Alleviating NB conditional independence using Multi-stage variable selection (MSVS): Banking customer dataset application

R Siva Subramanian1*, D Prabha2, J Aswini3, B Maheswari4, M Anita5

1 Research Scholar, Anna University, Chennai-600025, India
2 Associate Professor, Sri Krishna College of Engg and Tech, Coimbatore
3 Assistant Professor, Saveetha University, Chennai, India
4 Assistant Professor, Jawahar Engg College, Chennai, India
5 Assistant Professor, Jawahar Engg College, Chennai, India

E-Mail: sivasubramanian12@yahoo.com*, prabha@skcet.ac.in, aswini.jayaraman@gmail.com, mahesasi23@gmail.com, anitam.engg@gmail.com

Abstract. Customer research is one of the important aspects of understanding customer behavior patterns with business enterprises and predicate how consumer satisfaction is achieved. Customer analysis brings out various underlying information about the customer patterns with enterprises and analysis decision helps to make better marketing strategies to improve the customer lifetime and also enhance the business profit. To perform effective customer analysis in this research Naive Bayes an ML algorithm is applied. The efficiency of NB comes from its conditional independence assumption and the violation of NB assumption results in poor prediction. But in most real-time customer datasets, the NB assumption is violated due to the presence of correlated, irrelevant, and noisy variables. To improve NB prediction with these customer customers, in this research Multi-Stage Variable Selection (MSVS) is proposed to select the relevant variables from the customer dataset which helps to predicate the customer patterns wisely. The proposed approach consists of two stages in selecting the relevant variable subset from the customer datasets. Further variable subset obtained from the proposed MSVS approach is experimented with using the NB algorithm and the results obtained are compared using the wrapper and filter approaches. From the results, it clearly shows the proposed MSVS approach performs better in selecting the variable subset and improves the NB prediction in customer analysis efficiency compare to wrapper and filter approaches. Further, the proposed approach works efficiently in time and less computational compare to wrapper and filter approaches.

1. Introduction
Customer behavior analysis within the enterprises is considered an important one to understand the patterns of the customers with business enterprises and also helps to predicate how the customers are satisfied with the business services and products[1]. Since the enterprise's business purely depends on the customers, the need of understanding customer patterns is a very important one to improve the business and to enhance customer satisfaction. Performing better customer behavior patterns helps to understand the unpotencial customers within the enterprises and based upon the analysis helps to develop better marketing strategies to improve the less performing customers with the business[17]. To carry out an efficient analysis of customer patterns in this research NB an ML algorithm is performed. NB is a simple probabilistic classifier that performs customer analysis efficiently[3]. The efficiency of the NB algorithm comes from its conditional independence assumption - that is predictor variables in the dataset should conditional independent and all the predictor variables must be treated...
as equal[2]. This assumption makes the NB classifier to perform better analysis compare to other algorithms. But this NB assumption is not true with most customer datasets[4]. Since due to the advancement of the internet, technology, and CRM the collection of customer interaction data are enormous, and stored data may consist of correlated, irrelevant, noisy, and missing variables. This happens due to the collection of the same customer data from the different departments of CRM and storing the customer data in a centralized database. With these customer datasets, we cannot directly perform the customer analysis and if so proceed means then poor performance is witnessed. Further, this cause increases the complexity of the NB and results in lower prediction with these customer datasets[19]. To alleviate the NB assumption in this research multi-stage variable selection (MSVS) approach is proposed. The MSVS approach is based upon the variable selection mechanism to select the relevant variable subset and eliminate the correlated variables which help to improve the NB prediction and decrease the complexity of the NB model[5]. The main objective of variable selection is to obtain a relevant variable subset by eliminating correlated, and irrelevant variables from the whole dataset by using some sort of evaluation and search approach. The advantage of using variable selection includes a) reduce overfitting, b) minimize data dimension, c) remove irrelevant, correlated, noisy variables, d) choosing the best variable subset which helps to improve prediction accuracy and e) improves the NB prediction efficiency. Variable selection techniques can broadly be categorized into three types: filter, wrapper, and embedded approach[6].

The filter approach uses some evaluation method to evaluate the worthiness of the variables with the class label and rank accordingly to the variables scores. In the filter approach relevance of variables is obtained by using intrinsic properties of data and the variables with high rank or scores are considered for evaluation with the NB classifier. The main advantage of the filter approach is fast in execution and it is highly suitable for high-dimensional datasets and does not involve in interaction with the ML algorithm for choosing a variable subset. The main disadvantage of the filter approach is variable dependencies are ignored and this leads to poor performance in the NB classifier[7]. Likewise, the wrapper approach uses some search strategy and an ML algorithm to choose the best variable subset. The main advantage of the wrapper approach is variable subset obtained using the wrapper approach is an optimal one compare to the filter approach. This is possible due to the use of search strategies and ML algorithms to find the best variable subset. The main disadvantage of the wrapper approach is too computationally expensive due to the involvement of search strategies to find the best variable subset and has a high possibility of overfitting compare to the filter approach[8]. To address the above issue with filter and wrapper, multi-stage variable selection (MSVS) approach is proposed to select the efficient variable subset which improves NB prediction wisely with respect to time and computationally feasibility. The MSVS approach consists of two stages in variable selection: one is using filter approach(Symmetrical Uncertainty) and the other one is using wrapper approach(Sequential Forward Selection) to choose a best relevant variable subset to evaluate using NB classifier. Further performance of the MSVS approach is validated using the Banking dataset and experimental results obtained are compared using filter and wrapper approach. Experimental reveals the proposed MSVS approach works better than the filter and wrapper approach to select the relevant variable subset to improve the NB prediction wisely.

2. Proposed System

In this research, a multi-stage variable selection (MSVS) approach is proposed to select the best relevant variables subset to improve NB prediction better compare to the filter and wrapper approaches in efficient time and with less computationally. The MSVS approach consists of two-stage for variable selection: first using a symmetrical Uncertainty filter approach to rank the variables accordingly to relevance with the output class and based upon different threshold values different variable subset are selected. Further, the selected variable subset is evaluated using a sequential forward selection wrapper.
approach to select the best relevant variable subset. The objective of the MSVS approach to obtain an efficient variable subset in efficient time and with less computational. The idea of the MSVS approach is to overcome the drawbacks associated with the wrapper and filter approach in variable selection. The overall structure of the proposed MSVS approach is described in figure 1.

![Overall structure of proposed MSVS approach](image)

### 2.1 Variable selection using symmetrical Uncertainty first stage:
In the first stage, the customer dataset is evaluated using symmetrical Uncertainty filter approach to examine the worthiness of the variables with the class label. Based upon the evaluation the variable is ranked accordingly.

**Algorithm 1: Variable selection in first stage:**

1. Choose the customer dataset and divide into a learning set and testing set
2. Apply the symmetrical Uncertainty filter approach as a ranker
3. Compute the worthiness of variable with respect to the class label
4. Rank the variables from the higher score to lower score
5. Then apply different threshold values to choose a suitable variable subset.
6. Compute the performance of NB using the selected variable subset.
7. Lastly feed the variable subset obtained as input to the second stage in the wrapper approach.

#### 2.1.1 Symmetrical Uncertainty: Symmetrical Uncertainty is most commonly used feature selection approach to examine the worthiness of the feature with the class label and rank accordingly to correlation with class label. The quality of features is computed using entropy measure (information theory) and entropy of feature $F$ is calculated using [14]

\[
H(F) = - \sum P(f_i) \log_2(P(f))
\]

and entropy of $F$ after observing $B$ is computed using

\[
H(F|B) = - \sum_i P(b_j) \sum_j P(f_i|b_j) \log_2(P(f_i|b_j))
\]
here $P(f_i) - represents prior probabilities(F)$ and $P(f_i|h_j) - represents posterior probabilities$

Information Gain of the feature $F$ with target $B$ is measured using

$$IG(F, B) = H(F) - H(F|B)$$

(3)

But, IG is biased with the features which has large values. So SU is represented as

$$SU(A, B) = \frac{IG(A/B)}{H(A) + H(B)}$$

(4)

If SU value is $IG(F|B) = H(F) = H(B)$ & $SU(F, B) = 1$, then $F$ & $B$ are correlated and if SU value is $SU(F, B) = 0$ means uncorrelated[9].

Based upon the symmetrical Uncertainty measure the variables are ranked and arranged from higher rank to lower rank. Since the filter approach applied here only ranks the variables and to choose the suitable variable subset the use of threshold value is needed one. In this research different threshold is applied to check how the results vary accordingly to the threshold value and the performance of NB is computed using the variable subset obtained. Further, the variable subset obtained is given as input to the second stage where the variable subset is evaluated using a wrapper approach to find an optimal variable subset.

2.2 Variable selection using Sequential forward selection second stage:

The variable subset obtained from first stage is passed as input to second stage to find optimal variable subset using SFS wrapper approach.

Algorithm 1: Variable selection in second stage:

1. Choose the suitable variable subset(with high scores) obtained from the first stage
2. Apply SFS wrapper approach
   2.1: $F \leftarrow \emptyset$, repeat
   2.2: for each $Y_j \notin F$
   2.3: $R_j \leftarrow R(F \cup \{Y_j\})$
   2.4: Let $R' \leftarrow \text{argmax}\{R_j\}$
   2.5: $F' \leftarrow F \cup \{Y_j\}$
   2.6: if $R(F') > R(F)$ then
   2.7: $F \leftarrow F'$
   2.8: $R(F') \leftarrow R(F)$
   2.9: until $R(F') \leq R(F) || |F'| = d$ end
3. Based upon SFS approach optimal variable subset is selected.
4. Compute the performance of NB using the selected variable subset.
5. Compare the performance NB obtained using filter, wrapper and MSVS approach using various metrics scores.

2.2.1sequential Forward Selection: SFS belongs to the wrapper variable selection approach and the selection of variables starts with an empty set and lastly, the best variable subset in the dataset is selected based upon the search strategies and induction algorithm[15]. SFS algorithm starts with the null set and by evaluating the variables worthiness with other variables best variables are added. SFS approach is suitable for small variable set datasets and variables added cannot be removed back from the set. In the SFS approach, NB is applied as an induction algorithm to chose the best variable subset[10]. Further variable
subset obtained is applied with the NB classifier to check the efficiency of the proposed MSVS approach. The variable subset selected from the MSVS approach is efficient when compared to the filter and wrapper approach since the most irrelevant variable is eliminated in the filter approach and the reduced variable subset is passed into the wrapper approach, By this optimal variable subset is selected from the wrapper approach with less computational time. This makes the proposed MSVS approach select the best variable subset compare to the wrapper and filter approach.

3. Naive Bayes

Naive bayes is a simple probabilistic classifier based upon bayes theorem and the algorithm is applied widely due to its 1. simple computing, 2. time efficiency,3) works well with large datasets, and its conditional assumption[11]. Consider the dataset \( D = \{x_1, \ldots, x_n | C_k\} \) where \( \{x_1, \ldots, x_n\} \) are input predictors and \( C_k \) represents the output label were \( (k = 1, \ldots, K) \)

Consider the data \( X = \{x_1, \ldots, x_n\} \) with \( n \) variables; NB predicts the \( C_k \) class for \( X \) by

\[
p(C_k | X) = \frac{p(X|C_k)p(C_k)}{p(X)} = \frac{p(x_1, \ldots, x_n|C_k)p(C_k)}{p(x_1, \ldots, x_n)} = p(x_1, \ldots, x_n|C_k) = \prod_{i=1}^n p(x_i | C_k) \quad (5)
\]

By using chain rule, the \( p(x_1, \ldots, x_n | C_k) \) numerator can be represented as

\[
p(x_1, \ldots, x_n | C_k) = p(x_1 | x_2, \ldots, x_n, C_k)p(x_2 | x_3, \ldots, x_n, C_k)p(x_{n-1} | x_n, C_k)p(x_n | C_k) = \prod_{i=1}^n p(x_i | C_k) \quad (6)
\]

With respect to NB conditional assumption the above equation can be written like

\[
p(x_1 | x_2, \ldots, x_n | C_k) = p(x_1 | C_k) \Rightarrow p(x_1 \ldots, x_n | C_k) = \prod_{i=1}^n p(x_i | C_k) \quad (7)
\]

\[
p(C_k | x_1 \ldots, x_n) \propto p(C_k) \prod_{i=1}^n p(x_i | C_k) \quad (8)
\]

Using the above equation the NB classification procedure is carried out. The efficiency of the NB depends upon the customer dataset applied. NB implies two important assumptions of datasets: one is conditional independence- the variables must be independent of each other and the other one is all variable in the datasets should be treated as equal[18]. The violation of the NB assumption in the customer dataset leads to poor prediction[13]. Since in most real-time domain datasets, NB assumption is violated and to improve the NB prediction and to satisfy NB assumption, in this research MSVS approach is proposed. The objective of the approach is to select the best variable subset inefficient time and in less computational time.

4. Experimental Results And Discussion

The proposed MSVS approach is experimented with using the bank dataset which consists of 45211 instances with 17 variables (one target class with two outcomes)[12]. Details about the customer dataset is given in table 1. The objective of the bank dataset is to find the customer who is likely to avail of the bank service and based upon the analysis the various marketing strategies can be developed to improve the customer lifetime and customer satisfaction. Further results obtained from the proposed MSVS approach is compared using wrapper and filter approach using different metrics like[16]

\[
\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)
\]

\[
\text{Specificity} = \frac{TN}{TN+FP} \quad (10)
\]

\[
\text{Precision} = \frac{TP}{TP+FP} \quad (11)
\]
\[
\text{False Negative Rate} = \frac{FN}{FN+TP} \tag{12}
\]

Table 1. Description about the customer dataset applied

| S.NO | Variable description | Variable Type  |
|------|----------------------|----------------|
| 1    | Age                  | Numeric        |
| 2    | Job                  | categorical    |
| 3    | Martial              | categorical    |
| 4    | Education             | categorical    |
| 5    | Default               | categorical    |
| 6    | Housing               | categorical    |
| 7    | Loan                  | categorical    |
| 8    | Contact               | categorical    |
| 9    | Month                 | categorical    |
| 10   | Day                   | categorical    |
| 11   | Duration              | Numeric        |
| 12   | Campaign              | Numeric        |
| 13   | Pdays                 | Numeric        |
| 14   | Previous              | Numeric        |
| 15   | Poutcome              | categorical    |
| 16   | Balance               | Numeric        |
| 17   | Y(Target Variable)    | Categorical(Yes or No) |

Table 2. Accuracy, specificity, Precision and FNR of Naive Bayes using original Variable set

| Classifier | Accuracy | Specificity | Precision | FNR | Number of variables |
|------------|----------|-------------|-----------|-----|---------------------|
| NB         | 88.0073  | 0.926       | 0.488     | 0.472 | 16                 |

The results presented in table 2 shows modeling of Naive Bayes using original Variable set without using any preprocessing techniques.

Table 3. Accuracy, specificity, Precision and FNR of NB using symmetrical Uncertainty variable selection with different threshold value

| Methods | Percentage of Variables | Number of variables | Accuracy | Specificity | Precision | FNR  |
|---------|-------------------------|---------------------|----------|-------------|-----------|------|
|         |                         |                     |          |             |           |      |
| Symmetrical Uncertainty | 10 %                   | 2                   | 89.1951  | 0.969       | 0.572    | 0.697|
|         | 20%                     | 3                   | 89.0912  | 0.959       | 0.550    | 0.626|
|         | 30%                     | 4                   | 88.4077  | 0.950       | 0.506    | 0.615|
|         | 40%                     | 6                   | 88.463   | 0.944       | 0.508    | 0.569|
|         | 50%                     | 8                   | 88.8235  | 0.945       | 0.526    | 0.541|
|         | 60%                     | 10                  | 88.4519  | 0.939       | 0.507    | 0.528|
|         | 70%                     | 11                  | 88.4475  | 0.938       | 0.507    | 0.526|
|         | 80%                     | 13                  | 88.0582  | 0.928       | 0.490    | 0.480|
|         | 90%                     | 14                  | 87.9742  | 0.927       | 0.487    | 0.479|
|         | 100%                    | 16                  | 88.0073  | 0.926       | 0.488    | 0.472|

The results presented in the table 3 shows evaluating NB using variable subset obtained from symmetrical Uncertainty using different threshold value
Table 4. Accuracy, specificity, Precision and FNR of Naive Bayes using SFS Wrapper approach

| Methods                         | Accuracy | Specificity | Precision | FNR  | Number of variables |
|---------------------------------|----------|-------------|-----------|------|--------------------|
| Sequential Forward Selection NB | 89.8719  | 0.956       | 0.584     | 0.535| 7                  |

The results presented in the table 4 shows evaluating NB using variable subset obtained from SFS Wrapper approach.

Table 5. Accuracy, specificity, Precision and FNR of NB using symmetrical Uncertainty - SFS(MSVS approach) variable selection with different threshold value

| Methods                     | Percentage of Variables | Number of variables | Accuracy | Specificity | Precision | FNR  |
|-----------------------------|-------------------------|---------------------|----------|-------------|-----------|------|
| Symmetrical Uncertainty-Sequential Forward selection NB | 10 %                    | 1                   | 89.2858  | 0.986       | 0.647     | 0.815|
|                             | 20 %                    | 1                   | 89.2858  | 0.986       | 0.647     | 0.815|
|                             | 30 %                    | 1                   | 89.2858  | 0.986       | 0.647     | 0.815|
|                             | 40 %                    | 4                   | 89.7525  | 0.966       | 0.599     | 0.624|
|                             | 50 %                    | 5                   | 89.9759  | 0.964       | 0.605     | 0.587|
|                             | 60 %                    | 6                   | 89.8786  | 0.960       | 0.593     | 0.570|
|                             | 70 %                    | 6                   | 89.9449  | 0.964       | 0.604     | 0.591|
|                             | 80 %                    | 5                   | 89.852   | 0.956       | 0.584     | 0.542|
|                             | 90 %                    | 5                   | 89.852   | 0.956       | 0.584     | 0.542|
|                             | 100 %                   | 7                   | 89.8719  | 0.956       | 0.584     | 0.535|

The results presented in the table 5 shows evaluating NB using variable subset obtained from symmetrical Uncertainty - SFS(MVS approach) variable selection with different threshold value.

Table 6. Accuracy, specificity, Precision and FNR comparison of SU, SFS, SU-SFS and NB with original variable set

| Methods                | Accuracy | Specificity | Precision | FNR  | Number of variables |
|------------------------|----------|-------------|-----------|------|--------------------|
| SU_{NB}(10%)           | 89.1951  | 0.969       | 0.572     | 0.697| 02                 |
| SFS_{NB}               | 89.8719  | 0.956       | 0.584     | 0.535| 07                 |
| SU-SFS_{NB}(50%)       | 89.9759  | 0.964       | 0.605     | 0.587| 05                 |
| NB(original Variable subset) | 88.0073  | 0.926       | 0.488     | 0.472| 16                 |

The results presented in table 6 shows the best accuracy obtained from the MSVS, wrapper and filter approaches.
The experimental results obtained are presented in the table 2 to 6. Table 1 represents the evaluation of NB using original variable subset. Here we can see the NB gets accuracy up to 88.0073 accuracy with 16 variables. Likewise in table 3 represents the evaluation of NB using variable subset obtained from symmetrical uncertainty method. Here we have used different threshold value to select the variable subset and based upon the threshold value accuracy obtained from NB varies. From the different threshold value highest accuracy 89.1951 is obtained at 10 % threshold value with 02 variables. Table 4 represents the evaluation of NB using variable subset obtained from SFS method. Here we can see the NB gets accuracy up to 89.8719 accuracy with 07 variables. Table 5 represents the evaluation of NB using variable subset obtained from the proposed MSVS approach here we can see NB achieves higher accuracy of 89.9759 with 05 variables. From the overall results presented in table 6 represents the proposed MSVS approach performs better compare to wrapper and filter in selecting the best variable subset and improving the NB prediction. The accuracy comparison of SU, SFS, SU-SFS and NB with original Variable set is presented visually in the figure 2 and the no of variable selected for NB from filter, wrapper and MSVS approach is presented in the figure 3. Table 7 represents time taken to build the model and results reveals proposed MSVS approach performs better time complexity compare to Standard Naive Bayes.
4.1 Research Findings

1. This research aims to improve the naive bayes prediction in the customer analysis which consists of correlated and irrelevant variables in the customer datasets.

2. In filter approach the variable subset are obtained efficiently in time, but the variable subset generated are not satisfactory. Since the approach evaluate the correlation of input variables with the class label only and the dependency between the other input predictors are not examined and this make possible to choose correlated variables. Further to choose the best variables using of threshold value is considered.

3. In wrapper approach the variable subset obtained efficiently compare to filter approach. But compare to time complexity wrapper approach is infeasible and this brings back log to the wrapper approach.

4. Considering the problem with wrapper and filter approach, this research propose MSVS approach.

5. The variable subset selected from the MSVS approach is efficient when compared to the filter and wrapper approach since the most irrelevant variable is eliminated in the filter approach and the reduced variable subset is passed into the wrapper approach. By this optimal variable subset is selected from the wrapper approach with less computational time.

6. This makes the proposed MSVS approach select the best variable subset compare to the wrapper and filter approach.

7. Experimental results clearly shows the proposed MSVS-NB works better and improves the classification accuracy wisely compare to filter-NB and wrapper-NB.

5. Conclusion

In this research work to perform a better analysis of customer behavior patterns, the Multi-Stage variable Selection (MSVS) approach is proposed and experimented. The proposed approach overcomes the drawbacks associated with the wrapper and filter approaches and selects the best variable subset and improves the NB prediction better compared to both approaches. Further results obtained from the proposed MSVS approach are compared using the wrapper and filter approach using different metrics like accuracy, specificity, Precision, and FNR. The experimental results obtained reveals proposed MSVS approach achieves higher accuracy of 89.9759%, wherever wrapper approach achieves only accuracy of 89.8719% and the filter approach obtains only accuracy of 89.1951% compared to both approaches and the proposed MSVS approach is superior to both approaches in choosing best variable subset and improving the NB classifier wisely.

References

[1] Siva Subramanian, R., Prabha, D 2020 Customer behavior analysis using Naive Bayes with bagging homogeneous feature selection approach. J Ambient Intell Human Comput.

[2] Dr.D.Prabha, R.Siva Subramanian, Dr.S.Balakrishnan & Dr.M.Karpagam . 2019, Performance Evaluation of Naive Bayes Classifier with and without Filter Based Feature Selection, International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-8 Issue-10, pp.2154-2158.

[3] Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. 2018 A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. Journal of Business Research,101, 499-506

[4] Xu, W., Jiang, L., & Yu, L. 2018 An attribute value frequency-based instance weighting filter for naive Bayes. Journal of Experimental & Theoretical Artificial Intelligence,31(2) 1–12.
[5] Nazemi, A., & Fabozzi, F. J. 2018 Macroeconomic variable selection for creditor recovery rates. Journal of Banking & Finance, 89, 14–25.
[6] Zhang, C.-X., Xu, S., & Zhang, J.-S. 2018 A novel variational Bayesian method for variable selection in logistic regression models. Computational Statistics & Data Analysis, 133, 1–19.
[7] Martín, R., Aler, R. & Galván, I.M. 2018 A filter attribute selection method based on local reliable information. Appl Intell, 48, 35–45.
[8] Mafarja, M., & Mirjalili, S. 2018 Whale optimization approaches for wrapper feature selection. Applied Soft Computing, 62, 441–453.
[9] Piao, M., Piao, Y., & Lee, J. 2019 Symmetrical Uncertainty-Based Feature Subset Generation and Ensemble Learning for Electricity Customer Classification. Symmetry., 11(4), 498.
[10] Yulianti, Y., & Saifudin, A. 2020 Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes. IOP Conference Series: Materials Science and Engineering, 879, 012090.
[11] Yu, L., Gan, S., Chen, Y., & He, M. 2020 Correlation-Based Weight Adjusted Naive Bayes. IEEE Access., 8, 51377–51387.
[12] S. Moro, P. Cortez and P. Rita 2014 A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62,22-31
[13] Huan Zhang, Liangxiao Jiang, Liangjun Yu,2020, Attribute and instance weighted naive bayes, Pattern Recognition., 111, 107674
[14] Wang, C., Yao, H., & Liu, Z. 2018, An efficient DDoS detection based on SU-Genetic feature selection. Cluster Computing., 22,2505-2515.
[15] Zheng, K., Wang, X., Wu, B, Tong Wu, 2020, Feature subset selection combining maximal information entropy and maximal information coefficient. Appl Intell 50, 487–501
[16] Dhandayudam P, Krishnamuthi I 2013 Customer behavior analysis using rough set approach. J Theoret Appl Electron Commerce Res ,8,21–33
[17] R.Siva Subramanian, Dr.D.Prabha ,2017, A survey on customer relationship management, 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, Electronic ISBN: 978-1-5090-4559-4 pp. 1-5.
[18] R.Siva Subramanian, Dr.D.Prabha, 2018, Prediction of customer behaviour analysis using classification algorithms, AIP Conference Proceedings, ISBN: 978-0-7354-1647-5 Volume 1952. 020098.
[19] R.Siva Subramanian, Dr.D.Prabha, 2020, Customer Behavior Analysis using Weighted Selective Naive Bayesian, International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-9 Issue-4,pp.1110-1115.