# Decision Rule Mining Technique to improve customer satisfactory level based on review comments

[1]S.Shubha, [2]Dr. P. Suresh
[1]Research Scholar, [2]Research Supervisor &HOD, Dept. of Computer science
[1]Bharathiar University, Coimbatore, Tamil Nadu, [2]Salem Sowdeswari College, Salem, TN, India
[1]GFGC, Malleshwaram, Bangalore, India

***Abstract:*** Sentiment Analysis also known as Opinion Mining analyses customers opinions, emotions and attitude towards a product/service. With the growing development of internet, lot of product/service review data is generated online by potential customers. In this paper, different Decision Rule Mining techniques are developed to improve customer satisfactory levels based on the review comments which are performed in two steps. Initially the Machine Learning Based Sentiment classification (MLBSC) technique is used to classify the extracted words to a particular class. Opinion words and the class labels are used to obtain regressive factor by adapting various probabilistic rules. This helps to create a final decision to improve customer satisfaction. The opinions and sentiment reviews are evaluated by the customers using probabilistic rules thereby the associative regressive decision rule helps the service providers to improve the customer satisfactory level.

***IndexTerms*** **– Sentiment Analysis, Opinion Mining, Decision Rule Mining, Regression factor, probabilistic rules.**

## I. INTRODUCTION

In sentiment analysis it is very difficult to discover like and dislike of people. Hence by using matrices for words, the method can handle unseen word compositions. In order to estimate their helpfulness text mining and predictive modeling techniques towards a more complete analysis of the information captured by user-generated online reviews has been presented by many researchers. Taxonomy Aware Catalog Integration (TACI) [12] integrated products coming from multiple providers by making use of provider taxonomy information ensuring scalability that are typical on the web. However, tuning parameters does not update unless significant improvement in accuracy to avoid over fitting and it does not use any target or source taxonomy during training or application of classifier.

Tweet Analysis for Real-Time Event Detection and Earthquake (TA-RTED) [17] designed a classifier based on the keywords in order to improve the earthquake detection extracted through tweets and tweet makes difficult for the customers to identify the best reviews. However, the selection of feature set is limited. Comparative study of sentiment classification of Chinese reviews was presented for improving the classification accuracy of user review derived from various domains. Though classification accuracy was improved, but consumes large amount of time.

Many researchers have published their study of machine learning approach. Machine learning approach was developed using naïve Bayes [11] for identifying and distributing healthcare information. However, the syntactic rule-based relation extraction systems are complex based on additional tools.

Sara Hajian and Josep Domingo-Ferrer et al., [15] handles discrimination prevention in data mining and it also used for direct or indirect discrimination prevention. However, it failed to address the data distribution. An efficient algorithm was designed in [8] for detecting the top-k totally and partially unsolved sequences. This algorithm also used for reducing the running time and improving the accuracy while preserving data quality. However, it does not increase the detection accuracy of similar pattern at a required level.

Opinion mining analyzed people’s opinions, sentiments, and attitudes toward entities products, services, and their attributes. Characterization of event and prediction based on temporal patterns are detected using multivariate reconstructed phase space (MRPS) [20] using fuzzy clustering unsupervised method. However, the MRPS method provides more difficult event function for different applications. Intrinsic and extrinsic domain relevance criterion was developed in [22] aimed at improving the feasibility and effectiveness of the approach. However, it difficult to detect opinion features, including non-noun features, infrequent features, and implicit features collectively.

Probabilistic Generative Classifiers [1] used two or more classifiers resulting in the improvement of similarity measure. However, it does not address the various prior distribution investigations. The classification of trajectories on road networks was analyzed in [4] using frequent pattern-based classification which improves the accuracy. However, it does not address the pattern-based classification. The multi-class sentiment classification using that Extreme Learning Machine (ELM) methods were described in [19] for detecting their respective performance in multi-class sentiment classification of tweets. However, but it does not increases the classification accuracy effectively.

The contribution of the paper is organized as follows. Decision Rule Mining Technique (DRMT) is presented to predict the pattern for service owner and increasing their customer satisfaction based on their review comments. The Machine Learning Bayes Sentiment Classifier is subjected in DRMT to classify the class labels for each service reviews. By applying the various probabilistic rules, the regressive factor of the opinion words and Class labels are verified between the words. This helps to increase the review detection accuracy.

The rest of the paper is organized as follows. Section 2 introduces several data mining models. Section 3 introduces our Associative Regression Decision Rule Mining technique based on the customer review comments. Section 4 presents the experimental setting and Section 5 presents the results of performance evaluation. Finally, the concluding remark is presented in Section 6.

## II. Related work

In [2], aimed to apply classification technology to construct an optimum cerebrovascular disease, predictive model from that predictive model cerebrovascular disease classification rules were extracted and used to improve the prediction of cerebrovascular disease. They adopted three classification algorithms called decision tree, Bayesian classifier and back propagation neural network was presented. Another predictive model using gradient-boosted regression tree [10] to make prediction aiming at reducing the execution flow. However the prediction accuracy did not effectively increase. They found that predication and vectorization coupled with a more compact memory layout can significantly accelerate the runtime performance for tree-based models both on synthetic data and on real-world learning-to-rank dataset.

Many research works were conducted to answer top-k queries using Pareto Based Dominant Graph (DG) [5] aiming at improving the search efficiency. However, the relationship analysis remained unaddressed. Fast Distributed Mining (FDM) algorithm a version of Apriori algorithm was designed in [18] for mining of association rules in horizontally distributed databases in a secured manner aiming at minimizing the communication rounds, communication and computational cost. The protocol could improve significantly upon the leading protocol in terms of privacy and efficiency.

With the emergence of social media, web users have opened with a venue for expressing and sharing their thoughts and opinions related to different topics and events. Twitter feeds classification based on a hybrid approach was presented in [3] to achieve higher accuracy. However, this approach does not increase the accuracy at a required level.

In [14], a unified framework called, HOCTracker presented a novel density-based approach to identify hierarchical structure of overlapping communities. Probabilistic neural network and general regression neural network (PNN/GRNN) data mining model was planned in [9] for detect and preventing the oral cancer at earlier stage and also provides higher accuracy.

An incremental classification algorithm in [21] with feature space heterogeneity efficiently removed the outliers and extracted the relevant features at an early time period. In [6], an extensible method to mine experiential patterns from increasing game-logs was designed by utilizing the growing patterns.

An enhanced k-means clustering was applied in [13] to reduce the coefficient of variation and execution time using greedy approach for efficient discovery of patterns in health care data. In [7], random forest predictions were made using random forest algorithm to display prediction uncertainty. However, the true positive rate was not addressed.

Based on the aforementioned issues such as lack of detection in classification accuracy and failure in detecting the specified event in customer reviews, Decision Rule Mining technique (DRMT) is presented. The DRMT technique helps the service provider for improving the hotel customer satisfactory level at different cities. The detailed explanation is presented in forthcoming section.

## III. DESIGN OF DECISION RULE MINING TECHNIQUE

Decision Rule Mining technique (DRMT) is been done in two step process. First, the Machine Learning Bayes Sentiment Classification (MLBSC) [23] uses a base classifier where the class labels for each customer's reviews are classified. Then the opinion words and class labels are used to obtain the regressive factor using various probabilistic rules to produce a final decision on improving the customer satisfaction referred to as the Decision Rule model. These two steps are now discussed in detail.

Figure 1 shows the workflow of Decision Rule Mining technique. Given a domain-dependent review comments (i.e. opinion words) extracted from OpinRank dataset that includes the reviews of hotels in 15 different cities, we first extract a list of class labels from the Machine Learning Bayes Sentiment Classification via semantic equivalence of sentiments classification.

For each extracted class labels, we estimate its regression factor which represents the statistical association between opinion words and class labels. The resultant regressive sequence inferred is then applied with probabilistic rules to arrive at specific set of services preferred by the customers.

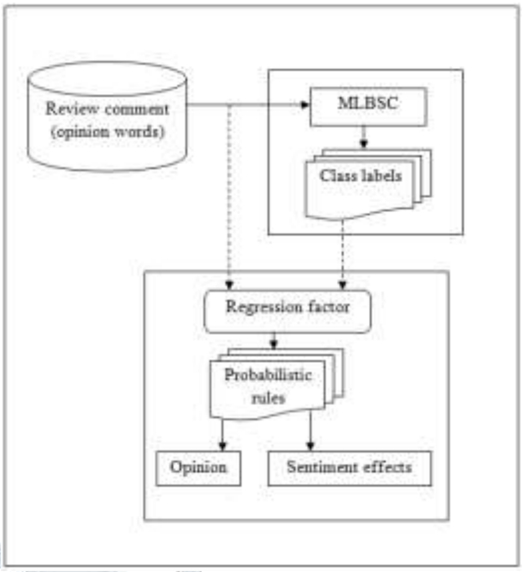


**Figure 1. Workflow of Decision Rule Mining technique**

### 3.1. Design of Regressive Sequencing Model

The first step in the design of DRMT is to obtain the class labels generated from Machine Learning Bayes Sentiment Classification (MLBSC) [23] technique. Here the sentiment class labels are extracted using Probabilistic Bayes Classifier. In MLBSC, Probabilistic Bayes Classifier is applied on the semantic opinion words to evaluate sentiment class label using the maximum likelihood estimates (MLE). The MLE of a training list (i.e. bag of words extracted from OpinRank dataset) belonging to a specific class are mathematically evaluated as given in equation (1).

$$MLE\left(\frac{B_i}{C}\right) = \frac{Count\ of\ B_i\ in\ semantic\ opinion\ words\ of\ Class\ C}{Total\ number\ of\ words\ in\ semantic\ opinion\ words\ of\ Class\ C} \qquad (1)$$

The maximum likelihood estimates is the ratio of count of semantic opinion words of class '$C$' to the total number of words. Followed by this, the class labels generated from MLBSC are subjected to regressive sequencing to infer the sentiments reflected in the customer reviews. The regressive sequencing in DRMT technique is produced with the aid of support and confidence value.

Let us assume that '$I = i_1, i_2, \dots, i_n$' represents a binary set consisting of opinion words with '$i_1, i_2, \dots, i_n$' referred to as items. Let us further assume that Transaction '$T$' (i.e. review comments) is the itemset with '$T \in I$'. Let '$P$' be the set containing items in '$I$' and transaction '$T$'contains '$P$' if '$P \in T$', then the support denotes the probability of frequent itemsets' occurrence. Smaller value of minsup results in larger number of rules whereas larger value of minsup results in smaller number of rules.

The support of rule '$P \rightarrow Q$' in the transaction database '$TD$' is the ratio between the transaction number including '$P$' and '$Q$' in the transaction sets and all the transaction number, which is then written as '$SUP\ (P \rightarrow Q)$'.

$$SUP\ (P \rightarrow Q) = \frac{Prob\ (PQ)}{N} \qquad (2)$$

The confidence of the rule '$P \rightarrow Q$' in the transaction sets is the ratio between the transaction number including '$P$' and '$Q$' and those including '$P$', which is written as '$CONF\ (P \rightarrow Q)$'. Therefore,

$$CONF\ (P \rightarrow Q) = \frac{Prob\ (PQ)}{Prob\ (P)} \qquad (3)$$

From equation (2) and (3), the sentiments reflected in customer reviews are obtained by using support and confidence value. This aids in achieving the true positive rate of customer reviews in an extensive manner. Once, the support and confidence value for customer reviews are generated, the regressive sequencing is designed. In MLBSC, the regressive sequencing model uses two variables '$y_1$' and '$y_2$' where '$y_1$' represents '$minsup$' and '$y_2$' represents '$minconf$' to infer the sentiments reflected in the customer reviews. The mathematical formulates for '$y_1$' and '$y_2$' is as given below.

$$x = \delta_0 + \delta_1\left(\frac{1}{y_1^2}\right) + \delta_2\left(\frac{1}{y_2^2}\right) \qquad (4)$$

$$x = \delta_0 + \delta_1\left(\frac{1}{y_1}\right) + \delta_2\left(\frac{1}{y_2}\right) \qquad (5)$$

$$x = \delta_0 + \delta_1 (y_1) + \delta_2 (y_2) \quad (6)$$

The regressive factor (i.e. '$y_1$' and '$y_2$') of the opinion words with class labels are checked for the association between the words. This in turn improves the associative regression factor in a significant manner. Figure 2 shows the algorithmic description of Regressive Sequencing algorithm.

| **Input**: opinion words '$I = i_1, i_2, \dots, i_n$', Transaction '$T$', |
|---|
| **Output**: Optimized true positive rate |
| Step 1: Begin<br>Step 2: For each Transaction '$T$' with opinion words '$I$'<br>Step 3: Measure the value for support using (2)<br>Step 4: Measure the value for confidence using (3)<br>Step 5: Measure '$minsup$' and '$minconf$' using (6)<br>Step 6: End for<br>Step 7: End |

**Figure 2. Regressive Sequencing algorithm**

The Regressive Sequencing algorithm performs three steps. For each transaction, customer reviews obtained from OpinRank dataset that includes hotel reviews is given as input. The first step evaluates the support value, followed by the measure of confidence value in order to identify the sentiments reflected in customer reviews. Finally, with the objective of improving the associative regression factor, sentiments reflected in customer reviews '$minsup$' and '$minconf$' are evaluated to check the association between the words.

### 3.2. Design of Decision Rule Mining Technique

The second step in the design of DRMT is to construct Decision Rule. Various probabilistic rules are generated for the class objects in the corresponding classes with more similar patterns together. Based on the probabilistic rules, the opinion and sentiments effect on customer reviews are analyzed to arrive at specific set of services preferred by the customers with their review comments. The Decision Rule helps the service providers to take decision on how to improve the hotel customer satisfactory level. Next, the frequent itemset generation algorithm is designed to the regressive sequenced dataset which is obtained through regressive sequencing model. Redundant regressive rules generated are eliminated using redundant regressive decision rule testing.

#### 3.2.1. Redundant Regressive Decision Rule Testing

Redundant regressive decision rule testing is performed in DRMT aiming at minimizing the regressive decision rule generation time and removes the redundancy involved. This is performed through elimination of redundant decision rule through regressive model.

$$P_a, P_b, P_c \ldots, P_n \rightarrow \sum_{i=1}^{n} (y_1, \mu_i, \sigma_i) \quad (7)$$

$$\text{First set of } variance\ (P_{ab}) = P_a - P_b \quad (8)$$

$$\text{Second set of variance } (P_{bc}) = P_b - P_c \quad (9)$$

From equation (7), '$\boldsymbol{\mu_i}$' symbolizes the mean of the target review for the class objects and '$\boldsymbol{\sigma}_i$' symbolizes the variance of the target review for the class objects. In (7), the mean and variance are evaluated. The variance of the association rules are calculated using equation (8) and (9) where $P_a, P_b, P_c$ are the target reviews for class object. If the variance (first set) of the association rule is lower than the variance (second set) of the association rule, then redundancy is said to be occurred in the first set. On contrary, if the variance (first set) of the association rule is greater than the variance (second set) of the association rule, then redundancy is said to be occurred in the second set. By using specified threshold value, the redundant rules are eliminated. If the identified redundancy value is obtained within the threshold value, the redundant rules are eliminated. The redundancy value is possibly occurred within the threshold value. This in turn minimizes the regressive decision rule generation time.

#### 3.2.2. Associative Regressive Decision Model

Once the redundant rules are eliminated using redundant regressive decision rule testing then, finally associative regressive decision model is designed to arrive at specific set of service preferred by the customers. Building an associative regressive decision model requires selection of a smaller, representative set of rules in order to provide an accurate representation of the training data.

The frequent itemset generation algorithm is shown in figure 3 to select the rule in an efficient manner by first sorting the rule and then remove the occurrences covered by the rule. For each itemset, the algorithm starts with the elimination of redundant rule. Followed by this rule redundant removal, the occurrence of redundancy is observed and removed in specified threshold value. Then rule sorting is performed based on the pair of rules. Finally, Associative Regressive Decision model is applied to the generated rules that help the service provider to take decision on improving the customer satisfactory level.

Let us consider a pair of rules, '$Rule_1$' and '$Rule_2$' where '$Rule_1 \gg Rule_2$'. This implies that '$Rule_1$ ' has higher preference over '$Rule_2$' and on contrary, if '$Rule_2 \gg Rule_1$', then '$Rule_2$ ' has higher preference over '$Rule_1$' is formulated as given below.

$$if(Rule_1 \gg Rule_2) \rightarrow ARD = Rule_1, Rule_2 \quad (8)$$

$$if(Rule_2 \gg Rule_1) \rightarrow ARD = Rule_2, Rule_1 \quad (9)$$

| |
|---|
| **Input**: mean of the target review for the class objects '$\mu_i$', variance of the target review for the class objects '$\sigma_i$', first set $P_{ab}$, second set $P_{bc}$, |
| **Output**: Improved customer satisfactory level |
| Step 1: Begin<br>Step 2: For each set $P_i$<br>Step 3: Perform redundant rule elimination through<br>Step 4: If $\sigma_i(P_{ab}) < \sigma_i(P_{bc})$<br>Step 5: Redundancy is found and said to be occurred in $(P_{ab})$<br>Step 6: End if<br>Step 7: If $\sigma_i(P_{ab}) > \sigma_i(P_{bc})$<br>Step 8: Redundancy is found and said to be occurred in $(P_{bc})$<br>Step 9 : End if<br>Step 10: If ( $T_h \geq redundancy\ value$ )<br>Step 11: The redundant rule is eliminated<br>Step 12: End if<br>Step 13: End for<br>Step 15: Perform rule sorting<br>Step 16: If $(Rule_1 \gg Rule_2)$<br>Step 17: $ARD = Rule_1, Rule_2$<br>Step 18: End if<br>Step 19: If $(Rule_2 \gg Rule_1)$<br>Step 20: $ARD = Rule_2, Rule_1$<br>Step 21: End if<br>Step 22: Perform Associative regressive decision model using equation (10)<br>Step 23: End |

**Figure 3. Associative regressive decision-based frequent itemset generation algorithm**

Once the sorted rules are obtained, the final step is to design Associative Regressive Decision model. The Associative Regressive Decision model is designed in such a way that, the rule has higher support value and has lower variance when '$Rule_1$' and '$Rule_2$' are applied. Then, the mathematical formulates

$$(Rule_1, Rule_2) \rightarrow (MaxSup\ (Rule_1, Rule_2), MinVar\ (Rule_1, Rule_2) \quad (10)$$

Based on equation (10), several probabilistic rules are generated for the class objects with more similar patterns together and also the rule the service preferred by the customers with their review comments. This in turn helps the service providers to take decision on improving the customer satisfactory level, thereby improving the review detection accuracy based on the review comments of the customers.

## IV. Experimental Settings

Decision Rule Mining Technique to improve customer satisfactory level based on review comments (DRMT) uses JAVA platform with WEKA tool to predict a pattern for service providers to improve their customer satisfaction based on their comments. The performance evaluation of this technique is performed using the standard benchmark datasets of services extracted from Hotel Customer Service Reviews (eg: OpinRank Dataset - Reviews from TripAdvisor). The training model for OpinRank dataset includes entire hotel reviews situated in 10 different cities (Dubai, Beijing, London, New York, New Delhi, San Francisco, Shanghai, Montreal, Last Vegas and Chicago) with the aid of Java platform with WEKA tool. This dataset has been chosen because it gives a correct picture in analyzing the comments made by tourists about services provided in the hotel rooms and food served. The total number of reviews in OpinRank dataset is approximately 250,000. For experimental purpose we have considered 350 reviews and the extracted field includes date of review, review title and full review made by the tourists.

The performance of Decision Rule Mining Technique to improve customer satisfactory level based on review comments (DRMT) technique is compared with the existing techniques namely Taxonomy-Aware Catalog Integration (TACI) [12], and Tweet Analysis for Real-Time Event Detection and Earthquake (TA-RTED) [17]. The tests on OpinRank dataset were conducted to evaluate four parameters: true positive rate, associative regression factor, regressive decision rule generation time and review detection accuracy of similar pattern.

## V. Discussion

The Decision Rule Mining Techniques to improve customer satisfactory level based on review comments (DRMT) technique is compared against the existing Taxonomy-Aware Catalog Integration (TACI) [12] and Tweet Analysis for Real-Time Event Detection and Earthquake (TA-RTED) [17]. The experimental results using JAVA on WEKA platform are compared and analyzed with the aid of graph.

### 5.1. Impact of True positive rate

The true positive rate is the ratio of sentiments that are correctly identified as belonging to a specific class in customer review words. It is mathematically formulated as.

$$TPR = \left(\frac{\text{sentiments correctly identified as belonging to a class}}{C}\right) * 100 \qquad (11)$$

From equation (11), the true positive rate '$TPR$' is obtained using the class '$C$' and is measured in terms of percentage (%). The convergence plot for 7 classes is depicted in table 1 and figure 4.

**Table 1. Tabulation for true positive rate**

| Class (C) | True Positive Rate (%) | | |
|---|---|---|---|
| | DRMT | TACI | TA-RTED |
| Class 1 | 84.32 | 73.15 | 68.21 |
| Class 2 | 88.15 | 77.12 | 71.08 |
| Class 3 | 91.35 | 80.32 | 74.28 |
| Class 4 | 85.21 | 74.18 | 68.10 |
| Class 5 | 87.57 | 76.54 | 70.46 |
| Class 6 | 89.32 | 78.29 | 72.22 |
| Class 7 | 92.14 | 81.11 | 75.04 |

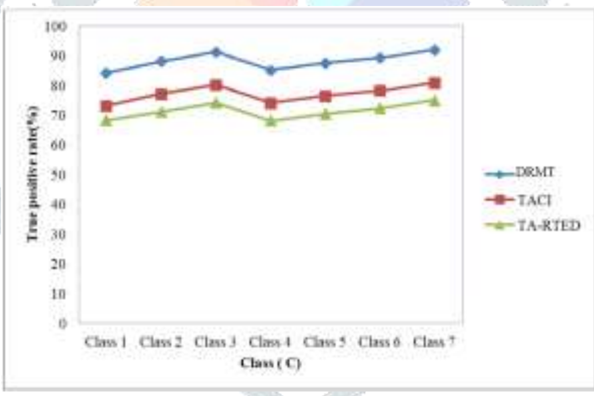


**Figure 4. Measure of true positive rate**

From figure 4 we observe that the proposed DRMT technique achieved maximum true positive rate on sentiments being correctly identified as belonging to a specific class with the application of maximum likelihood estimates when compared to other methods with the use of maximum likelihood estimates.

The maximum likelihood estimates in DRMT technique effectively constructs sentiment class labels for the testing and training data extracted from OpinRank dataset. Therefore, the true positive rate has improved by 12.52% compared to TACI [12]. Moreover, by evaluating the support and confidence value, probability of frequent itemsets occurrence are made in a significant manner. As a result, the true positive rate is increased by 19.21% compared to TA-RTED [17].

### 5.2. Impact of associative regression factor

The associative regression factor in table 2 was measured with the aid of 7 classes and 35 rules generated from 350 customer review words extracted from the OpinRank dataset. The table shows the associative regression factor using the three methods, DRMT, TACI [12] and TA-RTED [17] respectively.

**Table 2 Tabulation for associative regression factor**

| METHODS | ASSOCIATIVE REGRESSION FACTOR (%) |
|---|---|
| DRMT | 81.35 |
| TACI | 74.19 |
| TA-RTED | 65.18 |

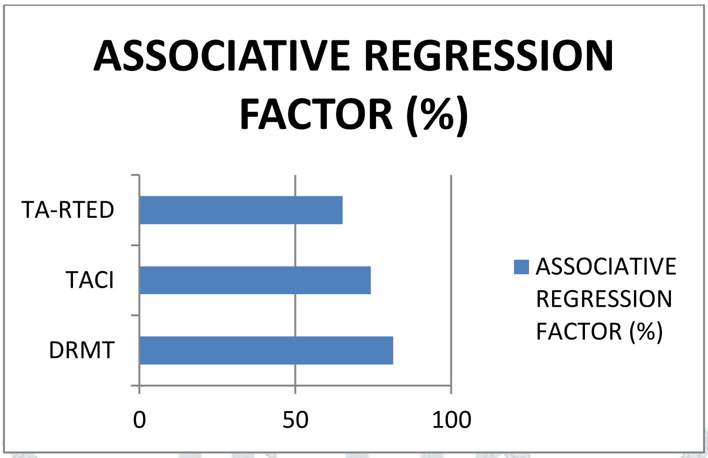


**Figure 5. Measure of associative regression factor**

Figure 5 shows the measure of associative regression factor with respect to 350 customer review words obtained from OpinRank dataset. The associative regression factor using DRMT has improved when compared to two other methods [12] and [17]. This is due to the application of Regressive Sequencing algorithm. By applying Regressive Sequencing algorithm, the support and confidence value are evaluated according to the sentiments reflected in the customer review. This in turn has improved the associative regression factor using DRMT by 8.80% and 12.14% when compared with TACI and TA-RTED respectively.

### 5.3. Impact of regressive decision rule generating time

The regressive decision rule generating time is measured using the number of rules and the time to extract single rule. It is mathematical formulated as.

$$DRGT = \sum_{i=1}^{n} Rule_i * Time\ (Rule_i) \quad (12)$$

From equation (12), the execution time '$DRGT$' is measured using the number of rules '$Rule_i$' in terms of milliseconds. Lower the regressive decision rule generation time more efficient the method is said to be. Convergence characteristics for the measure of time to extract opinions from customer reviews for 35 rules extracted from different customers are considered and compared with two other methods and is shown in table 3.

**Table 3. Tabulation of time to extract opinions from reviews**

| Number of rules | Time to extract opinions from customer reviews (ms) | | |
|---|---|---|---|
| | DRMT | TACI | TA-RTED |
| 5 | 1.31 | 1.68 | 1.85 |
| 10 | 2.51 | 2.81 | 3.05 |
| 15 | 3.79 | 4.02 | 4.22 |
| 20 | 4.96 | 5.26 | 5.48 |
| 25 | 5.85 | 6.15 | 6.35 |
| 30 | 6.3 | 6.60 | 6.80 |
| 35 | 8.32 | 8.62 | 8.82 |

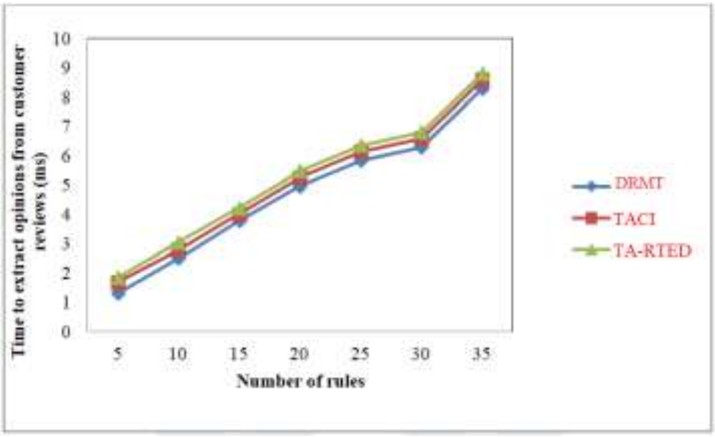


**Figure 6. Measure of Regressive decision rule generation time**

The targeting results of Regressive decision rule generating time for extracting predictive pattern using DRMT technique is compared with two other existing methods TACI [12] and TA-RTED [17]. Figure 6 is presented for visual comparison based on the number of rules. Our method differs from the TACI [12] and TA-RTED [17] in that we have incorporated associative regressive decision rule. The decision rule mining algorithm applies probabilistic rules using the mean and variance value for performing rule generation. As a result, the Regressive decision rule generating time generating decision rules using DRMT technique has increased by 9.40 to TACI [12]. Furthermore, by eliminating the redundant rule, further reduces the time for obtaining the regressive decision rule generation by 15.29% compared to TA-RTED [17].

### 5.4. Impact of Review detection accuracy of similar pattern

The review detection accuracy of similar pattern is the ratio of number of correct review patterns to the total number of test cases made. The mathematical formulation of review detection accuracy of similar pattern is formulated as given below.

$$A = \left(\frac{No.of\ correct\ review\ patterns}{Total\ no.of\ test\ cases}\right) * 100 \qquad (13)$$

From equation (13), the review detection accuracy '$A$' is measured in a significant manner in terms of percentage (%). Higher the detection accuracy more efficient the method is said to be.

**Table 4 Tabulation for review detection accuracy**

| Customer review words | Review detection accuracy (%) | | |
|---|---|---|---|
| | DRMT | TACI | TA-RTED |
| 50 | 87.53 | 74.11 | 68.21 |
| 100 | 89.31 | 77.27 | 71.25 |
| 150 | 92.14 | 80.10 | 74.08 |
| 200 | 85.14 | 73.10 | 67.07 |
| 250 | 88.21 | 76.17 | 70.15 |
| 300 | 88.15 | 86.11 | 80.11 |
| 350 | 91.35 | 79.31 | 73.21 |

The comparison of customer review detection accuracy is presented in table 4 with respect to different customer review words. Depending on the customer review words, the customer review detection accuracy either increases or decreases but found to be improved using the proposed DRMT technique. To ascertain the performance of customer review detection accuracy, comparison is made with two other existing works Taxonomy-Aware Catalog Integration (TACI) [12], and Tweet Analysis for Real-Time Event Detection and Earthquake (TA-RTED) [17].

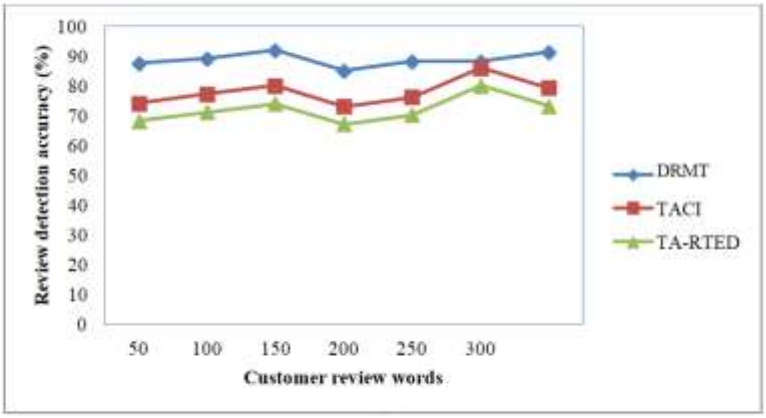


**Figure 7. Measure for review detection accuracy**

In figure 7, the customer review words are varied from 50 to 300. It is illustrative that the customer review detection accuracy has improved using the proposed DRMT technique when compared to two other existing methods. The increase is due to the application of associative regressive decision rule based frequent itemset generation algorithm, the DRMT technique chooses the rule in a greedy manner by first sorting the rule and detaches the occurrences covered by the rule.

In this way, the customer review detection accuracy has improved by using DRMT by 12.16% when compared to TACI [12]. Furthermore, by applying associative regressive decision model when applied to the generated rules, with higher support value and lower variance improves the customer satisfactory level, therefore improving the review detection accuracy based on their review comments of the customers by 18.93% than when compared to TA-RTED [17].

## VI. CONCLUSION

An efficient technique called Decision Rule Mining Technique to improve customer satisfactory level based on review comments (DRMT) is presented in this work. The technique improves the review detection accuracy that in turn improves the customer satisfaction based on their review comments and associative regression factor. The goal of Decision Rule Mining technique is to improve the true positive rate with sentiments correctly identified as belonging to a specific class and therefore to improve the associative regression factor using the customer review words extracted from OpinRank dataset which significantly contribute to the relevance. Regressive Sequencing algorithm was mainly written for improving the association regression factor in an extensive manner. In addition the DRMT algorithm eliminates the redundant rule and therefore reduces the time to extract opinions reviews and increase true positive rate. Finally, the decision rule mining technique improves the customer review detection accuracy. Extensive experiments were carried out using JAVA with WEKA tool and compared with existing methods. The results show that DRMT technique offers better performance with an improvement of review detection accuracy by 15.55% and reduces the time taken to extract opinions from reviewers by 12.34% compared to TACI and TA-RTED respectively.

## VII. ACKNOWLEDGMENT

We thank each one of those who are directly or indirectly helpful in the preparation of this paper. And also we thank our family members for their continuous support and encouragement.

## REFERENCES

**[1]** Dominik Fisch, Edgar Kalkowski, and Bernhard Sick, “Knowledge Fusion for Probabilistic Generative Classifiers with Data Mining Applications”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 3, March 2014, Pages 652-666.

**[2]** Duen-Yian Yeh, Ching-Hsue Cheng ,Yen-Wen Chen, “A predictive model for cerebrovascular disease using data mining”, Elsevier, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8970–8977.

**[3]** Farhan Hassan Khan, Saba Bashir , Usman Qamar, “TOM: Twitter opinion mining framework using hybrid classification scheme”, Elsevier, Decision Support Systems, Volume 57, January 2014, Pages 245–257.

**[4]** Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hong Cheng, “Mining Discriminative Patterns for Classifying Trajectories on Road Networks”, IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 5, May 2011, Pages 713-726.

**[5]** Lei Zou, and Lei Chen, “Pareto-Based Dominant Graph: An Efficient Indexing Structure to Answer Top-K Queries”, IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 5, May 2011, Pages 727-741.

**[6]** LiangWang, Yu Wang, and Yan Li, “Mining Experiential Patterns from Game-Logs of Board Game”, Hindawi Publishing Corporation, International Journal of Computer Games Technology, Volume 2015, December 2014 , Pages 1-21.

**[7]** Mareike Lie, Martin Hitziger, and Bernd Huwe, “The Sloping Mire Soil-Landscape of Southern Ecuador: Influence of Predictor Resolution and Model Tuning on Random Forest Predictions”, Hindawi Publishing Corporation, Applied and Environmental Soil Science, Volume 2014, February 2014, Pages 1-11.

**[8]** Massimiliano Albanese, Cristian Molinaro, Fabio Persia, Antonio Picariello, and V.S. Subrahmanian, “Discovering the Top-k Unexplained Sequences in Time-Stamped Observation Data”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 3, March 2014, Pages 577-594.

[9] Neha Sharma and Hari Om, “Usage of Probabilistic and General Regression Neural Network for Early Detection and Prevention of Oral Cancer”, Hindawi Publishing Corporation, The Scientific World Journal, Volume 2015(2015), May 2015, Pages 1-12.

[10] Nima Asadi, Jimmy Lin, Arjen P. de Vries, “Runtime Optimizations for Prediction with Tree-Based Models”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 9, September 2014, Pages 2281-2292.

[11] Oana Frunza, Diana Inkpen, and Thomas Tran, “A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts”, IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 6, June 2011, Pages 801-814.

[12] Panagiotis Papadimitriou, Panayiotis Tsaparas, Ariel Fuxman, and Lise Getoor, “TACI: Taxonomy-Aware Catalog Integration”, IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 7, July 2013, Pages 1643-1655.

[13] Ramzi A. Haraty, Mohamad Dimishkieh and MehediMasud, “An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data”, Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2015, December 2014, Pages 1-12.

[14] Sajid Yousuf Bhat and Muhammad Abulaish, “HOCTracker: Tracking the Evolution of Hierarchical and Overlapping Communities in Dynamic Social Networks”, IEEE Transactions on Knowledge and Data engineering, Volume 27, Issue 4,April 2014, Pages 1019-1032.

[15] Sara Hajian and Josep Domingo-Ferrer, “A Methodology for Direct and Indirect Discrimination Prevention in Data Mining”, IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 7, July 2013, Pages 1445-1459.

[16] Shubha.S and Dr. P. Suresh “Association Regression Decision Rule Mining for predicting customer satisfactory patterns”. Data Mining (DTMN 2016), pp. 121–134, 2016. © CS & IT-CSCP 2016 ISSN: 2231 – 5403, ISBN: 978-1-921987-49-6, DOI: 10.5121/csit.2016.60411

[17] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, “Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development”, IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 4, April 2013, Pages 919-931.

[18] Tamir Tassa, “Secure Mining of Association Rules in Horizontally Distributed Databases”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 4, April 2014, Pages 970-983.

[19] Wang, Zhaoxia, and Yogesh Parth "Extreme Learning Machine for Multi-class Sentiment Classification of Tweets." Proceedings of ELM-2015 Volume 1, Springer International Publishing, 2016. 1-11

[20] Wenjing Zhang, and Xin Feng, “Event Characterization and Prediction Based on Temporal Patterns in Dynamic Data System”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 1, January 2014, Pages 144-156.

[21] Yu Wang, “An Incremental Classification Algorithm for Mining Data with Feature Space Heterogeneity”, Hindawi Publishing Corporation, Mathematical Problems in Engineering, Volume 2014 (2014), February 2014, Pages 1-10.

[22] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang, “Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance”, IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 3, March 2014, Pages 623-634.

[23] Shubha.S and Dr. P. Suresh, “Machine Learning Bayes Sentiment Classification**”,** International Research Journal of Engineering and Technology (IRJET), Volume 02, Issue 09 Dec 2015, e-ISSN: 2395-0056, p-ISSN:2395-0072, Pg. No – 2439 to 2446.