International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 12 Issue 2, February-2023, Impact Factor: 7.957

Supermarket Classifying and Clustering Tb Racks Dataset Using Concepts from Data Mining Techniques

Mohini Dhummerkar[1], Dr. Neelesh Jain[2], Dr. Neeraj Gupta[3]

[1]M. Tech Research Scholar, Department Computer Science SAM College of Engineering and Technology
[2,3]Professor, Department Computer Science SAM College of Engineering and Technology

ABSTRACT

The objective of data mining is to extract appealing correlated information from bulky databases. This study proposal aims to comprehend the fundamental idea behind data mining technologies used in TB rack-based supermarket analysis. The Top to Bottom TB Racks Ratio (TBR)-based clustering technique is provided in a way that clarifies the use of data mining in supermarket analysis. The set of items that shoppers purchased at the grocery store were analyzed by the author using data mining software called Weka 3.8.6 for the TB racks application. The author attempted to relate the experiment and algorithm in this study. The author then presented the results by demonstrating a TB rack analysis application. The statistical outcome from Weka. The number of algorithms in the weka tool is used in this paper to analyze consumer data. We utilized the Random Tree algorithm and the BayesNet algorithm for classification in that algorithm. In the supermarket, consumer behavior analysis is important for making decisions because it can predict consumer behavior based on a variety of data. Consumer behavior analysis can also be used to find hidden relationships between data.

Keyboard:-Supermarket, BayesNet, Randoom Tree, TB racks, Data mining, Weka.

INTRODUCTION

Companies are now able to collect massive amounts of data thanks to our highly technological age. Most organizations of businesses have amassed tens of thousands to hundreds of millions of pieces of data that, if not converted into useful information, have no value[1]. The technology known as data mining is a tool that can be used by businesses to extract data from large databases. Knowledge discovery is a broader process that includes data mining.

A lot of people think of data mining as just one part of a larger process called Knowledge Discovery in Databases (KDD).As per Fayyad et.al, 'KDD is the nontrivial cycle of recognizing substantial, novel, possibly helpful and at last justifiable examples in information[6].'

RELATED WORK

Trnka A. et al 2010[1] presented how to apply Six Sigma methodology to Market Basket Analysis. Information Mining techniques give a ton of chances on the lookout area. One of them is a basket market analysis. Numerous statistical techniques are utilized in the Six Sigma methodology. We can alter the process's Sigma performance level and improve outcomes by incorporating Market Basket Analysis into one of Six Sigma's phases as part of Data Mining.

Chenyang M. et al 2019[2] Describe Nowadays, many people use data mining to find connections between items in huge datasets. Frequent itemset mining is an essential component of association rule mining, one of the most popular data mining techniques. The trustworthy but curious cloud service provider (CSP) receives large amounts of data. As a result, the CSP and third parties must be prevented from obtaining the raw data. In addition, supermarket transactions are too few to be mined using the same techniques as the majority of the other data. If these methods are applied to this particular dataset, they will require more computational power than they would for standard data. Under the encrypted mining query on supermarket transactions, we present an effective protocol to determine whether an item set is frequent. We develop a blocking algorithm to enhance mining efficiency. This

algorithm reduces the mining process's computation cost by separating encrypted transactions into blocks and only calculating bilinear pairings on ciphertexts of part blocks rather than all ciphertexts. Finally, we conduct theoretical analyses and simulator experiments to assess the efficiency of our protocol in terms of correctness, running time, cost of computation, and security. Our protocol clearly outperforms the previous solution in terms of efficiency while maintaining the same level of security, as demonstrated by the results.

Riccardo G. et al 2019 [3] Explains nowadays, offering personalized services to customers is a major challenge for supermarket chains. One of these services is market basket prediction, which provides customers with a shopping list for their next purchase based on their current requirements. The various factors that influence a customer's decision-making process cannot be simultaneously captured by current methods: co-occurrence, regularity, and recurrence of the purchased goods. We define a Temporal Annotated Recurring Sequence (TARS) pattern that can simultaneously and adaptively capture all of these factors to achieve this goal. TARS Based Predictor (TBP) is a predictor for the next basket that, in addition to TARS, is able to comprehend the level of the customer's stocks and recommend the set of the most essential items. We also define the method for extracting TARS. Supermarket chains could effectively expedite their customers' shopping sessions by implementing the TBP, which would allow them to create individualized recommendations for each customer. Extensive testing demonstrates that TBP outperforms the most recent competitors and that TARS is capable of explaining customer purchase behavior.

RESEARCH PROBLEM STATEMENT

The majority of businesses today do not recognize the significance of data mining techniques for the organizations' benefit. The study of shopping baskets has grown in popularity among retailers in recent years. They were able to collect data on their clients and their purchases thanks to cutting-edge technology. The use and application of transactional data in supermarket analysis increased with the introduction of electronic point-in sales[4]. Analyzing this kind of data is extremely helpful in retail businesses for comprehending buyer behavior. Mining buying designs permits retailers to change advancements, and store settings and serve clients better.

This is probably because hundreds of organizational-related software and tools have appeared in supermarkets, causing many corporate employees to become confused. As a result, the purpose of this study would be to investigate the significance of data mining to the organization, both directly and indirectly.

METHODOLOGY

Quantitative, qualitative, and demand research are the three main types of research strategies. Both experimental and non-experimental types of research are possible. The purpose of this subchapter is to investigate the TBR-based method of clustering. A set of data items will be divided into appropriate groups using the clustering algorithm. A collection of transactions serves as the data representation in the TB racks-based supermarket analysis. Product codes are represented by rows and columns in this dataset[5]. A “Yes/No” value indicates whether that product was purchased during that transaction in each cell.

Purpose of Research

The purpose of the proposed study is to examine how shopping patterns are related to rack layout and promotion in a sample supermarket store by analyzing customer purchases. The purpose of this study is to determine the outcomes of implementing TB racks analysis in a supermarket. The project's goals are as follows:

- ❖ To investigate the concept of the clustering algorithm in order to investigate the fundamental idea of data mining technology.
- ❖ To use a data mining tool known as weka 3.8.6 to carry out the application of TB racks–based supermarket analysis to discover hidden patterns among various supermarket products.
- ❖ The TB racks experiment aims to determine which of these products sells well together so that when a customer goes grocery shopping, the related products can be arranged together to increase the likelihood of a sale.

Research Methodology

There are a variety of relevant and utilized research methods in information systems research. Problems that haven't been studied before can benefit from exploratory research. It will aid in problem definition and comprehension. There will not be conclusive outcomes using this method. In this instance, the research will begin with a broad idea, and the findings can be applied to subsequent studies

DATA ANALYSIS & RESULT

Data Collection and attribute Selection

In this progression just those areas were selected which were required for data mining. A couple of determined factors were chosen. While a portion of the data for the factors was removed from the database. All the indicator

and reaction factors which were gotten from the database are given in Table I for reference. The data-set considered consists of 268 tuples and 13 attributes [15]. Each tuple represents the attribute values of a TB rack based ration data in depend on item type and itemset. Customer is consuming the product in purchase in supermarket. It describes the details of supermarket in item in retail or supermarket store of sale performance and social behavior.

Table 1: Item set description and parameter Value

Attribute	Description	Parameter Value
Customer_ID	Customer unique ID	Apla-Numeric
Customer_type	Customer Behaviour	Member, Normal
TID	Transition ID	Apla-Numeric
Gender	Customer Gender	Male, Female
Rack_No	Safe Number Location	Apla-Numeric (R1,R2,R3)
Rack_Type	Safe position	RT, RB
I1	Item Type	Yes ,No
I2	Item Type	Yes , No
I3	Item Type	Yes , No
I4	Item Type	Yes , No
I5	Item Type	Yes ,No
I6	Item Type	Yes , No

Working With Weka Tool

The graphical representation of the result obtained from the execution of a particular algorithm over the aforementioned data set is shown in below Figures 5.1 and 5.2. Every one of the diagrams is gotten for every calculation and the outcomes are deciphered. The graphs here not only represent some data but also assist in determining the built-in classifier's efficiency and prediction accuracy for each tool.

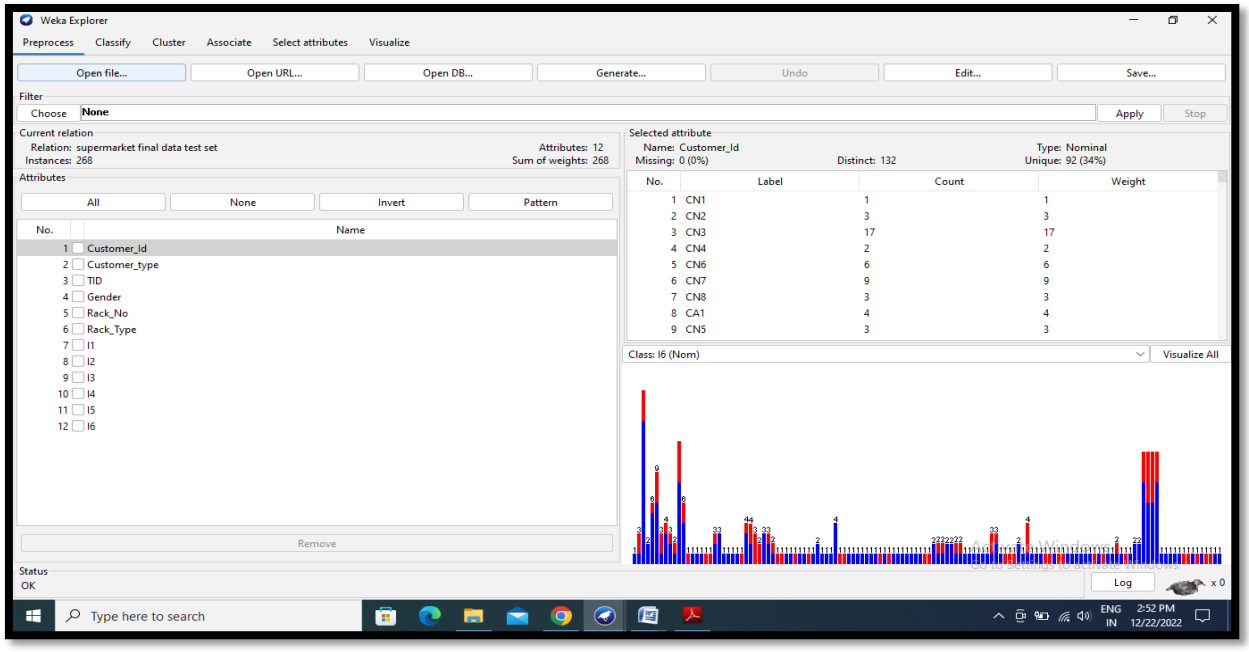

Figure 1:- Weka supermarket item set

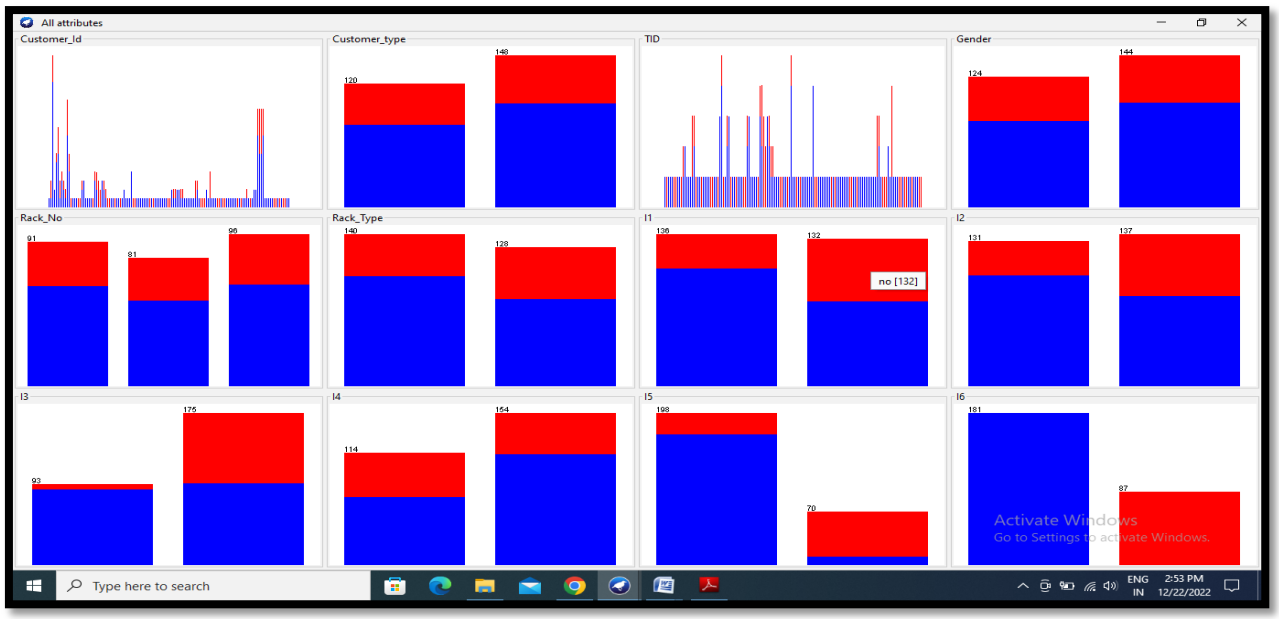

Figure 2 :-Weka tool 13 attribute visualize interface

Result Observation

In this proposed research we used consumer item set like Costumer ID, Customer Type, TID (transition ID), Gender, Rack Location, Rack Type, Item Name like as I1, I2,I3,I4, I5, and I6 Categories to 'Yes' level and 'No' level Categories to different budget levels Describes the customer resonance to different brand like the consume, What the items purchased customer, Payment mode on cash, like which super market, Based on customer satisfaction based on review in mainly two data is classify and compare BayesNet and Random tree discuss in details .

Aspects of research

Classification methods are used to classify customer datasets. It is implemented that BayesNet appears in the sequence of right categorized occurrence in the instances based on rack_loacation 218 and based on rack_type 220 of with accuracy of percentage 81.34% and 82.09%. The number of incorrectly classified instances based on rack_loacation 50 and based on rack_type 48 is 18.65% and 17.91%. Random Tree classified indicates the number of right class instances based on rack_loacation 265 and based on rack_type 267 with the accuracy of rack_loacation 98.88%, and based on rack_type accuracy 99.63% and the number of incorrectly classified instances based on rack_loacation 3 and based on rack_type 1 of that is incorrect percentage based on rack_loacation 1.11% and based on rack_type 0.37%. Table 6.2 indicate in correct the different BayesNet nad Random tree Classification.

Table 2: Comparison of Accuracy for BayesNet Vs Random tree classification Algorithms

Name of algorithms	Used for classified data	Correctly classified data		Incorrectly classified	
		Number of data	Percentage (%)	Number of data	Percentage (%)
BayesNet	Rack Location	218	81.34%	50	18.65%
Random Tree	Rack Location	265	98.88%	3	1.11%
BayesNet	Rack Type	220	82.09%	48	17.91%
Random Tree	Rack Type	267	99.63%	1	0.37%

Table 5.3 displays the most recent calculations, such as the False Positive Rate, True Positive Rate, Precision, Recall, F-Measure, MCC area, F-Measurement, and ROC Area of each BayesNet and RandomTree algorithm in terms of three classes based on rack_location: R1, R2, and R3, and two classes based on rack_type: RT (rack top) and RB (rack bottom).

Table 5.3: BayesNet and Random tree classification Algorithms Final Statistics

Name of algorithms	Used for classified data	TP Rate	FP Rate	Precision	Recall	F-measure	MCC	ROC	PCR	Classes

BayesNet algorithm	Rack Location	0.824	0.102	0.806	0.824	0.815	0.719	0.945	0.91	R1
		0.79	0.053	0.865	0.79	0.826	0.757	0.949	0.87	R2
		0.823	0.128	0.782	0.823	0.802	0.688	0.933	0.91	R3
	Rack Type	0.843	0.203	0.819	0.843	0.831	0.641	0.921	0.92	RT
		0.797	0.157	0.823	0.797	0.81	0.641	0.921	0.92	RB
Random tree algorithm	Rack Location	1	0.006	0.989	1	0.995	0.992	1	1	R1
		1	0.011	0.976	1	0.988	0.983	1	1	R2
		0.969	0	1	0.969	0.984	0.976	1	1	R3
	Rack Type	1	0	1	1	1	1	1	1	RT
		1	0	1	1	1	1	1	1	RB

The comparison between Weighted Average different BayesNet and Random Tree Classification is shown in Table 6.4 based on rack_loacation and rack_type of calculation.

Table 4: Comparison of Weighted Average

Different Measurements	BayesNet Algorithms		Random tree	
	Rack_Location	Rack_Type	Rack_	Rack_
TP Rate	0.813	0.821	0.989	1
FP Rate	0.097	0.181	0.005	0
Precision	0.815	0.821	0.989	1
Recall	0.813	0.821	0.989	1
F-measure	0.814	0.821	0.989	1
MCC	0.719	0.641	0.983	1
ROC	0.942	0.921	1	1
PCR	0.898	0.922	0.999	1

Based on the rack_loacation and rack_typepair of the BayesNet algorithm and the Random tree algorithm, Table 5.5(a) and 5.5(b) calculations as the Confusion Matrix.

Table 5(a) Calculations the Confusion Matrix based on Rack Location

Name of Classification algorithms	Rack_location				Classified Data
	a	b	c	Variable	
BayesNet	75	4	12	a	R1
	7	61	10	b	R2
	11	6	79	c	R3
Random Tree	91	0	0	a	R1
	0	81	0	b	R2
	1	2	93	c	R3

Table 5(b) calculations the Confusion Matrix based on Rack type

Name of Classification algorithms	Rack_Type			Classified Data
	a	b	Variable	
BayesNet	118	22	a	RT
	26	102	b	RB
Random Tree	140	0	a	RT
	0	128	b	RB

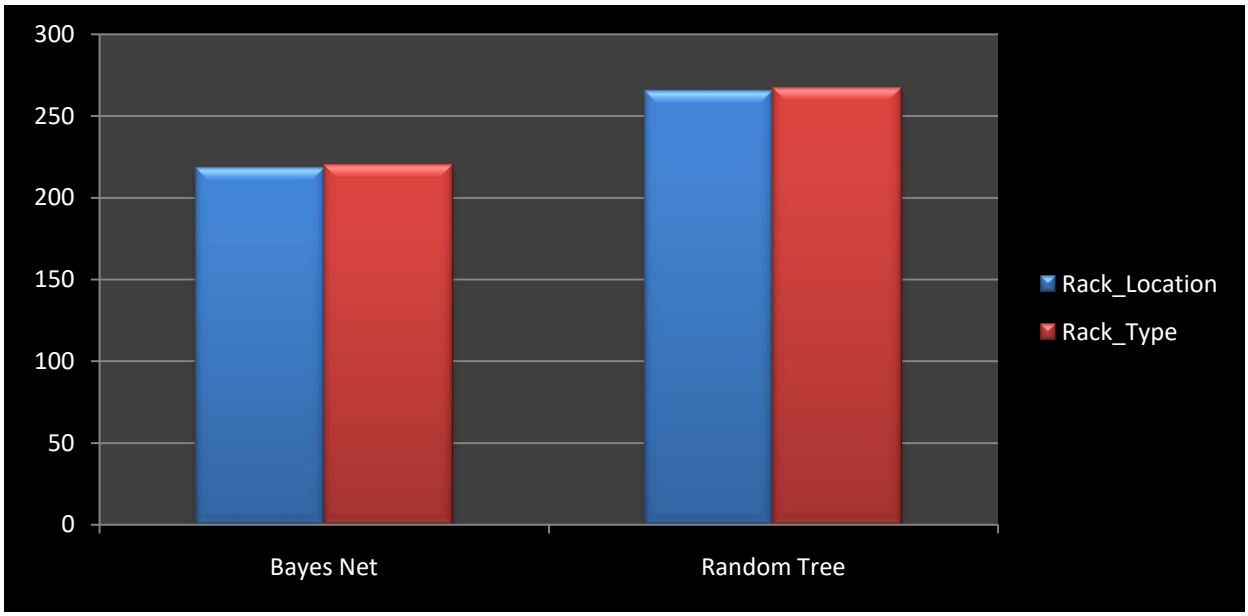

Figure 3:-Comparison of accuracy

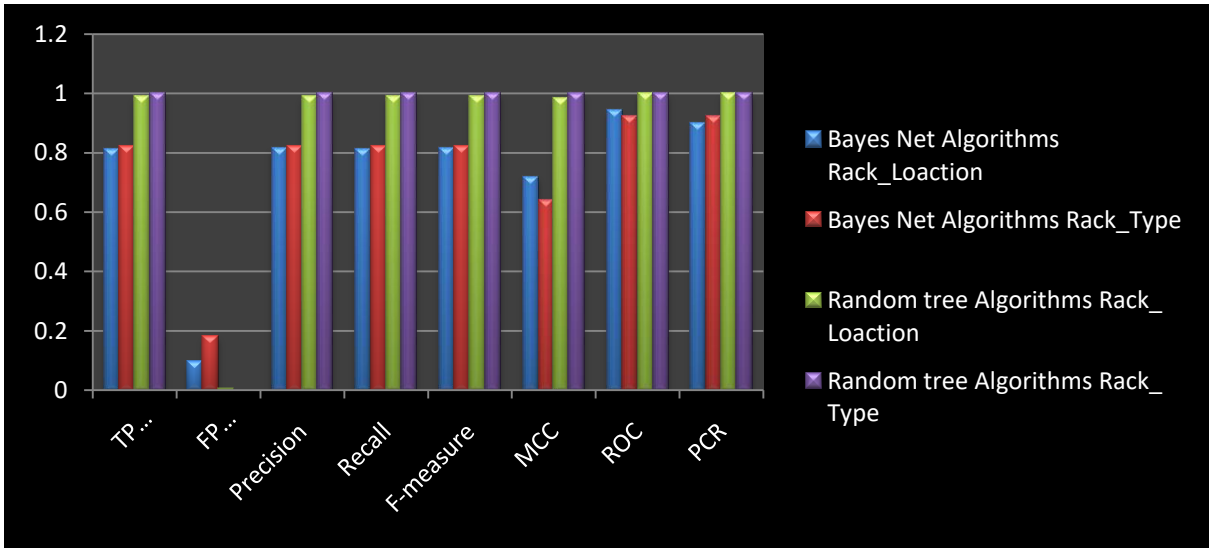

Figure 4:-Comparison Detailed Accuracy By Class

CONCLUSION

There are a number of aspects of retailing that would be difficult without supermarket analysis. Product tracking not only aids in the management and processing of inventory, but it can also be used in cross-sale campaigns and promotional strategies because it provides an overview of co-occurrence products. Marketers and retailers are

informed of a possible influential strategy that can affect the sale of one or both of the products when they know which ones are more likely to be chosen together.

The classification approach is a supervised learning algorithm that is used in data mining. It identified the categorized data, allowing for the predetermined classification. One of the most important studies in data mining is the data classification problem, in which the minimum classification of the data of interest is used and very little rack bottom sample data is used in comparison to rack top classes. Because this causes classifier prediction to be based on the majority class, solutions to this issue must be found. On the consumer behavior dataset, we used WEKA to evaluate solutions to class imbalance issues. In this paper, we contrast the two classification algorithms. This analysis aims to determine that the RandomTree algorithm produces fewer accurate classified data than the BayesNet algorithm.

REFERENCES

[1]. Trnka Andrej, "Market Basket Analysis with Data Mining Methods Six Sigma methodology improvement", International Conference on Networking and Information Technology, pp 446-449, IEEE 2010.
[2]. Chenyang Ma, Baocang Wang , Kyle Jooste, Zhili Zhang , and Yuan Ping, "Practical Privacy-Preserving Frequent Itemset Mining on Supermarket Transactions" IEEE SYSTEMS JOURNAL Personal use is permitted, but republication/redistribution requires IEEE permission, pp-1-11 , IEEE 2019.
[3]. Riccardo Guidotti , Giulio Rossetti , Luca Pappalardo , Fosca Giannotti, and Dino Pedreschi, "Personalized Market Basket Prediction with Temporal Annotated Recurring Sequences", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 31, NO. 11, pp- 2151-2163 November 2019.
[4]. A. M. Khattak, A. M. Khan, Sungyoung Lee and Young-Koo Lee. "Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka", International Journal of Database Theory and Application Vol. 3, No. 1.
[5]. Andreas Mild, Thomas Reutterer, "An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data", Journal of Retailing and Consumer Services, Volume 10, 123-133, 2003.
[6]. David R. Bell and James M. Lattin, "Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket", Marketing Science, Vol. 17, No. 1, 66-88, 2008.
[7]. Kumar, N., & Rao, R., "Using Basket Composition Data for Intelligent Supermarket Pricing" . Marketing Science, 25(2), 188-199,2006.
[8]. Ayinde A. Adetunji A., Bello M., and Odeniyi O., "Presentation evaluation of naive bayes and decision stump algorithm s in mining students educational data.," Interna-tional Journal of Computer Science Issues (IJCSI), vol. 10, no. 4, 2013.
[9]. Joachims T., Freitag D., and Mitchell T., "Web-watcher: A tour guide for the world wide web," in IJCAI (1), Citeseer, 1997, pp. 770–777,.
[10]. Romero C. and Ventura S., "Educational data mining : a review of the state of the art, Systems, Man, and Cybernetics, Part C: Applications and Reviews," IEEE 2010 Transactions on, vol. 40, no. 6, pp. 601–618.
[11]. Mendes R. R., Voznika F. B. de, Freitas , and Nievola J. C., "Discovering fuzzy classification rules with genetic programming and co-evolution," Principles of Data mining and Knowledge Discovery , Springer, 2001, pp. 314–325,.
[12]. Zhang B., Legible Discovering And Readable Chinese Typefaces For Reading Digital Documents. PhD thesis, Concordia University, 2011.