Frequent Item set Mining Using Global Profit Weight Algorithm

ASHA RAJKUMAR M.phil, Reasearch Scholar
Computer Science Department
P.S.G.R.Krishnammal College for Women
Coimbatore, India

G.SOPHIA REENA M.C.A., Mphil.,
HOD, BCA Department
P.S.G.R.Krishnammal College for Women
Coimbatore, India

Abstract— The objective of the study focused on weighted based frequent item set mining. The base paper has proposed multi criteria based frequent item set for weight calculation. Contribution towards this project is to implement the global profit weight measure and test the performance over utility based mining. For this project the data consist of 90 products from automobile shop including unit price, quantity sold and profit margin for transaction set (one month data). Algorithm has been implemented in Visual Basic for visualizing step by step process calculations**.** Supervised machine learning techniques namely Naïve Bayes Decision tree classifier, VFI and IB1 Classifier are used for learning the model. The results of the models are compared and observed that Naïve Bayes performs well. WEKA tool is used to classify the data set and accuracy is calculated.

Keywords- Global Profit Weight Algorithm; Classification Algorithm; WEKA Tool

I. INTRODUCTION

Discovery of efficient association rules has been found useful in many applications. However, without fully considering the importance and significance of items and transactions, it is noted that some discovered rules may be expired from users' interest. Utility measures play an important role in data mining, regardless of the kind of patterns being mined. These measures are intended for selecting and ranking patterns according to their potential interest to the user. This study focused to implement the Global Profit Weight (GPW) for the frequent item set. Generally the profit can be measured in a traditional way. In this study they proposed multi criteria based profit calculation.

This work employs global profit weight algorithm is implemented using visual basic to find the profit of the item set in the transaction. Classification algorithm is used to evaluate the profit measure such as high (H), medium (M) and low (L). Widely used supervised machine learning techniques namely Naïve Bayes Decision tree classifier, VFI and IB1 Classifier are used for learning the model. The results of the models are compared and observed that Naïve Bayes performs well. WEKA tool is used to classify the data set and accuracy is calculated.

II. FREQUENT ITEM SET MINING

Frequent item sets may only contribute a small portion of the overall profit, whereas non-frequent item sets may contribute a large portion of the profit. In reality, a retail business may be interested in identifying its most valuable customers (customers who contribute a major fraction of the profits to the company). Hence, frequency is not sufficient to answer questions, such as whether an item set is highly profitable, or whether an item set has a strong impact. Profit mining is thus useful in a wide range of practical applications and was recently studied in [15].

A. Analytical Hierachical Process

This paper presents Analytic Hierarchy Process and it uses pair wise comparisons and computes the weighting factors. The utility measure [6], was proposed to overcome the shortcomings of support. It reflects the impact of the quality sold on the cost or profit of an item set. Lu et al [12], Proposed a scheme for weighting each item using a constant value without regard to the signification of transactions [12]. In this scheme, the utilities are attached to the items rather than the transactions. Wang et al [14] suggested that it remains unclear to what extent patterns can be used to maximize the business profit for an enterprise.

The project generalizes previous work on profit measure. The profit measure of the items in the transaction is defined by using the characteristic of the item. Considering the profit of an item, there are a number of important factors to consider as well. This paper defines five variables for items to compute the Global Profit Weight. (i.e. Damage, Offer, Quality, Margin and Frequency). The preference of the Quality factor is calculated according to the level of Quality measures.

III. ASSOCIATION RULE

Association Rule is an important type of knowledge representation revealing implicit relationships among the items present in large number of transactions. Given I={$i_{1,}i_2,i_3,\ldots,i_n$} as the items' space, which is a set of items, a transaction may be defined as a subset of I, and a dataset may therefore be defined as a set D of transactions. X and Y are non-empty subsets of I. The support of an item set X in a dataset D, denoted as support D(X), is defined as count D(X)/|D|, where count D(X) is the number of transactions in D containing X. An item set is said to be frequent (large) if its support is larger than a user-specified value (also called minimum support (min_sup)). An association is an implication of the form [X → Y, sup, conf], where $X \subset I$, $Y \subset I$, and $X \cap Y = Ø$. The support of $X \cup Y$ (sup) in the transactions is larger than min_sup, furthermore when X appears in a transaction; Y is likely to appear in the same transaction with a probability conf. Given a threshold of minimum support and confidence, methods of discovering association rules [4, 5, 6, 9] have become active research topics since the publication of Agarwal, Imielinski and Swami and Agarwal and Srikant papers [2, 3].

However, the traditional Association Rule Mining (ARM) model assumes that items have the same significance without taking account of their weight/attributes within a transaction or within the whole item space. But this is not always the case. For example, [wine, salmon, 1%, 80%] may be more important than [bread, milk, 3%, 80%] even though the former holds a lower support. This is because those items in the first rule usually come with more profit per unit sale, but the standard ARM simply ignores this difference.

B Problem Classification

The mining of association rules for the unweighted case has been done for several years. However, for the above reasons, association rules have been developed for weighted items. To begin with, the association rule must be defined first. Similar to [1], [2] consider a database with a set of transactions D, a set of attributes or items T, and each transaction is assigned a transaction identifier <TID>.

In this study, the major problem is to mine the association rules with weighted items, based on the different types of the association rules, which are binary association rules and quantitative association rules. New algorithms are required to solve such problems since the available algorithms cannot be solved.

IV. ASSOCIATION RULE MINING

The search space of all association rules contains exactly $3^{|I|}$ different rules. However, given all frequent item sets, this search space immediately shrinks vastly. Indeed, for every frequent item set I, there exists at most $2^{|I|}$ rules of the form $X \Rightarrow Y$, such that $X \cap Y = I$. Again, in order to efficiently traverse this search space, sets of candidate association rules are iteratively generated and evaluated, until all frequent and confident association rules are found. The underlying technique to do this is based on a similar monotonicity property as was used for mining all frequent item sets.

Proposition 1. (Confidence monotonicity) Let $X, Y, Z \subseteq I$ be three item sets, such that $X \cap Y = \{\}$. Then,

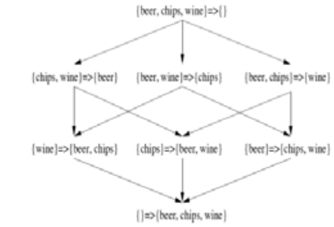


Figure 1. Example of Association Rule.

Association rules (ARs) have been widely used to determine customer buying patterns from market basket data. The task of mining association rules is mainly to discover association rules (with strong support and high confidence) in large databases. Classical Association Rule Mining (CARM) deals with the relationships among the items present in transactional databases. The typical approach is to first generate all large (frequent) item sets (attribute sets) from which the set of ARs is derived. A large item set is defined as one that occurs more frequently in the given data set than a user supplied support threshold. To limit the number of ARs generated a confidence threshold is used. The number of ARs generated can therefore be influence by careful selection of the support and confidence thresholds, however great care must be taken to ensure that item sets with low support, but from which high confidence rules may be generated, are not omitted.

$$confidence(X \setminus Z \Rightarrow Y \cup Z) \leq confidence(X \Rightarrow Y).$$

Proof. Since $X \cup Y \subseteq X \cup Y \cup Z$, and $X \setminus Z \subseteq X$, we have

$$\frac{support(X \cup Y \cup Z)}{support(X \setminus Z)} \leq \frac{support(X \cup Y)}{support(X)}.$$

C Weighted Association Rule Mining

One possible problem with the definition is that when the number of items in an item set is large, then the total weight may be large, even if each item has a small weight. In this section, we focus on the mining of weighted association rules for which the weight of an item set is normalized by the size of the item set. The choice of using unorganized or normalized weight will depend on the individual need of each application.

Although the semantics of the rules will be different, previous algorithm MINWAL(O) can be applied for this case, with a modification of the definitions of large weighted item sets and k-support bound However, MINWAL(W) new algorithm is applied.

D Frequent Pattern Mining Problem

Definition 10**.** Let D be a transaction database over a set of items I, and σ a minimal support threshold. The collection of frequent item sets in D with respect to σ is denoted by

$$\mathcal{F}(\mathcal{D},\sigma) := \{X \subseteq \mathcal{I} \mid support(X,\mathcal{D}) \geq \sigma\}$$

Example 1. Consider the database shown in Table 1 over the set of items I = {beer, chips, pizza, and wine}.

Tid	**X**
100	{beer, chips, wine}
200	{beer, chips}
300	{pizza, wine}
500	{chips, pizza}

Table 1: An Example Transaction Database d.

Table 1 shows all frequent item sets in D with respect to a minimal support threshold of 1. Table 3 shows all frequent and confident association rules with a support threshold of 1 and a confidence threshold of 50%.

V. MULTI CRITERIA DECISION MAKING

E Analytical Hierachical Process

This process uses pair wise comparisons and then computes the weighting factors through evaluation of a set of criteria elements. The decision maker starts by laying out the overall hierarchy of the decision. This hierarchy reveals the factors to be considered as well as the various alternatives in the decision. Then a number of pair wise comparisons are done, which result in the determination of factor weights and factor evaluations [18]. The process has been used to assist numerous corporate and government decision makers.

In this process the problems are decomposed into a hierarchy of criteria and alternatives. An important part of the process is accomplished by the three steps. The steps are stating the Objective, Defining criteria, and Pick the alternatives. This information is then arranged in a hierarchical tree and synthesized to determine relative rankings of alternatives. Both qualitative and quantitative criteria can be compared using informed judgments to derive weights and priorities

For Example, We have three mobiles Nokia, Motorola and Sony. While Comparing Nokia with Motorola, Nokia slightly favored then Motorola, thus we put 1/3 in the row 1 column 2 of the matrix. Comparing Nokia and Sony, Nokia strongly more preferred (likes) than Sony, thus we put actual judgment 5 on the first row, last column of the matrix. Comparing Motorola and Sony, Motorola is very strongly preferred. Thus we put his actual judgment 7 on the second row, last column of the matrix. Because we have three comparisons, thus we have 3 by 3 matrix. The diagonal elements of the matrix are always 1 and we only need to fill up the upper triangular matrix. To fill the lower triangular matrix, we use the reciprocal values of the upper diagonal. If aij is the element of row ith column jth of the matrix, then the lower diagonal is filled using this formula

$$a_{ji} = 1/a_{ij}$$

$$A = \begin{array}{cccc} & \text{Nokia} & \text{Motorola} & \text{Sony} \\ \text{Nokia} & \mathbf{1} & \mathbf{1/3} & \mathbf{5} \\ \text{Motorola} & \mathbf{3} & \mathbf{1} & \mathbf{7} \\ \text{Sony} & \mathbf{1/5} & \mathbf{1/7} & \mathbf{1} \end{array}$$

Fig 2: Pair wise Matrix

F Global Profit Weight Algorithm

In this section we propose new GPW Algorithm, which is used to derive the list of Global Profit Weight for the set of items. The inputs are product (P), Quantity (Q), Unit Price (UP). The output derived is called as Global Profit Weight (GPW).

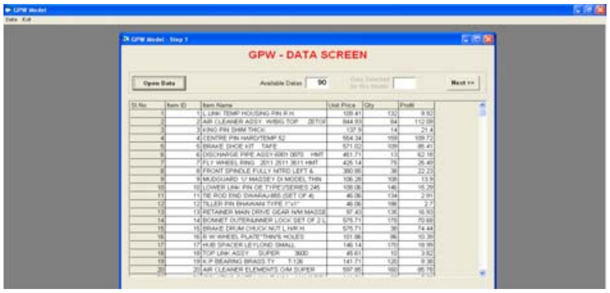
```
Algorithm       :       GPW
Input           :       Product (P) , Quantity (Q),
Unit Price (UP)
Output          :       The set of Global Profit Weight
(GPW)

Procedure Global Profit Weight (P, Q, UP)
        Begin
        GTP ← Ø;
        For each Pi ≤ n do
                        TPi ← Qi * UPi;
        GTP ← GTP + TPi;
        End
For each Pi ≤ n do
        TPPi ← TPi / GTPi;
End
Cwr ← ((∏Cr)1/Cn) / (∑(∏Cr)1/Cn)
Pwk ← ((∏CPk)1/CPn)/ (∑(∏CPk)1/CPn)
For each i ≤ n
    PWi ← Ø;
    For each x ≤ r
        For each y ≤ k
                PWi ← PWi  + (Pwxy * Cwyx) ;
        End
End
For each Pi ≤ n do
GPWi ← PWi + TPPi;
End
End
```

G GPW IMPLEMENTATION

The implementation is achieved with a real time transaction set collected from automobile shop.

Fig 3: GPW Model

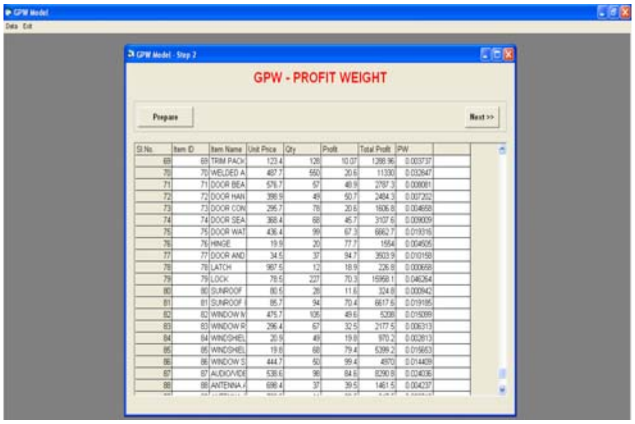


Sl No	Item ID	Item Name	Unit Price	Qty	Profit
1	1	L LINK TEMP HOUSING PIN R H	109.41	132	9.92
2	2	AIR CLEANER ASSY W/BIG TOP 2ETOI	844.93	84	112.09
3	3	KING PIN SHIM THICK	137.9	14	21.4
4	4	CENTRE PIN HARD/TEMP 52	554.34	159	109.72
5	5	BRAKE SHOE KIT TAFE	571.02	109	85.41
6	6	DISCHARGE PIPE ASSY 6901 0070 HMT	461.71	13	62.18
7	7	FLY WHEEL RING 2011 2511 3511 HMT	425.14	75	25.40
8	8	FRONT SPINDLE FULLY NITRD LEFT &	300.86	38	22.23
9	9	MUDGUARD 1/ MASSEY DI MODEL THIN	106.26	108	13.9
10	10	LOWER LINK PIN OE TYPE 2/SERIES 245	108.06	146	15.29
11	11	TIE ROD END SWARAJ 855 (SET OF 4)	46.06	134	2.91
12	12	TILLER PIN BHAWANI TYPE 1"x1"	46.06	198	2.7
13	13	RETAINER MAIN DRIVE GEAR N/M MASSE	97.43	136	16.93
14	14	BONNET OUTER&INNER LOCK SET OF 2 L	575.71	170	70.68
15	15	BRAKE DRUM CHUCK NUT L H/R H	575.71	36	74.44
16	16	R W WHEEL PLATE THIN/6 HOLES	101.86	86	10.39
17	17	HUB SPACER LEYLOND SMALL	146.14	170	18.99
18	18	TOP LINK ASSY SUPER 3600	45.61	10	3.92
19	19	K P BEARING BRASS TY T-126	141.71	120	9.38
20	20	AIR CLEANER ELEMENTS O/M SUPER	597.86	160	85.76

Fig 4: Transaction Set

The above screen shows the transaction set collected for this study. The original is kept except the product ID. When a GPW model option is selected it shows the step 1 screen. Open data button open the data from file and display it in a grid.

The following Figure 5 shows the profit weight calculation from the transaction set. As discussed in the base paper, the calculation is made on total profit and profit weight.

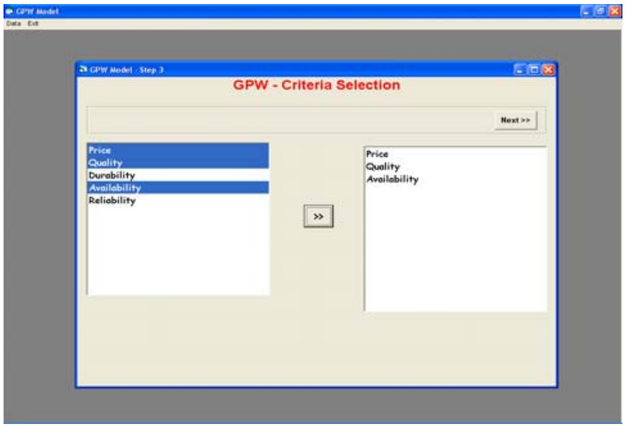


Sl.No	Item ID	Item Name	Unit Price	Qty	Profit	Total Profit	PW
69	69	TRIM PACK	123.4	128	10.07	1288.96	0.003737
70	70	WELDED A	487.7	550	20.6	11330	0.032847
71	71	DOOR BEA	576.7	57	48.9	2787.3	0.008081
72	72	DOOR HAN	398.9	49	50.7	2484.3	0.007202
73	73	DOOR CON	295.7	78	20.6	1606.8	0.004658
74	74	DOOR SEA	368.4	68	45.7	3107.6	0.009009
75	75	DOOR WAT	436.4	99	67.3	6662.7	0.019316
76	76	HINGE	19.9	20	77.7	1554	0.004505
77	77	DOOR AND	34.5	37	94.7	3503.9	0.010158
78	78	LATCH	987.5	12	18.9	226.8	0.000658
79	79	LOCK	78.5	227	70.3	15958.1	0.046264
80	80	SUNROOF	80.5	28	11.6	324.8	0.000942
81	81	SUNROOF I	85.7	94	70.4	6617.6	0.019185
82	82	WINDOW M	475.7	105	49.6	5208	0.015099
83	83	WINDOW R	296.4	67	32.5	2177.5	0.006313
84	84	WINDSHIEL	20.9	49	19.8	970.2	0.002813
85	85	WINDSHIEL	19.8	68	79.4	5399.2	0.015653
86	86	WINDOW S	444.7	50	99.4	4970	0.014409
87	87	AUDIO/VIDE	538.6	98	84.6	8290.8	0.024036
88	88	ANTENNA A	698.4	37	39.5	1461.5	0.004237

Fig 5: Profit Weight (PW)

As name described, the unit price shows the item price, qty column denotes number of units sold in one month, profit is marginal profit per item collected from the organization. Total profit is calculated through multiplying quantity and profit margin. The profit weight is calculated from total profit for the item divided by total profit earned from all the items.

The following Figure 6 describes the criteria selection. The criteria are built according to our objective of the study.



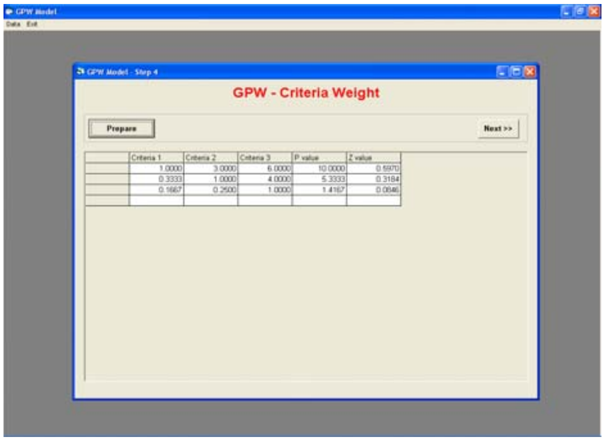
Price
Quality
Durability
Availability
Reliability

Price
Quality
Availability

Fig 6: Criteria Selection

The main objective of our study is to predict the global profit weight with its importance measured through criteria. Here, the weight can be measured up to five criterions such as price, quality, durability, availability and reliability. But in our experiment we would like to measure the products with price, quality and availability band

The following Figure 7 shows the AHP calculation for criteria weight.

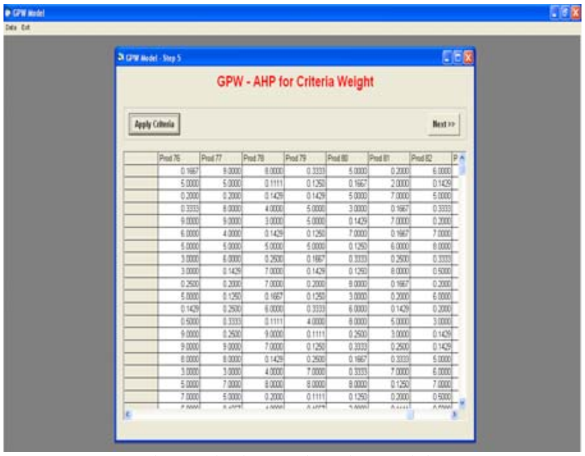


Criteria 1	Criteria 2	Criteria 3	P value	Z value
1.0000	3.0000	6.0000	10.0000	0.5970
0.3333	1.0000	4.0000	5.3333	0.3184
0.1667	0.2500	1.0000	1.4167	0.0846

Fig 7: Criteria Weight

As shown in Figure 7, we have chosen only three criteria. In this screen we would like to calculate the criteria weight through AHP pair wise matrix. We have applied the same rule proposed in AHP nine point scaling technique. The p value is product weight and the z value is a priority ratio.

The following Figure 8 shows the criteria based product weight.



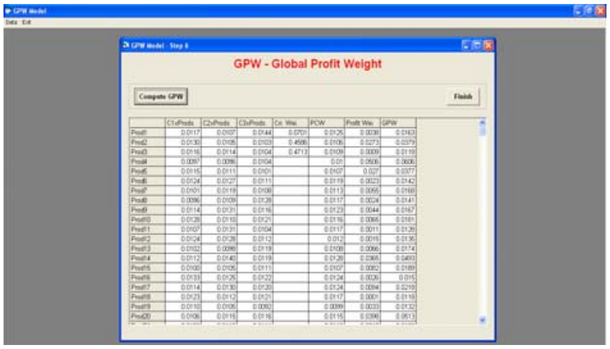
GPW - AHP for Criteria Weight

	Prod 76	Prod 77	Prod 78	Prod 79	Prod 80	Prod 81	Prod 82	P
	0.1667	9.0000	8.0000	0.3333	5.0000	0.2000	6.0000	
	5.0000	5.0000	0.1111	0.1250	0.1667	2.0000	0.1429	
	0.2000	0.2000	0.1429	0.1429	5.0000	7.0000	5.0000	
	0.3333	8.0000	4.0000	5.0000	3.0000	0.1667	0.3333	
	9.0000	9.0000	3.0000	5.0000	0.1429	7.0000	0.2000	
	6.0000	4.0000	0.1429	0.1250	7.0000	0.1667	7.0000	
	5.0000	5.0000	5.0000	5.0000	0.1250	6.0000	8.0000	
	3.0000	6.0000	0.2500	0.1667	0.3333	0.2500	0.3333	
	3.0000	0.1429	7.0000	0.1429	0.1250	8.0000	0.5000	
	0.2500	0.2000	7.0000	0.2000	8.0000	0.1667	0.2000	
	5.0000	0.1250	0.1667	0.1250	3.0000	0.2000	6.0000	
	0.1429	0.2500	6.0000	0.3333	6.0000	0.1429	0.2000	
	0.5000	0.3333	0.1111	4.0000	8.0000	5.0000	3.0000	
	9.0000	0.2500	9.0000	0.1111	0.2500	3.0000	0.1429	
	9.0000	9.0000	7.0000	0.1250	0.3333	0.2500	0.1429	
	8.0000	8.0000	0.1429	0.2500	0.1667	0.3333	5.0000	
	3.0000	3.0000	4.0000	7.0000	0.3333	7.0000	6.0000	
	5.0000	7.0000	8.0000	8.0000	8.0000	0.1250	7.0000	
	7.0000	5.0000	0.2000	0.1111	0.1250	0.2000	0.5000	

Fig 8: Criteria Based Product Weight

In this screen it shows three matrixes with twenty five products each and also calculated the p value and z value. This screen describes that it is calculated for quality wise product priority, price wise product priority and availability wise product priority.

The following Figure 9 shows the global profit weight measure.



GPW - Global Profit Weight

	C1xProds	C2xProds	C3xProds	Cr. Wei.	PCW	Profit Wei.	GPW
Prod1	0.0117	0.0107	0.0144	0.0701	0.0125	0.0038	0.0163
Prod2	0.0130	0.0105	0.0103	0.4586	0.0106	0.0273	0.0379
Prod3	0.0116	0.0114	0.0104	0.4713	0.0109	0.0009	0.0118
Prod4	0.0097	0.0098	0.0104		0.01	0.0506	0.0606
Prod5	0.0115	0.0111	0.0101		0.0107	0.027	0.0377
Prod6	0.0124	0.0127	0.0111		0.0119	0.0023	0.0142
Prod7	0.0101	0.0119	0.0108		0.0113	0.0055	0.0168
Prod8	0.0096	0.0109	0.0128		0.0117	0.0024	0.0141
Prod9	0.0114	0.0131	0.0116		0.0123	0.0044	0.0167
Prod10	0.0128	0.0110	0.0121		0.0116	0.0065	0.0181
Prod11	0.0107	0.0131	0.0104		0.0117	0.0011	0.0128
Prod12	0.0124	0.0128	0.0112		0.012	0.0015	0.0136
Prod13	0.0102	0.0098	0.0119		0.0108	0.0066	0.0174
Prod14	0.0112	0.0140	0.0119		0.0128	0.0365	0.0493
Prod15	0.0100	0.0105	0.0111		0.0107	0.0082	0.0189
Prod16	0.0133	0.0125	0.0122		0.0124	0.0026	0.015
Prod17	0.0114	0.0130	0.0120		0.0124	0.0094	0.0218
Prod18	0.0123	0.0112	0.0121		0.0117	0.0001	0.0118
Prod19	0.0110	0.0105	0.0080		0.0099	0.0033	0.0132
Prod20	0.0106	0.0115	0.0116		0.0115	0.0398	0.0513

Fig 9: Global Profit Weight

It is inferred from the above screen that criteria wise product priority is shown in first three columns. The fourth column denotes the general criteria weight calculated in Figure 8. The Profit Criteria Weight (PCW) is computed as matrix multiplication of criteria wise product priorities and criteria weight. The global profit weight measure is an abstraction of PCW and Profit Weight (Figure 5).

The global profit weight is built up with two distinct values; marginal profit earned through sales and sentimental attribute values for all products. It helps to decide the best item in both senses. The product priority can be checked with the strategy as higher the value (GPW) is the best product.

H Performance Evaluation

In the previous section describes about the model developed for global profit weight. The objective of GPW is an alternative for the traditional weighted association rule mining. An efficient algorithm is required because a significant amount of processing is undertaken to the application of weighted association rule mining. Experiment is performed with the real time data set collected from the automobile company. The GPW algorithm was implemented and presented in step wise manner.

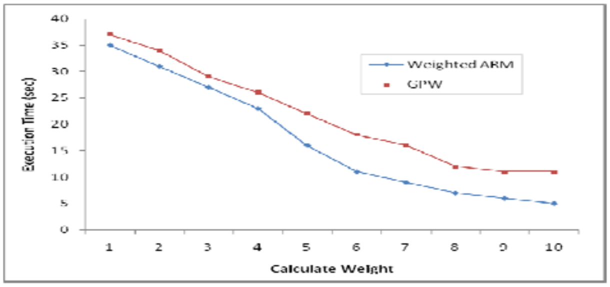


Fig 10: Execution time to generate weight for products

The above figure exhibits that the traditional weighted association rule mining algorithm has taken lesser time than global profit weight measure. The experiment has taken for 10 items. It is also noted that there is some similarity found between WARM and GPW on its execution time. The proposed algorithm is required high computation power to predict the weight measure.

To measure the product weight with its quality of output for 10 products. The quality is measured in five-point scaling like very poor quality, poor quality, fair quality, good quality and very good quality. The following figure 11 exhibits the scaling measures adopted to the weight data.

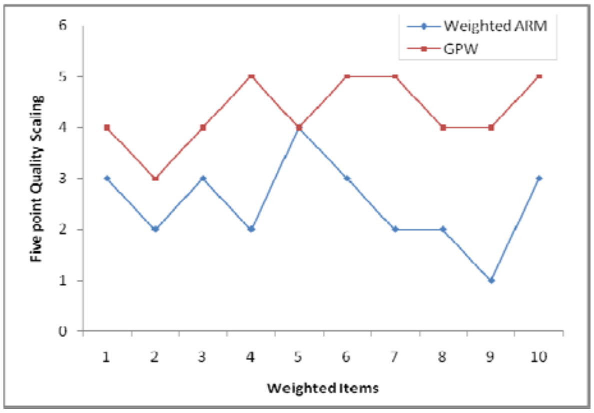
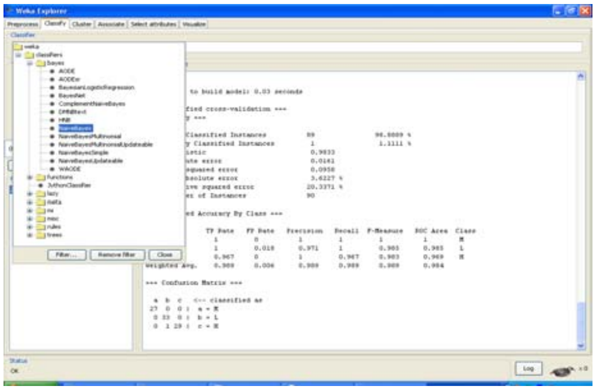


Fig 11: Five-point quality scaling

It is inferred from the above table that GPW has got highest quality scaling grade for most data. The result exhibits GPW is powerful and has more meaningful result than traditional weighted association rule mining. It is also understood from the GPW measurement that frequency of computation will be less than weighted ARM. Due to high frequency items, changes in criteria and changes in profit for item situation only needed to recompute the GPW, else can reuse the data with previous weight measures.

I Training And Testing



```
 to build model: 0.03 seconds

fied cross-validation ===
y ===

Classified Instances          89               98.8889 %
y Classified Instances         1                1.1111 %
istic                          0.9833
ute error                      0.0161
squared error                  0.0958
bsolute error                  3.6227 %
ive squared error             20.3371 %
er of Instances               90

ed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0         1           1       1           1        H
                 1         0.018     0.971       1       0.985       0.985    L
                 0.967     0         1           0.967   0.983       0.969    M
Weighted Avg.    0.989     0.006     0.989       0.989   0.989       0.984

=== Confusion Matrix ===

  a  b  c   <-- classified as
 27  0  0 |  a = H
  0 33  0 |  b = L
  0  1 29 |  c = M
```

Fig 12: Selection of Naïve Bayes from the Classifier Tab

The comparative evaluation results are summarized in Table 2. The performance of the three models were evaluated based on the three criteria, the prediction accuracy, learning time and error rate are illustrated in Figures 13, 14 and 15.

Evaluation Criteria	NAÏVE BAYES	IB1	VFI	BF TREE
Timing to build model (in secs)	**0.02 sec**	0.03 sec	0.5 sec	0.36 sec
Correctly classified Instances	**98**	88	74	71
Incorrectly Classified Instances	**2**	10	23	26
Prediction Accuracy (%)	**98.8%**	88.8%	74.4%	71.1%

Table 2: Comparative results of the classifiers

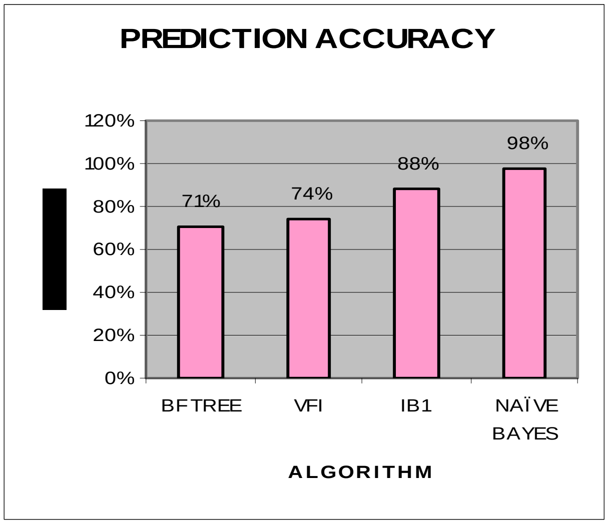


Fig 13: Prediction Accuracy

As shown in Figure 13, Naïve Bayes predicts better than other algorithms. Among the four classifiers used for the experiment, the IB1 classifier is more or less the same prediction accuracy. The accuracy rate of BF Tree and VFI classifier is the lowest among the four machine learning techniques.

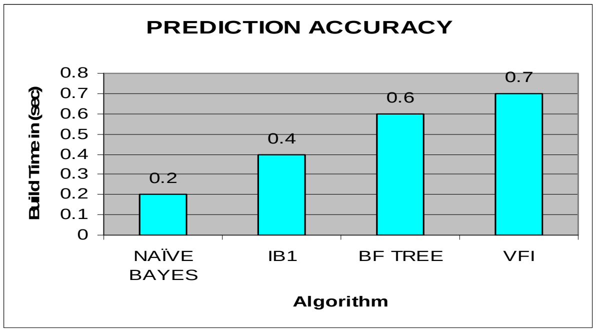


Fig 14: Time Taken to Build

As shown in Figure 14, the build times of the four schemes are under consideration. The Naïve Bayes, the probabilistic classifier tends to learn more hastily for the given dataset. There is a little statistical variation in the time taken to build the IB1, VFI and BF Tree classifier model and probabilistic model.

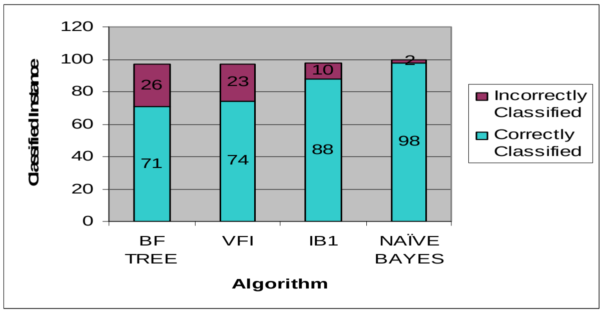


Fig 15: Classified Instance

Figure 15 shows the correctly classified instances and incorrectly classified instances from number of instances of four classifiers.

VI.CONCLUSION

The objective of study is to implement the global profit weight measure and test the performance of the algorithm with traditional weighted association rule mining. The implementation is expressed as step wise visual presentation and its performance is measured with weighted ARM.

The accuracy is classified using classification algorithm such as Navie Bayes, VFI, BF Tree and IB1and the results are compared using WEKA. It can be concluded from the study result that GPW is required high computation power to generate the weight. The result is compromised with its quality. According to the research problem, the calculated weight can be reused it for many times and as required.

REFERENCES

[1] C. Aggarwal. Towards long pattern generation in dense databases. SIGKDD Explorations, 3(1):20{26, 2001.

[2] .R. Agrawal, C. Aggarwal, and V. Prasad. Depth First Generation of Long Patterns. In 7th Int'l Conference on Knowledge Discovery and Data Mining, Aug. 2000.

[3] .R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in largedatabases.In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pages 207-216. ACM Press, 1993.

[4] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal. Mining frequent patterns with counting inference. SIGKDD Explorations, 2(2), Dec. 2000.

[5] R. J. Bayardo. Efficiently mining long patterns from databases. In ACM SIGMOD Conf. Management of Data, June 1998.

[6] Barber and H.J. Hamilton. Extracting share frequent itemsets with infrequent subsets. Data Mining and Knowledge Discovery, 7(2):153-185,2003

[7] F. Bodon. A fast apriori implementation. In Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 2003.

[8] B. Goethals. E±cient Frequent Pattern Mining. PhD thesis, transnational University of Limburg, Belgium, 2002.

[9] K. Gouda and M. J. Zaki. Efficiently mining maximal frequent itemsets. In 1st IEEE Int'l Conf. on Data Mining,

[10] Nov. 2001.

[11] G. Grahne and J. Zhu. High performance mining of maximal frequent itemsets. In 6th International Work- shop on High Performance Data Mining, May 200

AUTHORS PROFILE

ASHA RAJKUMAR M.phil, Reasearch Scholar.
Computer Science Department
P.S.G.R.Krishnammal College for Women, Coimbatore, India

G.SOPHIA REENA M.C.A., Mphil.,
HOD, BCA Department
P.S.G.R.Krishnammal College for Women
Coimbatore, India