*International Journal of Engineering & Technology*, 7 (4) (2018) 5856-5861

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*
*doi: 10.14419/ijet.v7i4.14875*
***Research paper***

# Choosing a spectacular feature selection technique for telecommunication industry using fuzzy TOPSIS MCDM

**M. Durairaj [1] *, T. S. Poornappriya [2]**

[1] *Assistant Professor, Department of Computer Science and Engineering, Bharathidasan University Tiruchirappalli, India*
[2] *Research Scholar, Department of Computer Science and Engineering, Bharathidasan University Tiruchirappalli, India*
**Corresponding author E-mail: poorna.priya23@gmail.com*

**Abstract**

Sampling and Feature Selection may be employed to depreciate processing time and hence diminishing discovery time of churn customer in the telecommunication industry. This article intends to evaluate the feature selection methods based on the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) ranking method. An Entropy-based TOPSIS method has used to recommend the one or more selections between choices, having multiple attributes. The five Ranking feature selection methods like Information Gain, Gain Ratio, Chi-Square, ReliefF and Fisher score are utilized to decrease the size of the telecommunication dataset. The classification technique like Artificial Neural Network, Naïve Bayes and Support Vector Machine are applied to determine the performance of the feature selection techniques. This Entropy-based TOPSIS method is utilized to examine and rank the feature selection methods to improve the churn customer prediction.

***Keywords***: *Feature Selection; Information Gain; Gain Ratio; Chi-Square Analysis; Relieff; Fisher; ANN; NB; SVM; Fuzzy TOPSIS.*

## 1. Introduction

Customer churn moving from individual service provider to neighboring opponent in the business is a developing problem for many server-based enterprises and especially for the telecommunication business [1]. It is an example of the vital problem for project supervisors because missing a consumer is a low-cost chance for opponents to get consumers. It has summarized that the different cost with the procurement of new consumers in ten eras added for an industry as compared to maintaining the current consumer. Maintaining present customer's heads to a significant improvement in businesses and diminished purchasing cost. These factors finally concentrated on customer churn prediction as an essential portion of telecom organizations' important decision making and outlining the process that is the primary aim of Customer Relationship Management (CRM) [2][16]. The effect of the increasing problem has also pointed to the growth of many predictive mechanisms that help some necessary tasks in the classification process, and predictive modeling.

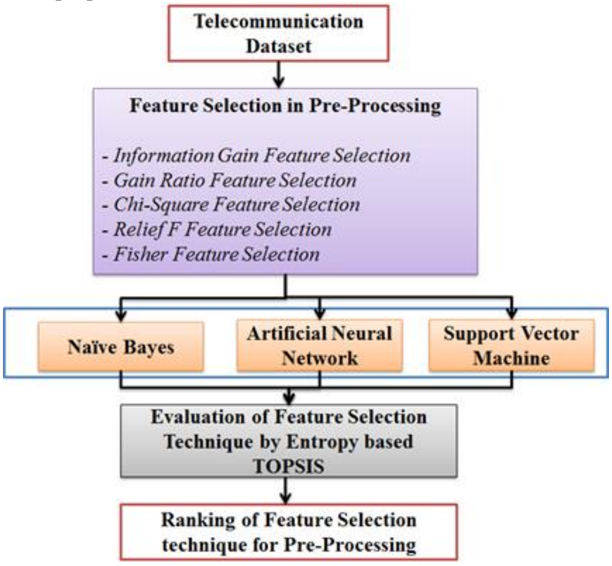
## 2. Importance of feature selection in pre-processing

The quantity of high-dimensional data that supports and is publically available on the internet has developed mainly in the past several ages. Hence, machine learning methods have difficulty in dealing with the critical amount of input attributes, which is creating an attractive problem for researchers. To use machine learning methods expeditiously, preprocessing of the input is necessary. Feature selection [3] is an example of the common prominent and frequent techniques in data preprocessing and has converted a crucial part of the machine learning method it has also distinguished as attribute selection, variable subset or variable selection in statistics and machine learning. It is the process of identifying relevant and eliminating irrelevant features, redundant or noisy data [24]. This method rushes up data mining algorithms, improves predictive accuracy and understandability. Irrelevant features are those that provide no valuable knowledge, and irrelevant features present no further information than the presently selected features. About supervised inductive learning, feature selection performs a collection of candidate features utilizing one of the three strategies [4]:

- The specific dimension of the subset of attributes that optimizes the measure of valuation.
- The little dimension of the subset that provides a critical limitation on the measures of evaluation.
- In general, the subset with the best confinement between the evaluation measure and size.

In the method of feature selection [5], redundant or noise attributes in the data may be obstructed in various situations, because they are not necessary and essential for the class theory such as microarray data examination. When the amount of samples is decidedly less than the attributes, then machine learning becomes especially tricky, because the exploration space will be distributed population. Therefore, the design will not be ready to distinguish precisely between relevant and noise data. There are two essential strategies to feature selection [16-18]. The first is Valuation of Individual, and the second is Valuation of Subset. Ranking of the attributes has identified as Evaluation of Individual. In the Individual Evaluation, the importance of an individual attribute has allocated according to its order of relevance. In the evaluation of

Subset, candidate feature subsets are formed utilizing search procedure [25].




## 3. Multi criteria decision making technique

The Decision Making is greatly natural while in considering the single standard issues; subsequent to the option picked it has the most noteworthy inclination rating [19-22]. Notwithstanding, when decision makers assess the options with the numerous criteria, and issues, for example, clashes among criteria, inclination reliance and weights of criteria, appear to confound the choice issues and ought to be overcome the complex techniques. Consequently, to manage the MCDM (Multiple Criteria Decision Making) [6] issues, the initial step need to make the sense in what number of traits or criteria exist in the issues and how to overwhelm the method for the issues (i.e., distinguishing the issues). Next, the proper information or data is gathered in which the inclinations of leaders can be effectively considered and reflected (i.e., developing the inclinations). The further work is to assemble an arrangement of conceivable choices or methodologies keeping in mind the end goal to ensure that the objective will be come to (i.e., assessing the choices). As the endeavors over, the following stride is to choose the fitting technique that assesses and outrank the conceivable options or systems (i.e., deciding the best option). The feature selection methods are general and could be utilized for various kinds of the dataset. One technique may give the superior result for one dataset while under-perform for another dataset. A multi-criteria decision-making technique like TOPSIS [7] is used to choose the top ranked feature selection techniques for the prediction of the churn customer in the telecommunication industry.

## 4. Framework for evaluation of feature selection technique

The following figure 1 depicts the framework for evaluating the feature selection technique for predicting the churn customers in the telecommunication industry. In this framework, the ranking of the feature selection methods are done by using Entropy based TOPSIS multi criteria decision making technique.

### 4.1. Information gain feature selection method

Entropy is frequently used in the information theory measure, which exemplifies the transparency of a random collection of samples [8]. It is in the establishment of Gain Ratio, Information Gain and Similarity Uncertainty (SU) [9]. Therefore, the entropy quantity is measured as a parameter of the classification's randomness. The entropy of B is

$$H(B) = \sum_{b\in Y} p(b)\log_2(p(b)) \quad (1)$$

Where p(b) is the marginal probability density function for the arbitrary variable B. If the experimental values of B in the training data set S are segregated in bestowing to the values of a second feature A, and the entropy of B in reference to the segregations persuaded by A is less than the entropy of B prior to segregating, at that point there is an association between features B and A. The entropy of B after spotting A is then:

$$H(B|A) = \sum_{a\in A} p(a)\sum_{b\in B} p(b|a)\log_2\big(p(b|a)\big) \quad (2)$$

where p(b |a ) is the conditional probability of b given a.
As given the entropy is a measure for contamination in a training set S, we can state a measure replicating supplementary data nearly B provided by A that epitomizes the quantity which the entropy of B decreases. This amount is known as IG. It is known by

$$IG = H(B) - H(B|A) = H(A) - H(A|B) \quad (3)$$

IG [8] is a proportioned measure, and it is known by equation (3). The information gained about B after observing A is alike the information gained approximately A after detecting B. A flaw of the IG criterion is that it is subjective in accord of features with further values even once they are not highly instructive.
Figure 1: Framework for Evaluating the Feature Selection methods by Entropy based TOPSIS decision making technique
IG is an asymmetrical pattern (refer to equation 3)). The information gained about Y after examining X is equal to the information gained about X after scrutinizing Y. A delicacy of the IG measure is that it is predetermined in support of features with high values even when they are not more informative.

### 4.2. Gain ratio feature selection method

The Gain Ratio [10] is the non-symmetrical measure that is introduced to compensate for the bias of the Information Gain (IG). GR is given by

$$GR = \frac{\text{Information Gain (IG)}}{H(A)}$$

Information Gain (IG) is a symmetrical measure.

$$IG = H(B) - H(B|A) = H(A) - H(A|B)$$

The information gained about B after observing A is equal to the information gained about A after observing B. A weakness of the IG criterion is that it is biased in favor of features with more values even when they are not more informative.
As in the equation (3.5) presents, when the variable B has to be predicted, then normalize the IG by dividing by the entropy of A, and vice versa. Due to this normalization, the GR values always fall in the range [0, 1]. A value of GR = 1 indicates that the knowledge of A completely predicts B, and GR = 0 means that there is no relation between B and A. In opposition to IG, the GR favors variables with fewer values.

### 4.3. Chi-square feature selection method

Feature Selection via chi-square $\chi^2$test [11] is another, very commonly used method. Chi-squared attribute evaluation evaluates the worth of a feature by computing the value of the chi-squared statistic for the class. The initial hypothesis $H_0$ is the assumption that the two features are unrelated, and the chi-squared formula tests it:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{O_{ij} - E_{ij}}{E_{ij}}\right)^2$$

Where $O_{ij}$ is the observed frequency, and $E_{ij}$ is the expected (theoretical) frequency, asserted by the null hypothesis. The greater the value of $\chi^2$, the greater the evidence against the hypothesis $H_0$.

### 4.4. Relief feature selection method

ReliefF [12] is a sort of ranking algorithm that attempts to assign a fair rank to each feature iteratively. The algorithm in the first step considers a zero vector based on the number of features. Then in each step, the algorithm selects two samples, one of them must be the nearest neighbor with respect to the class of the selected sample and the other must be the nearest sample from the other class, they update the aforementioned vector in each iteration. The algorithm will run m times, where m is lower than the sample size.

### 4.5. Fisher feature selection method

Fisher: [13] Fisher attempts to assign a value to each feature by considering how much discrimination capability the feature has and how much data of each class are scattered based on the respected feature. The higher rank of the Fisher score indicates the higher feature discrimination power of that feature.

## 5. TOPSIS decision making model

TOPSIS is to decide the best option in light of the ideas of the cooperation procedure. The cooperation procedure can be viewed as to pick an answer with the briefest Euclidean separation from the perfect arrangement and the most distant Euclidean separation from the negative-perfect arrangement. The methodology of TOPSIS can be portrayed as takes after.
Given a set of alternatives, $A = \{A_i \mid i = 1, \cdots, n\},$ and a set of criteria, $C = \{C_j \mid j = 1, \cdots, m\},$ where $X = \{x_{ij} \mid i = 1, \cdots, n; j = 1, \cdots m\}$ denotes the set of ratings and $W = \{w_j \mid j = 1, \cdots, m\}$ is the set of weights. Then the information table $I = (A, C, X, W)$ can be represented as:

**Table 1:** The Information Table of TOPSIS

| $I$ | $C_1$ | $C_2$ | $\cdots$ | $C_m$ |
|---|---|---|---|---|
| $A_1$ | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1m}$ |
| $A_2$ | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2m}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $A_n$ | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{nm}$ |
| $W$ | $w_1$ | $w_2$ | $\cdots$ | $w_m$ |

The first step of TOPSIS is to calculate the normalized ratings by

$$\tilde{S}_i^+ = \sqrt{\sum_{j=1}^{m}[\tilde{v}_{ij}(\boldsymbol{x}) - \tilde{v}_j^+(\boldsymbol{x})]^2}, i = 1, \cdots, n \qquad r_{ij}(\boldsymbol{x}) = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, m$$

and then to calculate the weighted normalized ratings by
$v_{ij}(\boldsymbol{x}) = w_j r_{ij}(\boldsymbol{x}), \quad i = 1, \cdots, n; \quad j = 1, \cdots, m.$
Next the positive ideal point (PIS) and the negative ideal point (NIS) are derived as:

$$\text{PIS} = \boldsymbol{A}^+ = \{v_1^+(x), v_2^+(x), \cdots, v_j^+(x), \cdots, v_m^+(x)\}$$

$$= \{(\max_i v_{ij}(\boldsymbol{x}) \mid j \in J_1), \ (\min_i v_{ij} \mid j \in J_2) \mid i = 1, \cdots, n\}$$

$$\text{PIS} = \boldsymbol{A}^- = \{v_1^-(\boldsymbol{x}), v_2^-(\boldsymbol{x}), \cdots, v_j^-(\boldsymbol{x}), \cdots, v_m^-(\boldsymbol{x})\}$$

$$= \{(\min_i v_{ij}(\boldsymbol{x}) \mid j \in J_1), \ (\max_i v_{ij}(\boldsymbol{x}) \mid j \in J_2) \mid i = 1, \cdots, n\}$$

where $J_1$ and $J_2$ are the benefit and the cost attributes, respectively.
The following step are to calculate the separation from the PIS and the NIS between the alternatives. The separation values can be measured by using the Euclidean distance which is given as:

$$S_i^+ = \sqrt{\sum_{j=1}^{m}[v_{ij}(\boldsymbol{x}) - v_j^+(\boldsymbol{x})]^2}, i = 1, \cdots, n$$

and

$$S_i^- = \sqrt{\sum_{j=1}^{m}[v_{ij}(\boldsymbol{x}) - v_j^-(\boldsymbol{x})]^2}, i = 1, \cdots, n$$

The similarities to the PIS can be derived as:

$$C_i^* = S_i^- \big/ (S_i^+ + S_i^-) \quad , i = 1, \cdots, n$$

where $C_i^* \in [0,1] \quad \forall i = 1, \cdots, n$.
Finally, the preferred orders can be obtained according to the similarities to the PIS ($C_i^*$) in descending order to choose the best alternatives.

### 5.1. Fuzzy TOPSIS decision making model

Once the last criteria weights are received, they can be connected with a numerous criteria choice setting to rank an arrangement of options for which the execution measures on the diverse criteria are known. All the more correctly, MCDM alludes to settling on choices within the sight of various, clashing criteria in general. For this research, the TOPSIS (Technique for Order Preference Similarity to Ideal Solution) is received. The principle thought originated from the idea of the compromise answer for pick the option where the closest to the positive of the perfect classification (ideal feature selection method) and the most distant to the negative perfect arrangement (mediocre feature selection method). At that point, pick is the best one of sorting, which will be the best option. The brief methods of TOPSIS can be portrayed as takes after. Since the preferred ratings usually refer to the subjective uncertainty, it is natural to extend TOPSIS to consider the situation of fuzzy numbers. Fuzzy TOPSIS [14] can be intuitively extended by using the fuzzy arithmetic operations as follows:
Given a set of alternatives, $A = \{A_i \mid i = 1, \cdots, n\},$ and a set of criteria, $C = \{C_j \mid j = 1, \cdots, m\},$ where $\tilde{X} = \{\tilde{x}_{ij} \mid i = 1, \cdots, n; j = 1, \cdots m\}$ denotes the set of fuzzy ratings and $\tilde{W} = \{\tilde{w}_j \mid j = 1, \cdots, m\}$ is the set of fuzzy weights.
The first step of TOPSIS is to calculate the normalized ratings by

$$\tilde{r}_{ij}(\boldsymbol{x}) = \frac{\tilde{x}_{ij}}{\sqrt{\sum_{i=1}^{n} \tilde{x}_{ij}^2}}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, n$$

and then to calculate the weighted normalized ratings by

$$\tilde{v}_{ij}(\boldsymbol{x}) = \tilde{w}_j \tilde{r}_{ij}(\boldsymbol{x}), \quad i = 1, \cdots, n; \quad j = 1, \cdots, m.$$

Next the positive ideal point (PIS) and the negative ideal point (NIS) are derived as

$$\text{PIS} = \tilde{\boldsymbol{A}}^+ = \{\tilde{v}_1^+(\boldsymbol{x}), \tilde{v}_2^+(\boldsymbol{x}), \cdots, \tilde{v}_j^+(\boldsymbol{x}), \cdots, \tilde{v}_m^+(\boldsymbol{x})\}$$

$$= \{(\max_i \tilde{v}_{ij}(\boldsymbol{x}) \mid j \in J_1), \ (\min_i \tilde{v}_{ij}(\boldsymbol{x}) \mid j \in J_2) \mid i = 1, \cdots, n\}$$

$$\text{PIS} = \tilde{\boldsymbol{A}}^- = \{\tilde{v}_1^-(\boldsymbol{x}), \tilde{v}_2^-(\boldsymbol{x}), \cdots, \tilde{v}_j^-(\boldsymbol{x}), \cdots, \tilde{v}_m^-(\boldsymbol{x})\}$$

$$= \{(\min_i \tilde{v}_{ij}(\boldsymbol{x}) \mid j \in J_1), \ (\max_i \tilde{v}_{ij}(\boldsymbol{x}) \mid j \in J_2) \mid i = 1, \cdots, n\}$$

where $J_1$ and $J_2$ are the benefit and the cost attributes, respectively.
Similar to the crisp situation, the following step is to calculate the separation from the PIS and the NIS between the alternatives. The separation values can also be measured using the Euclidean distance given as:
And

$$\tilde{S}_i^- = \sqrt{\sum_{j=1}^{m}[\tilde{v}_{ij}(\boldsymbol{x}) - \tilde{v}_j^-(\boldsymbol{x})]^2}, i = 1, \cdots, n$$

Where

$$\max\{\tilde{v}_{ij}(\boldsymbol{x})\} - \tilde{v}_j^{+}(\boldsymbol{x}) = \min\{\tilde{v}_{ij}(\boldsymbol{x})\} - \tilde{v}_j^{-}(\boldsymbol{x}) = 0.$$

Then, the defuzzified separation values should be derived by using one of the defuzzified methods, such as CoA to calculate the similarities to the PIS.
Next, the similarities to the PIS is given as

$$C_i^* = {D(S_i^-)} \big/ {[D(S_i^+) + D(S_i^-)]} \quad , i = 1, \cdots, n$$

where

$$C_i^* \in [0,1] \quad \forall i = 1, \cdots, n$$

Finally, the preferred orders are ranked according to $C_i^*$ in descending order to choose the best alternatives.
Step by Step procedure for the Fuzzy TOPSIS for ranking the feature selection techniques
Step 1: Construct the Decision Matrix for all the qualities.
Step 2: Construct the Normalized Decision Matrix.
Step 3: If elective quality is fuzzy then deciding the ENTROPY of worth parameters taking into account of connection for the Fuzzy information else decide the ANP weight.
Step 4: Calculate the Weighted Normalized Decision Matrix with weighting vector and choice lattice.
Step 5: Determine irrefutably the Positive Ideal and Negative Ideal arrangements.
Step 6: Calculate the Separation Measures.
Step 7: Calculate the Relative closeness to the Ideal arrangement.
Step 8: Rank the Preference Order.
Output: Feature Selection techniques taking into account of relative closeness for perfect classification

## 6. Dataset description

Following table 2 depicts the description about the dataset. This dataset is composed of 21 features [15]. It is the combination of churner and non-churners customer details. Their call rates, Message rates, International call rates etc. are included in this dataset.

**Table 2:** Description of the Dataset

| Feature Number | Feature Name | Description of the feature |
|---|---|---|
| 1 | Account Length | This holds the account number of the customer |
| 2 | VMail Message | Number of messages to the Voice Mail number |
| 3 | Day Minutes | Holds the total day minutes of the customer |
| 4 | Evening Minutes | Holds the total evening minutes of the customer |
| 5 | Night Minutes | Holds the total night minutes of the customer |
| 6 | International Minutes | Holds the total international minutes of the customer |
| 7 | Customer Service Calls | Total number of customer service calls |
| 8 | International Plan | 1-Yes, 0-No |
| 9 | Vmail Plan | 1-Yes, 0-No |
| 10 | Day Calls | Holds the total calls for that day by the customer |
| 11 | Day Charge | Total Charge for that day |
| 12 | Evening Calls | Total calls made in the evening |
| 13 | Evening Charge | Total evening call charge in the evening |
| 14 | Night Calls | Total calls made in the night |
| 15 | Night Charge | Total night charge |
| 16 | International Calls | Total international call |
| 17 | International Charge | Charge made for the international calls |
| 18 | State | Customer belonging to the state |
| 19 | Area code | Area code of the customer |
| 20 | Phone Number | Phone number of the customer |
| 21 | Churn | 1-Yes, 0-No |

## 7. Result and discussion

The following parameters are considered to evaluate the feature selection methods. In the ideal situation, some parameters like accuracy, the true positive rate should have maximum values while others like the number of features, errors like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE) should have the least amount. All settings are considered equivalently relevant, and unit weight has allotted to each of them. However in exceptional circumstances, some parameters may have more effect than the others, so weight has to conform accordingly. The following step by step procedure represents the mining of vague information pattern of heart disease diagnosis dataset.
Table 3 depicts the performance of the original dataset, different feature selection methods like Chi-Squared analysis, Gain Ratio, Information Gain, ReliefF and Fisher using Naïve Bayes classification. From the table 3, it is clear that the feature selection methods like Chi-Squared analysis, Gain Ratio, Information Gain performs better in all aspects than the other two techniques like ReliefF and Fisher feature selection methods.

**Table 3:** Performance Analysis of the Original Dataset and Feature Selection Methods Like Chi-Squared, Gain Ratio, Information Gain, Relieff and Fisher by Using Naïve Bayes Classification

| Performance Metrics | Original Dataset | Feature Selection Methods | | | | |
|---|---|---|---|---|---|---|
| | | CS | GR | GR | RF | Fisher |
| Accuracy | 49.333 | 55.33 | 51.33 | 50.00 | 45.66 | 52.66 |
| Kappa Statistics | 0.0397 | 0.033 | 0.031 | 0.045 | -0.044 | 0.018 |
| MAE | 0.5201 | 0.445 | 0.489 | 0.504 | 0.534 | 0.497 |
| RMSE | 0.6301 | 0.544 | 0.603 | 0.6 | 0.638 | 0.608 |
| RAE | 238.15 | 204.1 | 224.0 | 129.8 | 224.6 | 290.6 |
| RRSE | 191.56 | 165.4 | 183.3 | 136.2 | 194.6 | 209.0 |
| TPR | 0.493 | 0.553 | 0.513 | 0.5 | 0.457 | 0.52 |
| FPR | 0.396 | 0.481 | 0.44 | 0.439 | 0.564 | 0.466 |
| Precision | 0.806 | 0.798 | 0.8 | 0.637 | 0.76 | 0.84 |
| Recall | 0.493 | 0.553 | 0.513 | 0.5 | 0.457 | 0.52 |
| F-Measure | 0.574 | 0.629 | 0.594 | 0.524 | 0.544 | 0.616 |
| ROC Area | 0.554 | 0.581 | 0.554 | 0.544 | 0.467 | 0.512 |

Table 4 represents the performance analysis of the original dataset, feature selection methods like Chi-Squared analysis, Gain Ratio, Information Gain, ReliefF and Fisher using Artificial Neural Network classification method. From the table 4, it is showed that the methods like Chi-Squared analysis, Gain Ratio, Information Gain gives the increase in accuracy, decreased error rates than ReliefF and Fisher feature selection methods by using Artificial Neural Network.

**Table 4:** Performance Analysis of the Original Dataset and Feature Selection Methods Like Chi-Squared, Gain Ratio, Information Gain, Relieff and Fisher by Using Artificial Neural Network Classification

| Performance Metrics | Original Dataset | Feature Selection Methods | | | | |
|---|---|---|---|---|---|---|
| | | CS | GR | GR | RF | Fisher |
| Accuracy | 51.66 | 73.33 | 85.33 | 73.66 | 63.33 | 53.66 |
| Kappa Statistics | 0.1465 | -0.006 | 0.028 | 0 | -0.015 | -0.597 |
| MAE | 0.0358 | 0.266 | 0.146 | 0.2633 | 0.108 | 0.963 |
| RMSE | 0.1892 | 0.516 | 0.383 | 0.513 | 0.329 | 0.981 |
| RAE | 96.619 | 68.5 | 67.16 | 67.72 | 99.52 | 247.7 |
| RRSE | 145.170 | 112.2 | 116.4 | 116.5 | 141.3 | 222.8 |
| TPR | 0.517 | 0.733 | 0.853 | 0.737 | 0.133 | 0.037 |
| FPR | 0.741 | 0.738 | 0.833 | 0.737 | 0.149 | 0.962 |
| Precision | 0.488 | 0.542 | 0.793 | 0.543 | 0.128 | 0.074 |
| Recall | 0.517 | 0.733 | 0.853 | 0.737 | 0.133 | 0.037 |
| F-Measure | 0.502 | 0.623 | 0.817 | 0.625 | 0.112 | 0.044 |
| ROC Area | 0.388 | 0.498 | 0.51 | 0.5 | 0.492 | 0.037 |

Table 5 gives the performance analysis of the original dataset and feature selection methods like Chi-Squared analysis, Gain Ratio, Information Gain, ReliefF and Fisher using Support Vector Machine classification. From the table 5, it is clear that CS, IG and GR feature selection methods gives more accuracy, precision, true positive rate, Kappa Statistics value and lessen the error rates than the other two techniques by using SVM classifier.

**Table 5:** Performance analysis of the Original Dataset and Feature Selection Methods like Chi-Squared, Gain Ratio, Information Gain, ReliefF and Fisher by using Support Vector Machine Classification

| Performance Metrics | Original Dataset | Feature Selection Methods | | | | |
|---|---|---|---|---|---|---|
| | | CS | GR | GR | RF | Fisher |
| Accuracy | 73.667 | 87.33 | 86.66 | 69.66 | 73.66 | 66.33 |
| Kappa Statistics | 0 | 0.032 | 0.018 | 0.047 | 0.043 | -0.024 |
| MAE | 0.0692 | 0.126 | 0.133 | 0.003 | 0.069 | 0.368 |
| RMSE | 0.1845 | 0.355 | 0.365 | 0.250 | 0.184 | 0.569 |
| RAE | 186.758 | 58.00 | 61.05 | 78.01 | 186.7 | 162.5 |
| RRSE | 141.538 | 87.20 | 111.0 | 125.0 | 141.5 | 212.3 |
| TPR | 0.737 | 0.873 | 0.867 | 0.697 | 0.737 | 0.663 |
| FPR | 0.737 | 0.854 | 0.855 | 0.702 | 0.69 | 0.579 |
| Precision | 0.543 | 0.812 | 0.794 | 0.608 | 0.552 | 0.567 |
| Recall | 0.737 | 0.873 | 0.867 | 0.697 | 0.737 | 0.663 |
| F-Measure | 0.625 | 0.823 | 0.82 | 0.632 | 0.631 | 0.612 |
| ROC Area | 0.442 | 0.51 | 0.506 | 0.497 | 0.464 | 0.492 |

Table 6 depicts the confidence level of the original dataset, Chi-Squared analysis, Gain Ratio, Information Gain, ReliefF and Fisher feature selection methods using various classification methods like Naïve Bayes, Artificial Neural Network and Support Vector Machine by using Entropy based TOPSIS multi criteria decision making technique. The confidence value is high for CS analysis, IG and GR feature selection methods than the existing methods using different classification techniques.

**Table 6:** Confidence Level of the Original Dataset, Chi-Squared Analysis, Gain Ratio, Information Gain, Relieff And Fisher Feature Selection Methods by Using NB, ANN and SVM Classifiers

| Feature Selection Methods | Classification Techniques | | |
|---|---|---|---|
| | Naïve Bayes | Artificial Neural Network | Support Vector Machine |
| Original Dataset | 0.1577 | 0.1483 | 0.1986 |
| Chi-Squared | 0.8587 | 0.9621 | 0.7891 |
| Gain Ratio | 0.9216 | 0.8034 | 0.7104 |
| Information Gain | 0.8011 | 0.8314 | 0.8959 |
| ReliefF | 0.4462 | 0.6117 | 0.2959 |
| Fisher | 0.2575 | 0.3226 | 0.4161 |

Table 7 represents the ranking of the feature selection methods using different classification techniques like NB, ANN and SVM

**Table 7:** Fuzzy TOPSIS Based Ranking of the Original Dataset, Chi-Squared Analysis, Gain Ratio, Information Gain, Relieff And Fisher Feature Selection Methods by Using NB, ANN and SVM Classifiers

| Feature Selection Methods | Classification Techniques | | | Final Rank |
|---|---|---|---|---|
| | Naïve Bayes | Artificial Neural Network | Support Vector Machine | |
| Original Dataset | 6 | 6 | 6 | 6 |
| Chi-Squared | 2 | 1 | 1 | 1 |
| Gain Ratio | 1 | 3 | 3 | 3 |
| Information Gain | 3 | 2 | 2 | 2 |
| ReliefF | 5 | 5 | 4 | 5 |
| Fisher | 4 | 4 | 5 | 4 |

## 8. Conclusion

Alternatives can be performed quickly when there is only one feature, but selection among many possible methods sometimes become difficult job because they have multiple metrics. The strength of feature selection methods mostly depends on the variety of dataset. One technique may give excellent effect on one type of dataset but may underperform on a different kind of dataset. TOPSIS can be utilized to recommend one, among some possible alternatives where each choice has several features. In this paper, Entropy-based TOPSIS is used to rank the feature selection techniques like Chi-Squared analysis, Gain Ratio, Information Gain, ReliefF and Fisher. The feature selection techniques have estimated by Naïve Bayes, Artificial Neural Network, Support Vector Machine classification methods. From the above result and discussion, it has confirmed that feature selection methods like CS, IG and GR performs well in the ANN classification technique.

## References


[1] Coussement, Kristof, Stefan Lessmann, and Geert Verstraeten. "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry." *Decision Support Systems* 95 (2017): 27-36. https://doi.org/10.1016/j.dss.2016.11.007.

[2] Mahajan, Vishal, Richa Misra, and Renuka Mahajan. "Review on factors affecting customer churn in telecom sector." *International Journal of Data Analysis Techniques and Strategies* 9.2 (2017): 122-144. https://doi.org/10.1504/IJDATS.2017.085898.

[3] Stripling, Eugen, et al. "Profit maximizing logistic model for customer churn prediction using genetic algorithms." *Swarm and Evolutionary Computation* 40 (2018): 116-130. https://doi.org/10.1016/j.swevo.2017.10.010.

[4] Zhang, Rui, FeipingNie, and Xuelong Li. "Self-weighted supervised discriminative feature selection." *IEEE transactions on neural networks and learning systems* 29.8 (2018): 3913-3918. https://doi.org/10.1109/TNNLS.2017.2740341.

[5] Sheikhpour, Razieh, et al. "A survey on semi-supervised feature selection methods." *Pattern Recognition* 64 (2017): 141-158. https://doi.org/10.1016/j.patcog.2016.11.003.

[6] Kurth, Margaret H., et al. "Trends and applications of multi-criteria decision analysis: use in government agencies." *Environment Systems and Decisions* 37.2 (2017): 134-143. https://doi.org/10.1007/s10669-017-9644-7.

[7] Gupta, Charu, and Amita Jain. "Fuzzy Multi-Criteria Decision Making and Fuzzy Information Gain Based Automotive Recommender System." *North American Fuzzy Information Processing Society Annual Conference*. Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-67137-6_30.

[8] Jadhav, Swati, Hongmei He, and Karl Jenkins. "Information gain directed genetic algorithm wrapper feature selection for credit rating." *Applied Soft Computing* 69 (2018): 541-553. https://doi.org/10.1016/j.asoc.2018.04.033.

[9] Venkataraman, Sivakumar, and Rajalakshmi Selvaraj. "Optimal and Novel Hybrid Feature Selection Framework for Effective Data Classification." *Advances in Systems, Control and Automation*. Springer, Singapore, 2018. 499-514. https://doi.org/10.1007/978-981-10-4762-6_48.

[10] Liu, Yang, Jian-Wu Bi, and Zhi-Ping Fan. "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms." *Expert Systems with Applications* 80 (2017): 323-339. https://doi.org/10.1016/j.eswa.2017.03.042.

[11] Thaseen, Ikram Sumaiya, and CherukuriAswani Kumar. "Intrusion detection model using fusion of chi-square feature selection and multi class SVM." *Journal of King Saud University-Computer and Information Sciences* 29.4 (2017): 462-472. https://doi.org/10.1016/j.jksuci.2015.12.004.

[12] Urbanowicz, Ryan J., et al. "Relief-based feature selection: introduction and review." *Journal of biomedical informatics* (2018). https://doi.org/10.1016/j.jbi.2018.07.014.

[13] Senawi, Azlyna, Hua-Liang Wei, and Stephen A. Billings. "A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking." *Pattern Recognition*67 (2017): 47-61. https://doi.org/10.1016/j.patcog.2017.01.026.

[14] Onu, Pascal U., et al. "Evaluation of sustainable acid rain control options utilizing a fuzzy TOPSIS multi-criteria decision analysis model frame work." *Journal of cleaner production* 141 (2017): 612-625.

[15] https://www.kaggle.com/becksddf/churn-in-telecoms-dataset. https://doi.org/10.1016/j.jclepro.2016.09.065.

[16] Durairaj, M, and T.S Poornappriya. "A Review on Big Data Analytics Tools for Telecommunication Industry." International Journal of Control Theory and Applications. 9.27 (2016): 185-193. Print.

[17] LakshmanaprabuS.K, Sachi Nandan Mohanty, K. Shankar, Arunkumar N, GustavoRamireze. "Optimal deep learning model for classification of lung cancer on CT images", *Future Generation Computer Systems*. 2018. https://doi.org/10.1016/j.future.2018.10.009.

[18] Lakshmanaprabu SK, K. Shankar, Deepak Gupta, Ashish Khanna, Joel J. P. C. Rodrigues, Plácido R. Pinheiro, Victor Hugo C. de Albuquerque. "Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering". *Complexity*, 2018: 1-9. https://doi.org/10.1155/2018/3569351.

[19] K. Shankar, Lakshmanaprabu S.K, Deepak Gupta, Andino Maseleno, Victor Hugo C. de Albuquerque. Optimal Features Based Multi Kernel SVM Approach for Thyroid Disease Classification. The Journal of Supercomputing, 2018. https://doi.org/10.1007/s11227-018-2469-4.

[20] Lakshmanaprabu SK, K. Shankar, Ashish Khanna, Deepak Gupta, Joel J. P. C. Rodrigues, Plácido R. Pinheiro, Victor Hugo C. de Albuquerque. Effective Features to Classify Big Data using Social Internet of Things. IEEE Access, 6 (2018): 24196-24204. https://doi.org/10.1109/ACCESS.2018.2830651.

[21] Andino Maseleno, Alicia Y.C. Tang, Moamin A. Mahmoud, Marini Othman, Suntiaji Yudo Negoro, Soukaina Boukri, K. Shankar, Satria Abadi, Muhamad Muslihudin. "The Application of Decision Support System by Using Fuzzy Saw Method in Determining the Feasibility of Electrical Installations in Customer's House". International Journal of Pure and Applied Mathematics, 119. 16 (2018): 4277-4286.

[22] Muhammad Muslihudin, Risma Wanti, Hardono, Nurfaizal, K. Shankar, Ilayaraja M, Andino Maseleno, Fauzi, Dwi Rohmadi Mustofa, Muhammad Masrur, Siti Mukodimah, "Prediction of Layer Chicken Disease using Fuzzy Analytical Hierarcy Process", International Journal of Engineering & Technology, 7. 2.26 (2018): 90-94.

[23] Eka Sugiyarti, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, K. Shankar, Andino Maseleno, "Decision Support System of Scholarship Grantee Selection using Data Mining", International Journal of Pure and Applied Mathematics, 119.15 (2018): 2239-2249.

[24] Tri Susilowati, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, K. Shankar, Andino Maseleno, Anis Julia, Sucipto, "Determination of Scholarship Recipients using Simple Additive Weighting Method", International Journal of Pure and Applied Mathematics, 119. 15 (2018): 2231-2238.

[25] K. Shankar. "Prediction of Most Risk Factors in Hepatitis Disease using Apriori Algorithm", Research Journal of Pharmaceutical, Biological and Chemical Sciences, 8. 5 (2017): 477-484.

[26] Nur Aminudin, Eni Sundari, K. Shankar, P. Deepalakshmi, Fauzi, Rita Irviani, Andino Maseleno. "Weighted Product and Its Application to Measure Employee Performance", International Journal of Engineering & Technology, 7. 2.26 (2018): 102-108. https://doi.org/10.14419/ijet.v7i2.27.11574.