL-Ghamdi, Ayman I. Madbouly, in the all authors to compare the different likeC4.5 algorithms, Random Forest algorithm , IBK-E algorithms, LibSupport Vector machine algorithms , MLP algorithms, Multilayer Perceptron algorithms , Naïve Bayes algorithms , and PART algorithms . Comparative study and analysis related to classification predict included Recall, Precision, F-Measure, Matthews’s Correlation Coefficient (MCC), Precision-Recall Curve (PRC), ROC curve, FP-rate, and TP-rate have been analysis by simulation using Weka Toolkit. Experimental Results given that C4.5 gives the best performance and accuracy and lowest absolute errors, then PART, Random Forest, Multilayer Perceptron, and Naïve Bayes, respectively[6].Yugal kumar and G. Sahoo compare the all algorithms BayesNet, Naïve Bayes, Naïve Bayes Up table, Multilayerperceptron, Voted perceptron and J48 But, it is not easy to say which one is the best.mean absolute value is false rate of J48 is small cancer data for breast i.e. small data and large data from the table one, seen give the action of the J48 classification is best compare to another classification technique[7]. PrudhviRaj G,Arka Halder , Lakshmi S.V.S.S, Arka Halder ,Prudhvi Raj G, Lakshmi S.V.S.S, all authors make to compare this three algorithm decision tree ,naïve bayes,zero R using weka tool so that final result is naivebayes algorithm is best one running time is very low and accurate value is given[8]. Md. Faisal Kabir, Alamgir Hossain, Keshav Dahal in these three authors to identify the classification algorithm is the best one of the algorithm[9].

3.Classification Algorithms

A classification algorithm in common, is a role that weighs the input value so that the output separates one class into positive data values and the other into negative data values. Linear Classification algorithms, Logistic regression classification algorithms, Naive Bayes classification algorithm, Fisher’s linear discriminate classification algorithms, Support vector machines classification algorithm, Least squares support vector machines classification algorithms, Quadratic classification algorithm , Kernel estimation classification algorithm , k-nearest neighbor classification algorithm , Decision trees classification algorithm, Random forests classification algorithm, Neural networks, Learning vector quantization different classification algorithms

3.2 Naïve Bayes algorithm

One of the "supervised learning "algorithms is naïve bayes algorithm .it is give the both input data and output classified data's (i.e., the data having the right input and output for every point).Naïve Bayes classification algorithms based on Bayes' theorem which is powerful to the predicted the variables . Naïve Bayes algorithm is to is classified the group of data items to efficient and, correct, fast. It is more accept in different group of data prediction analysis .When we assume of non dependence data variables is handle , a Naive Bayes algorithm to perform the good compare to other model like regression analysis. It is good performinggive to the different input data compare the numeric value of variable in data, for numericvalue of distribution is predicted. In the assumption is they have now appear of character isindependent to the current of any other character in the adjusted class. For example, a food item may be considered as different quality of food if it is having quality of food and no quality of food . all thoseproperty are independent or dependent of each other, they contribution of independent to thevariable probability distribution of that the food is a quality or no quality. Bayesian classification, the main aim is to find the posterior probabilities i.e. the probability of a label given some observed features, $P (B | characteristic)$. With the help of Bayes theorem, we want to express this in unit form as follows [12].

$$P(B|char) = \frac{P(B)P(characteristic|B)}{P(characteristic)}$$

$P(B|char) = (P(B)P(characteristic|B))/(P(characteristic))$

Here, $(B | characteristic)$ - posterior probability.

$P(B)$ -prior probability .

$P(characteristic |B)$ - probability prediction class.

$P(characteristic)$ - prior probability of prediction.

In the given example of the data is climate and equal of given class variable possibility is playing.In there we want to identify the players playing or not playing based on climate condition. See following pace to performingit.fitstransfer data into transaction table second Create table to calculate the probability like cloudy probability = 0.29 and playing probability is 0.64.and next we use Naive Bayes theorem equation to the posterior probability for everyclass. The class with the top level posterior probability is the output come of prediction

Problem: Players will play if climate in sunny. In this statement is true?

Table.3: Weather

climate	play
Sunny	No
Cloudy	Yes
Rain	Yes
Sunny	Yes
Sunny	Yes
Cloudy	Yes
Rain	No
Rain	No
Sunny	Yes
Rain	Yes
Sunny	No
Cloudy	Yes
Cloudy	Yes
Rain	No

Transaction table		
climate	Yes	No
Sunny	3	2
Rain	2	3
Cloudy	4	-
Total	9	5

Table.4: Transaction

Potentiality table		
Climate	Yes	No
Sunny	3	2
Rain	2	3
Cloudy	4	-
Total	9	5
Total no yes& no	=9/14	=5/14
Total no yes& no	0.64	0.34

Table.5: Potentiality

$$\text{Total no of Sunny} = 5/14 = 0.36$$

$$\text{Total no of rain} = 5/14 = 0.36$$

$$\text{Total no of cloudy} = 4/14 = 0.26$$

We apply the formula of posterior probability.

$P(\text{Yes OF Sunny}) = P(\text{Sunny OF Yes}) * P(\text{Yes}) / P(\text{Sunny})$, Here $P(\text{Sunny OF Yes}) = 3/9 = 0.33$, $P(\text{Sunny}) = 5/14 = 0.36$, $P(\text{Yes}) = 9/14 = 0.64$, Now $(\text{Yes OF Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$, which has higher one probability ratio.

Naïve Bayes use the equal method of predicting the probability ratio of different class an entity. In that algorithm frequently used in classification of data [13].

4.Experimental Setup

In this experimental setup, using the weka tool different analysis done by the classification algorithm. Model of WEKA has been implemented to the executing classification algorithms using the consumer Data for csv file format and arff file format also used. When startup Weka software, Explorer and experiment is the first and second display the various menu display Weka. First select Explorer menu the aim of analysis the Consumer data in this is analysis in two process one Preprocess and second one is Classify the data

4.1. Preprocess

Starting Weka Explorer, first we selected and loaded the data after that filtering the data by using supervised attribute selection option. Before the filtering, 16 attributes have selected for classification and after the filtering identify the 8 attributed only selected. In this step is basic step for all classification algorithms. How the data set to select and loading, filter into weka tool. Select the Classify option in the first step of Explorer by choosing the Bayesian method of classification algorithm Bayes Network are selected the ten times multiple Confirmation of experiment method setup the classified. If you want change the fold cross change it. using classification for different data to showing in explanation of execution method selecting and set up the classify data has been show the Fig.3,4 For(eg Gender ,review)[14].

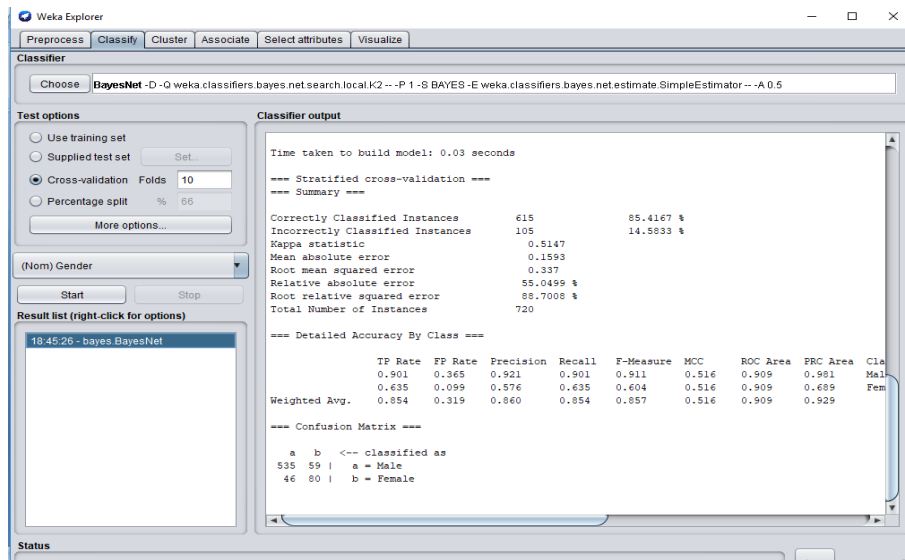


Fig.3: Bayesian Network Classifier algorithm and Result for Gender

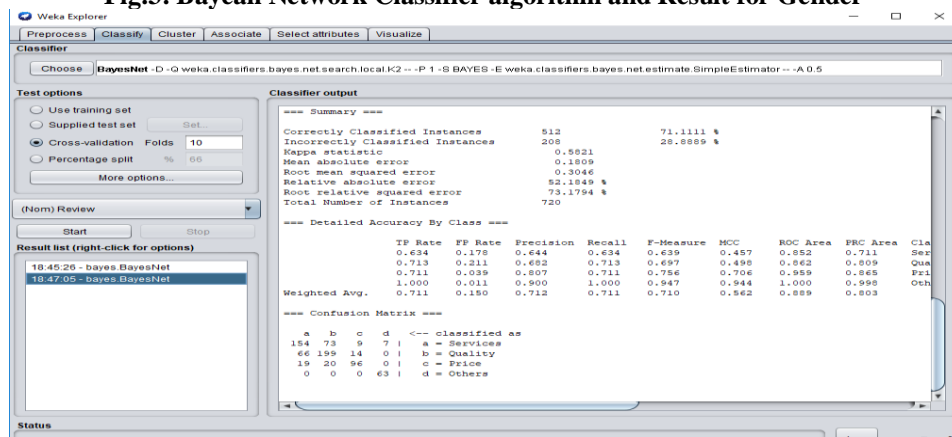


Fig.4: Bayesian Network Classifier algorithm and Result for review

The weka Tool, Select the Classify option in the first step of Explorer by choosing the Bayesian method of classification and Bayes Network algorithm and selected the ten times multiple Confirmation of experiment menu and set up the classified data. If you want to change the fold cross change it using classification for different data to show in explanation of execution method selecting and set up the classify data has been shown in Fig.5,6 For (eg Gender,review)[15].

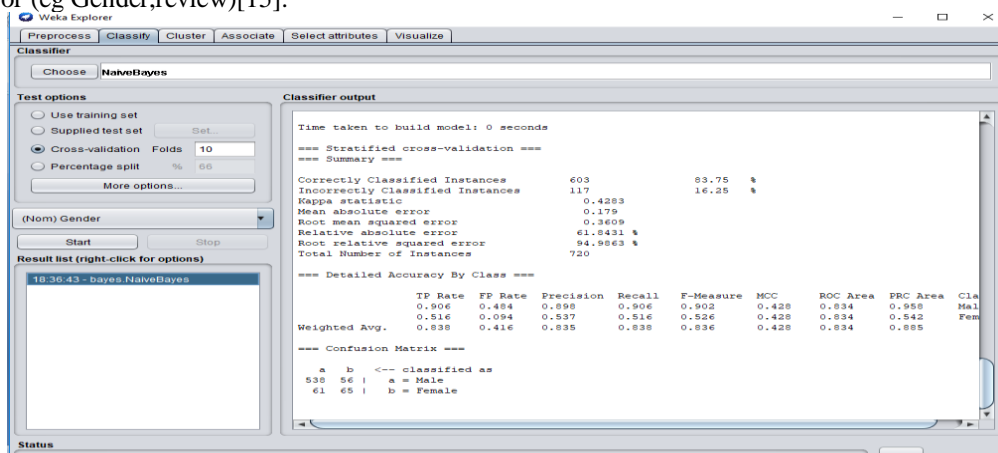


Fig.5: Naive Bayes Classifier algorithm and Result for Gender

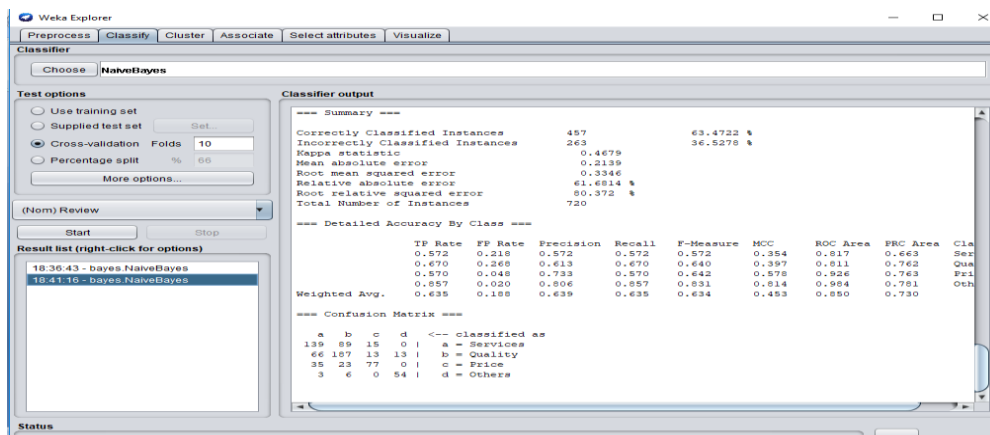


Fig.6: Naïve Bayes Classifier algorithm and Result for Review

5. Result Analysis and Discussion

In this paper we used consumer data set like Consumer Name, Age, Gender, City, Categories to low level(income10000), middle level (income50000), high level(income100000 above),Categories to different budget levels Describes the price level, Categories to different brand like the consumer, What the items purchased consumer, Payment mode on cash, card, net bank, or other payment mode Consumer like which super market ,Based on consumer satisfaction based on review in mainly two data is classify and compare Naïve Bayes and Bayean Networkdiscuss in details [16].

5.1 Method for Evaluation

The conduction is applied in the method of evaluation:

1. ConfusionMatrix: it is the used to analysis how predict to your data classifier can recognize attribute of different attribute class member. In this matrix an M*M, where M is the number of attribute class memberbeing calculated. Forthe issue in our hand, we have M equal to 2and M equal to 4, and hence we get a 2*2 and 4*4 matrix.
2. Kappa: Measurements theconnection link together categorizes in the data elements and true attribute. In the value intermediate of (ZERO,ONE). The mean value of ONE is correct value and ZERO means dynamic value.
3. True Positive: in this measurement that give the correct classification attribute.
4. False Positive: in this detail data of inaccurate instances labels as correct instances.
5. Recall Measurement: Measurement of recall in the rate of all relevant datasets provided by the classifier. Highest –level of recall is a model of a data set that provides relevant data.
6. Precision Calculation: Measurement of the precision calculation exactingthe applicable knowledge data is gathering. High level precision model return the more applicableattribute and inapplicable attribute [17].

Name of algorithms	Used for classified data	Correctly classified data		Incorrectly classified data	
		Number of data	Percentage (%)	Numbers of data	Percentage (%)
Bayes Net	Gender	615	85.4167	105	14.5833
Naïve Bayes	Gender	603	83.75	117	16.25
Bayes Net	Review	512	71.1111	208	28.8889
Naïve Bayes	Review	457	63.4722	263	36.5278

Table.6: comparison of Accuracy for Bayes classification Algorithms

5.2. Considerations of study

The consumer dataset categorized using classification techniques. It is implemented that Bayes Networks appear in the sequence of right categorized occurrence in the instances based on gender 615 and based on review 512 of with the accuracy of percentage 85.45% and 71.11%. The number of incorrectly classified instances based on gender 105 and based on review 208 that is 14.58% and 28.89%. Naïve Bayesian classified indicate the number of right class instances based on gender 603 and based on review 457 with the accuracy of

gender 83.75%,and based on review accuracy 63.47% and the number of incorrectly classified instances based on gender 117 and based on review263 of that is incorrect percentage based on gander and review16.25%, 36.53% .Table one indicate in correct the different BayesianClassification .Table two display the last calculation like False Positive Rate calculation, True Positive Rate measurements , calculation ofPrecision, measurement ofRecall, calculation of F-Measure, measurement of MCC area, measurementROC Area of all the Bayes Net algorithm and the Naïve Bayes algorithm in the terms of two classes of based on gender is male and female, based on reviewservices, quality, price, others. Table three indicates the Comparison based on gender and review of calculation in the Weighted Average different BayesClassification. Table four calculations given the Confusion Matrix based on gender and reviewpair of the Bayesian Network algorithm and the Naïve Bayes algorithm. [18].

Name of algorithms	Used for classified data	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC	PCR	CLASSES
Bayes Net algorithms	Gender	0.901	0.365	0.921	0.901	0.911	0.516	0.909	0.981	Male
		0.635	0.099	0.576	0.635	0.604	0.516	0.909	0.689	Female
	Review	0.634	0.178	0.644	0.634	0.639	0.457	0.852	0.711	Services
		0.713	0.211	0.682	0.713	0.697	0.498	0.862	0.809	Quality
		0.711	0.039	0.807	0.711	0.756	0.706	0.959	0.865	Price
		1.000	0.011	0.900	1.000	0.947	0.944	1.000	0.998	Others
Naïve Bayes algorithms	Gender	0.906	0.484	0.898	0.906	0.428	0.428	0.834	0.958	Male
		0.516	0.094	0.537	0.516	0.428	0.428	0.834	0.542	Female
	Review	0.572	0.218	0.572	0.572	0.572	0.354	0.817	0.633	Services
		0.670	0.268	0.613	0.670	0.670	0.397	0.811	0.762	Quality
		0.570	0.048	0.733	0.570	0.570	0.578	0.926	0.763	Price
		0.857	0.020	0.806	0.857	0.857	0.814	0.984	0.781	Others

Table.7: Bayes classification AlgorithmsFinal Statistics

Different Measurements	Name of classification algorithms			
	Bayes Net algorithms		Naïve Bayes algorithms	
	Gender	Review	Gender	Review
TP Rate	0.854	0.711	0.838	0.635
FP Rate	0.319	0.150	0.416	0.188
Precision	0.860	0.712	0.835	0.639
Recall	0.854	0.711	0.838	0.635
F-Measure	0.857	0.710	0.836	0.634
MCC	0.516	0.562	0.428	0.453
ROC	0.909	0.889	0.834	0.850
PCR	0.929	0.803	0.885	0.730

Table.8: Comparison of Weighted Average.

Name of classification algorithms	Gender attribute			Classified data
	a	b	variables	
Bayes Net algorithms	535	59	a	Male
	46	80	b	Female
Naïve Bayes algorithms	a	b	variables	Classified data
	538	58	a	Male
	61	65	b	Female

Table .9: Confusion Matrix for Gender

Name of classification algorithms	Gender attribute					Classified data
	a	b	c	d	variables	
Bayes Net algorithms	154	73	9	7	a	Services
	66	119	14	0	b	Quality
	19	20	96	0	c	Price
	0	0	0	63	d	Others
Naive Bayes algorithms	139	89	15	0	a	Services
	66	187	13	13	b	Quality
	35	23	77	0	c	Price
	3	6	0	54	d	Others

Table.10: Confusion Matrix for Review

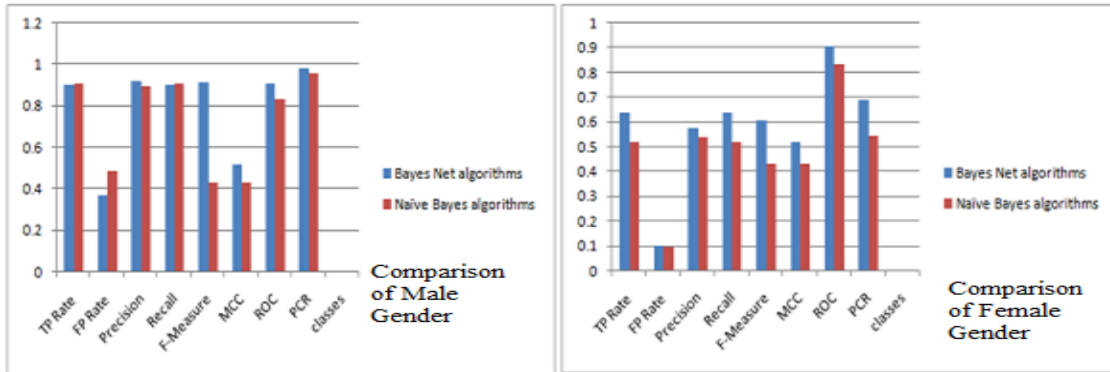


Fig.7: Comparison of Final Statistics Bayes classification Male and Female

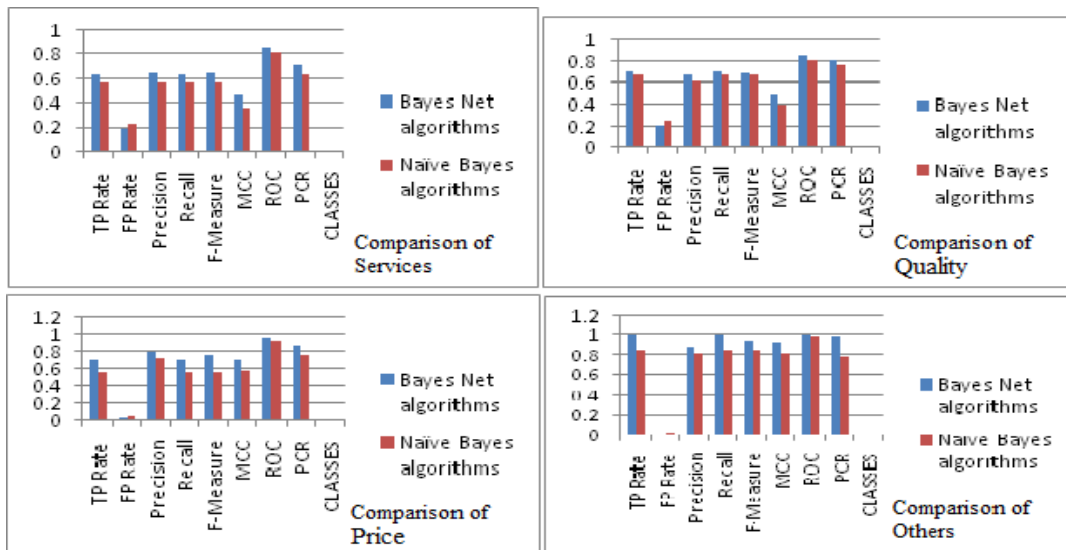


Fig.8: Comparison of Final Statistics Bayes classification Services, Quality, Price, Others

6. Conclusion

Classification approach is data mining approach it is supervised learning algorithm. It identified the categorized data, so that the classification of the pre-determined classes. Data classification problem is one of the most important researches in data mining , in which the classification of data of interest is the minimum classification, and has very low level sample data compared to the major classes. In this leads classifier prediction to be based on towards majority class, so solutions needs to be found out to handle this problem. Here we have evaluated solutions to class imbalance problem on consumer behaviour dataset using WEKA.In this paper compare the two classification algorithm. In this analysis to identify that Bayesian Network algorithm create the huge amount of correct classified data compared to the Naïve Bayes algorithm.

References

1. Rana Alaa El-Deen Ahmed, M.Elemam.Shehab, Shereen Morsy, Nermeen Mekawie,” Performance study of classification algorithms for consumer online shopping attitudes and behavior using data mining”,2015.

2. Kareena, Raj Kumar," A Consumer Behavior Prediction Method for E-Commerce Application",Volume.8(2S6), 2019, ISSN NO 2277-3878
3. Shaffy Goyal, Namisha Modi, "A Review of Various Classification Algorithms for Online Shopping Data", Volume 6(2),pp.2250-1797,2016.
4. Dr.Anil Sharma,Balrajpreet Kaur,"A Research Review on Comparative Analysis of Data Mining Tools, Techniques and Parameters", Volume .8(7), 2017, ISSN NO .0976-5697.
5. Gopala Krishna Murthy Nookala, Bharath Kumar Pottumuthu , Nagaraju Orsu, Suresh B. Mudunuri," Performance Analysis and Evaluation of Different Data Mining Algorithms used for Cancer Classification" , Volume 2(5), 2013.
6. Abdul Hamid M. Ragab, Abdullah S. AL-Ghamdi ,Amin Y. Noaman ,Ayman I. Madbouly," A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining".
7. Yugal kum, G. Sahoo," Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA ",Volume.05, pp. 359–369, 2012.
8. Arka Haldar , G.Prudhvi Raj , S.V.S.S Lakshmi."Comparison of Different ClassificationTechniques Using WEKA for DiabeticDiagnosis", Volume. 6(1), 2018, ISSNNO. 2320-9801.
9. Md. Faisal Kabir &Chowdhury Mofizur Rahman,Alamgir Hossain,Keshav Dahal" ,Enhanced Classification Accuracy on Naive Bayes DataMining Models", Volume 28(3),pp .0975 – 8888,2011.
10. Murat Koklu, Yavuz Unal," Analysis of a Population of Diabetic Patients Databases with Classifiers", Volume.7(8), 2013, ISNI NO.0000000091950263.
11. Md. Nurul Amin, Md. Ahsan Habib," Comparison of Different Classification Techniques Using WEKA for Hematological Data", Volume.4(3), pp.55-61,2015, ISSN NO. 2320-0847.
12. N.G.Sree Devi, M.Jeyanthi," Comparative Analysis Of Classification Algorithm Using MachineLearning Technique", Volume. 6 (2),2019, ISSN NO. 2349-5162.
13. Sourabh Shastri, Paramjit Kour, Ankush Gupta, Shakshi Sambyal, Arun Singh Bhadwal, Amardeep Sharma, Professor Vibhakar Mansotra, Dr. Anand Sharma," Development Of A DataMining Based Model For Classification Of Child Immunization Data", Volume. 08 (6), 2018, ISSN NO.2250 – 3005.
14. V. Vaithyanathan, K. Rajeswari, Kapil Tajane, Rahul Pitale," Comparison Of Different Classification Techniques Using Different Datasets", Volume.6 (2), Pp. 764-768, ISSN NO. 2231-1963.
15. Kaushik H. Raviya ,Biren Gajjar," Performance Evaluation of Different Data Mining Classification Algorithm Using WEKA", Volume. 2(1), 2013, ISSN NO. 2250-1991
16. Mohd Fauzi bin Othman,Thomas Moh Shan Yau," Comparison of Different Classification Techniques Using WEKA for Breast Cancer", Volume. 15, pp. 520-523, 2007.
17. Shivangi Gupta ,Neeta Verma," Comparative Analysis of classification Algorithms using WEKA tool", Volume. 7(8), 2016, ISSN NO. 2229-5518.
18. M. Purnachary, B. Srinivasa S P Kumar, Humera Shaziya," Performance Analysis of Bayes Classification Algorithms in WEKA Tool using Bank Marketing Dataset", Volume. 5(2), 2018, ISSN NO. 2394-2320. .
19. A.S. Arunachalam, T.Velmurugan. "A Survey on Educational Data Mining Techniques." International Journal of Data Mining Techniques and Applications 5.2 (2016): 167-171.
20. S. Perumal,T.Velmurugan, "Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization",International Journal of Engineering & Technology,7 (2.26) (2018) 74-79.
21. Arunarani, S., Gobinath, R. "A relative analysis of multimodal biometric fusing face, ear and fingerprint"International Journal of Scientific and Technology Research, 2020, 9(4), pp. 1996–2002.