Original Article

Stacked Feature Selection and C5.0 Classification Model with Tsallis Entropy for Medical Dataset

Dr.K.V. Uma[1], P. Juhi Padmaja[2], D. Vinoodhini[3]

[1]Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu. E-mail: kvuit@tce.edu

[2]Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu.

[3]Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu.

Abstract

Feature selection is considered to be one of the important tasks in data mining. It identifies the subset of features from the dataset that are best related to the response variable. Here, the five different medical dataset such as Pima diabetes, Liver, Hepatitis, Chronic kidney disease and Breast cancer are considered. The optimal subset of features for the various dataset is obtained by the intersection of top 'n' features returned by feature selection algorithms such as CMIM, JMI, mRMR, CFS, Boruta and SVM-RFE. Then, the method of C5.0 algorithm with Tsallis entropy and Association function is tested with these top n features selected. Accuracy value obtained for the proposed method is of 61% for Pima diabetes dataset, 85% for Liver dataset, 95% for Hepatitis dataset, 99.5% for Chronic Kidney disease dataset and 97% for Breast cancer dataset. The performance measures show that the proposed method works better than that of SVM, Naïve Bayes, KNN and Random forest.

Keywords: Feature Selection, Entropy, C5.0, Classification.

DOI: 10.47750/pnr.2022.13.S03.065

INTRODUCTION

When building a predictive model, feature selection is the process of minimizing the number of input variables. Reducing the number of input variables is desirable because it reduces the computational load of modeling and in some cases improves the performance of the model. The relationship between each input and the target is statistically evaluated and the input with the strongest relationship to the target is selected. The choice of statistical scale depends on the data types of both the input and output variables, but these methods can be fast and successful. As a result, it can be difficult for machine learning practitioners to choose an acceptable statistical measure of a dataset when performing filter-based feature selection.

RELATED WORK

Alzubi *et al.* (2018) proposed a feature selection method that combines Conditional Mutual Information Maximization (CMIM) and Support vector Machine recursive feature elimination for the Single Nucleotide Polymorphisms SNPs dataset. The objective of this method was to determine the important SNPs of the whole genome that distinguish the complex disease with healthy ones. Chormunge *et al.* (2018) described an important technique that reduces the issues of dimensionality in data mining task. The irrelevant features in the dataset are identified and removed by k-means clustering method. Then correlation measure was applied in each cluster to identify the non-redundant features. Ayar & Sabamoniri (2018) proposed a hybrid model that encompasses genetic algorithm for selecting the features for cardiac arrhythmias based on ECG data. Then the selected features are classified using C4.5 model. With the help of Map reduce framework, a novel gene feature clustering model was proposed for the document classification of PubMed and Medline documents (Bikku et *al.* 2018).Fuzzy Mutual Information-based Feature Selection combined with the method of Non-Dominated solution (FMIFS-ND) selects the features of the dataset by means of feature-class relation and feature-feature fuzzy mutual information. Then the selected features were classified using KNN (Hoque *et al.* 2016). Sasikala *et al.* (2016) proposed a Multi-Filtration Feature Selection (MFFS) comprises of four different stages such as feature extraction, feature subset selection, feature ranking and classification. Variance coverage was adjusted and thus the classifier provides better accuracy. Both low ranking and high ranking miRNA expression profiles were retrieved using two different methods. High ranking miRNAs were selected using correlation-based feature selection. Different measures such as chi-square test, attributes information gain, gain ratio and Pearson's correlation method were used to select m:n feature

subset (Li *et al.* 2014).Some of the other works related to feature selection was a meta heuristic approach of Binary Artificial Bee Colony also known as BABC algorithm for selecting the important features for disease identification (Subanya *et al.* 2015), ImpCHI method for classification of Arabic documents (Bahassine *et al.* 2018), LW-index along with the Sequence Forward Search algorithm (SFS-LW) (Liu *et al.* 2017), graph clustering method (Gangurde & Metre 2015), hybrid model of CFS and Filter subset evaluation for heart disease prediction (Peter & Somasundaram 2012), filter and wrapper approaches along with Particle Swarm Optimization (PSO) for the different medical datasets (Harb & Desuky 2014), F-Score and Recursive Feature Elimination for thyroid disease (Pavya & Srinivasan 2017).

Jin *et al.* (2009) proposed ID3 classification algorithm that uses the Association function. This method overcomes the shortcomings of ID3 algorithm. The need of each attribute present in the dataset is calculated by using Association function. It represents the relation between all the elements and their attributes in the dataset. The importance of attribute is combined with the Information gain measure. If the attribute in the dataset D is represented by A and C represents the category attribute present in the dataset, the relation degree function between the attribute A and Category attribute can be defined as in Equation (1).

$$F(A) = \frac{\sum_{i=1}^{n} |x_{i1}| - |x_{i2}|}{n} \quad (1)$$

where, x_{ij} in the formula denotes the attribute *A* of the dataset D takes the i^{th} value, and the category attribute *C* considered will assign the sample number of the j^{th} value, where *n* is the considered as the total number of values that a attributes *A* takes. Once the relation degree is estimated for each and every attribute present in the dataset, perform normalization. Its performed as that if there are *m* numb*er* of attributes present in the dataset and the attribute relation degree function of each and every attribute values are defined by *AF*(1), *AF*(2),…, *AF*(*m*), Equation (2) is defines the normalization which is defined as

$$V(k) = \frac{AF(k)}{AF(1)+AF(2)+ \ldots\ldots+AF(m)} \quad (2)$$

where, k ranges from 1 to m. Then this association function can be used during Gain calculation. Attribute importance combined with the Information gain measure has reduced the computational complexity of the algorithm.

Ming *et al.* (2009) proposed an improved version of decision tree algorithm that combines the Taylor formula with information entropy. The entropy was calculated using Maclaur in formula. Each attribute was assigned a value of N, which is used to balance the uncertainty of dataset. Rahman & Hasan (2011) proposed a decision tree classification technique that uses Gain ratio to find the best split attribute for the construction of tree. It maintains the uniformity of the attribute values. Gain ratio is defined as given in Equation (3)

$$Gain\ Ratio = \frac{Information\ gain}{Split\ Information} \quad (3)$$

Split information value is based on the column sums. Since the hospital surveillance data is used, it contains more number of attributes. Not all the attributes are needed for analysis. The relevant attributes are selected by using entropy measure. Then it is applied for decision tree construction.

Ashok *et al.* (2012) suggested that the information in the dataset, measured using different concepts will make the data mining algorithms to get desired results with more accuracy. In order to characterize the amount of information as a probability distribution, entropy measure is used. Hence the entropy measure and its various generalizations are mostly used in the field of statistics, computer science and communication theory. Here, Renyi entropy was used for extraction of information in ID3 algorithm for Census data.Mathur *et al.* (2012) proposed the use of Havrda and Charvat entropy instead of Shannon entropy in ID3 algorithm. The depth of the decision tree constructed need to be minimized. For that select the attribute with maximum entropy reduction as split attribute. Jayakameswaraiah & Ramakrishna (2015) proposed a Classification with Clustering method called as Improved ID3 and TkNN Clustering for decision making. The method was applied for Car dataset. It is essential to analyze the features of car in a short duration of time such that the manufacturer, customer and seller choose the right product. Decision tree is constructed that uses Shannon entropy form determining the best attribute. Then TkNN graph was constructed for the instances. Similarities on each edge are initialized as given in Equation (4) and it is normalized.

$$w_{iz} = \exp\left(\frac{||x_1 - x_2||^2}{2\sigma^2}\right) \quad (4)$$

Then compute the label for set prediction matrix and predict the label for instances that are unbalanced. Liang *et al.* (2015) proposed an improved ID3 algorithm that mainly considers the weightage of the attribute present in the dataset. The dataset is divided into subsets. The relation between the attributes in the subset and classification attribute was identified and combined with the Information gain for choosing the best split attribute.

Prasanthi *et al.* (2016) proposed Over Sampled ID3 (OSID3) algorithm for handling imbalance datasets. Initially noisy and outlier instances are removed from the minority and majority subset. Then oversampling of minority subclass and under sampling of majority subclass was performed to overcome the imbalance nature of the dataset. Then the tuned dataset was applied to ID3 algorithm.

Jose Suganya & Balasubramanian (2016) proposed the use of Attribute important entropy in the ID3 algorithm and applied it for Appraisal dataset. Attribute Important Entropy (AIE) of an Attribute A is calculated as in Equation ()

$$\boldsymbol{AIE(A) = \sum_{i=1}^{N} \frac{((S1_i + S2_i + \cdots + Sn_i}{S} + \rho(A))I(S1_i + S2_i + \cdots + Sn_i)} \quad (5)$$

where, ρ defines the attribute importance. Gain of an attribute is calculated as in Equation (6)

$$\boldsymbol{Gain(AIE) = I(S1_i + S2_i + \cdots + Sn_i) - AIE(A)}\ (6)$$

Here the attribute with highest $Gain(AIE)$ was chosen as splitting attribute.

From the survey it is inferred that feature selection algorithm is considered to be important in achieving higher accuracy for the Classifier. Hence in the proposed work, different feature selection algorithms are considered. The optimal features are selected by the intersection of features obtained from different algorithms. Then it is applied to the C5.0 classifier model that uses Tsallis entropy and Association function.

METHODOLOGY

The figure.1 shows the system design.

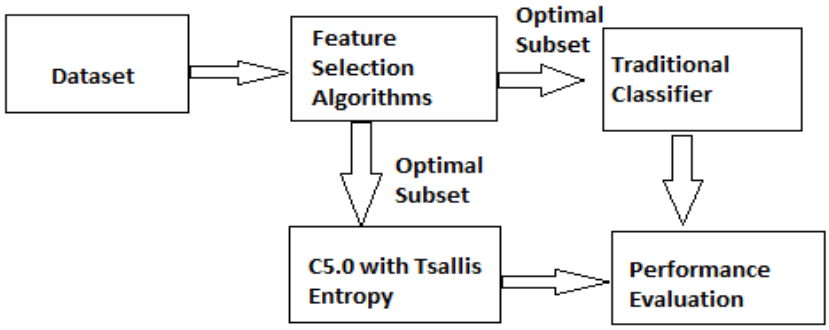

Figure 1; Feature Selection and C5.0 Classifier with Tsallis Entropy

The steps involved in the system design are given as follows.

Step 1: The input dataset contains missing values for the attributes and redundant tuples. The missing values present in the dataset are replaced with mean and the redundant tuples are removed.

Step 2: The preprocessed dataset is fed as an input to different feature selection algorithms such as CMIM, JMI, mRMR, Correlation Feature Selection, SVM-RFE and Boruta. The top n features are selected for a dataset using different algorithms. The optimal subset is obtained by the intersection of features selected by different algorithms.

Step 2a. Continuous Mutual Information Maximization (CMIM)

The feature that has a higher value of CMIM indicates that the attribute X_n is more relevant to the target variable Y and is highly to be considered for further classification than that of other feature X_j where j ∈ S.

The selection criterion for the feature is given in the Equation (7)

$$CMIM(X_n) = \min_{jEs} I(X_i; \frac{Y}{X_j})\ (7)$$

Step 2b SVM-RFE (Support Vector Machine-Recursive Feature Elimination)

SVM-RFE is type of wrapper method that applies backward function removal. Find a subset of features that lead to maximizing class separation margins. It starts with the entire set of features and recursively removes the least important features for the predictor in the backward removal process.

SVM-RFE algorithm consists of four steps.

1. SVM model is trained with the training dataset such that it fits to the simple linear SVM.
2. Order features of the resulting classifier using the weights
3. Remove the features that have the smallest weight
4. Repeat the steps with the training set for the remaining set of features

SVM-RFE

Input: *Training patterns* $[x_1, x_2, \ldots, x_k, \ldots, x_l]^T$
class labels $[y_1, y_2, \ldots, y_k, \ldots, y_l]^T$
Output: *Feature ranked list(r)*

Do the initialization of the features subset where s=[1,2,…,n]
Initialize the ranked list as empty $r = []$
For each feature

$$X = X_0(:, s)$$
$$\alpha = SVM\ train(X, y)$$
$$w = \sum_k \alpha_k y_k x_k$$
$$c_i = (w_i)^2$$

End

$$f = argmin(c)$$

Update the feature ranked list $r = [s(f), r]$
Eliminate the feature with small rank

$$s = s[1: f-1, f+1: length(s)]$$

Step 2c The Boruta Algorithm

This algorithm is suitable for classification and regression problems. Consider a multivariable relationship. This is an improvement in measuring the importance of random forest variables, which is a very common method of variable

selection.

Steps in Boruta Algorithm

Initially this algorithm will be generating the shadow features by the process of duplicating the dataset and shuffling the values present in each attribute of the dataset. Then it trains the classifier and identifies the mean decrease in impurity for each feature of the dataset.

Check whether the real features have higher importance. This is done by comparing the Z-score of the original and shadow features. Z-score is calculated using the equation (8).

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (8)$$

where, $\bar{x}$ denotes the sample mean and S denotes the standard deviation. Iterate the process repeatedly by comparing the Z-score of the shuffled copies and the original features and determine the features that are important.

Step 2d mRMR (Minimum Redundancy and Maximum Relevance Feature Selection)

mRMR is one of the feature selection algorithm which select the features that has high correlation with classification variable which is similar to that of CMIM algorithm.

Steps in mRMR algorithm

1. The features that are more relevant to the class attribute C are identified based on Mutual information difference, Mutual information Quotient and F-test etc.
2. Then the selected features are correlated and the redundant features are identified.
3. Combine relevance and redundancy to form the final feature subset.

First use mrMR to generate a short feature pool and then use a wrapper to get the least redundant feature set more accurately. Feature selection helps improve the performance of your learning model by removing most irrelevant and redundant features from your data.

Step 2e JMI

Mutual information is a measure between two (possibly multidimensional) random variables XX and YY, which quantifies the amount of information obtained by one random variable and the other. Mutual information is provided by Equation (9).

$$I(X;Y) = \int X \int Y \, p(x,y) \log(x,y) p(x) p(y) dxdy \quad (9)$$

Where, p(x,y) is considered to be the joint probability density function of XX and YY and, p(x)p(x) and p(y)p(y) are considered to be the marginal density functions.

Step 2f Correlation Feature Selection Algorithm (CFS)

CFS is an algorithm which contains the evaluation formula that follows the heuristic search strategy and uses an appropriate correlation measure.

The features that are commonly selected by all the above methods are considered for further processing.

Step 3a: The dataset with relevant features obtained is fed as an input to traditional classification algorithms such as KNN, SVM, Naïve Bayes, C4.5, C5.0 and Random Forest

Step 3b: Also, it is fed as an input to the proposed classifier C5.0 algorithm that uses Tsallis entropy and Association function.

Step 4: Performance evaluation of the different classifier models and proposed model is done. The most recent version of Quinlan's algorithm is C5.0 included the function of boosting and cross-validation. Advantage of C5.0 algorithm over C4.5 is that it is faster, memory efficient, builds small decision trees and reduce noise (winnowing). Important component in decision tree construction is to estimate the split attribute at each internal node of structure of the tree. C5.0 algorithm uses Information Gain Ratio in order to identify the best split attribute for the tree construction. Information gain ratio will measure the reduction in the value of entropy for the data partitioned by the split.

Let, T be the training dataset having the tuples belonging to any of the k classes $\{C_1, C_2, \ldots, C_k\}$. The information amount or the randomness of the dataset is measured as given in Equation (2.9).

$$\boldsymbol{Info(T)} = -\sum_{j=1}^{k} \frac{\boldsymbol{freq(C_j)}}{|\boldsymbol{T}|} \times \boldsymbol{log_2} \frac{\boldsymbol{freq(C_j,T)}}{|\boldsymbol{T}|} \quad (10)$$

where, $freq(C_j, T)$ takes the value equal to the number of tuples in the dataset T that belongs to class C_j, and |T| is considered as the total number of observations available in the dataset T. For an attribute X, that partitions the dataset T into n number of outcomes, the total information available after applying the attribute X is calculated as given in Equation (2.10)

$$Info_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} \times Info(T_i) \quad (11)$$

The information gained by splitting the dataset T using the attribute X is defined as given in Equation (12)

$$Gain(X) = Info(T) - Info_X(T) \quad (12)$$

Information gain selects the attribute with the maximum value. To compensate with more number of splits, the measure Gain(X) is normalized by SplitInfo(X) as given in Equation (13)

$$Split\ Info(X) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times log_2 \frac{|T_i|}{|T|} \quad (13)$$

Now the Gain Ratio is calculated as given in Equation (14).

$$Gain\ ratio\ (X) = \frac{Gain(X)}{SplitInfo(X)} \quad (14)$$

The dataset T is recursively split unless the Gain ratio of the attribute is maximized at each node of the tree. The procedure terminates when the leaf node contains tuples of a single class or there is zero gain in information. The decision tree generated must be pruned.C5.0 algorithm uses error-based pruning approach in order to remove the features that are considered to be spurious. (Wang et al. 2016)

Entropy with powers of probability will have such control. Tsallis entropy has the powers of probability. Tsallis entropy $S_q(X)$ is generalization of Shannon entropy done by adding an adjustable parameter called as q. It is defined by as in

Equation (15).

$$S_q(X) = \frac{1}{1-q}\left(\sum_{i=1}^{n} p(x_i)^q - 1\right), q \in R \quad (15)$$

where, X is the random variable having value (x_1,x_2,…….,x_n) and p(x_i) is considered as the probability of x_i. (Wang *et al*. 2017).

EXPERIMENT AND RESULTS

The different Medical dataset such as Pima Diabetes, Hepatitis, Liver, Kidney and Breast cancer are collected from the UCI Repository and Kaggle. Since the dataset contains missing values, they are replaced with Mean values. The table 1 shows the different dataset along with the name of existing features and the top n features obtained using the intersection of features obtained through the execution of feature selection algorithms such as CMIM, JMI, MRMR, Correlation Feature Selection, SVM-RFE and Boruta.

Table 1. Dataset, Existing Features and top N Features

Dataset	Existing features	Top-n features
Pima diabetes	Pregnancies, glucose, blood pressure, skin thickness, insulin, bmi, diabetes pedigree function, age, outcome[9 features]	Pregnancies, glucose, blood pressure, skin thickness, insulin, age, outcome [7 features]
Hepatitis	Class, age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver_big, liver_firm, spleen_pa, ascites, spiders, varices, bilirubin, alk_phos, sgot, albumin, protime, histology [20 features]	Fatigue, age, ascites, spiders, alk phosphate, sex, bilirubin, mmalaise, histology, sgot[10 features]
Liver	Age, gender, tot_bilirubin, direct_bilirubin, tot_proteins, albumin,ag_ratio,sgpt,sgot,alkphos,is_patient[10 features]	Age, direct_bilirubin, albumin, ag_ratiosgpt, tot_bilirubin, direct_bilirusbin [6 features]
Kidney	Age, blood pressure, specific gravity, albumin, sugar, redblood cell, pus cell, puss cell clumps, bacteriablood glucose random, blood urea, serum creatinine, sodium, potassium, haemoglobin, packed cell, white blood cell count, red blood cell count, hypertension, diabetes mellituscoronary artery disease, appetite, pedeledema, anemia, class [26 features]	albumin, sugar, age, bacteria, serum creatinine, potassium, diabetes mellitus, pus cell colume, appetite, pus cell, hypertension, coroonary artery disease, red blood cells[14 features]
Breast cancer	ID, idnumberdiagnosis, Radius_Mean, Texture_Mean, Perimeter_Mean, Area_Mean, Smoothness_Mean, Compactness_Mean, Concavity_ Mean, Concave Points_Mean, Symmetry_Mean, Fractal_Dimension_Mean, Radius_Se, Texture_Se, Perimeter_SE, Area_SE, Smoothness_SE, Compactness_SE, Symmetry_SE, Fractal_Dimension_SE, Radius_worst,Texture_ worst, Perimeter_worst, area_worst, smoothness_worst, compactness_ worst, concavity_worst, concave points_worst, symmetry_worst, Fractal_Dimension_worst [32 features]	Radius_mean, texture_mean, concavity_seconcave points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worstarea_ worst smoothness_worst, compactness_worst, concavity_worst, concave oints_worst, symmetry_worst fractal_dimension_worst [16 features]

Table 2 depicts the accuracy obtained for the various Classification algorithms namely SVM, Naïve Bayes, KNN, Random forest, C4.5 and C5.0 executed with full dataset and the dataset with top n features.

Table 2. Accuracy of different Classifier Executed with Full Dataset (F) and with Top N Features (R)

Dataset	SVM		C50		C45		Naïve Bayes		KNN		Random Forest	
	Full	Red	Full	Red	Full	Red	Full	Red	Full	Red	Full	Red
Pima Diabetes	50	51.0	55.3	59.7	52	53.6	52.3	54.4	52.15	54.15	48.48	50
Liver	68.9	69.2	79	81	78.1	80	47.2	48.3	68.59	75.86	50.3	56.7
Hepatitis	81.5	80.4	87	89	86	88.9	50.8	52.6	80.47	82.9	70	76.7
Chronic Kidney Disease	68.3	66.3	99.3	99.2	99.4	99.5	44.85	86.4	64.1	63.87	99.2	99.1
Breast Cancer	67.8	68.9	85	88.6	96.4	96.3	65	66.1	74.72	71.49	55.6	58.6

In most of the cases, dataset with optimal subset of features provides more classification accuracy. Next, the proposed C5.0 classifier with Tsallis entropy and Association function is executed with the optimal dataset and original dataset. Table.3 shows the accuracy obtained.

Table 3. Accuracy of C5.0 Algorithm-using Tsallis Entropy

Dataset	Full dataset	Top n features
Pima diabetes	59.6	61
Liver	80.1	85
Hepatitis	92	95
Chronic kidney disease	91	99.5
Breast cancer	93	97

Figure 2 shows the comparison of results obtained through various algorithms with the proposed methodology.

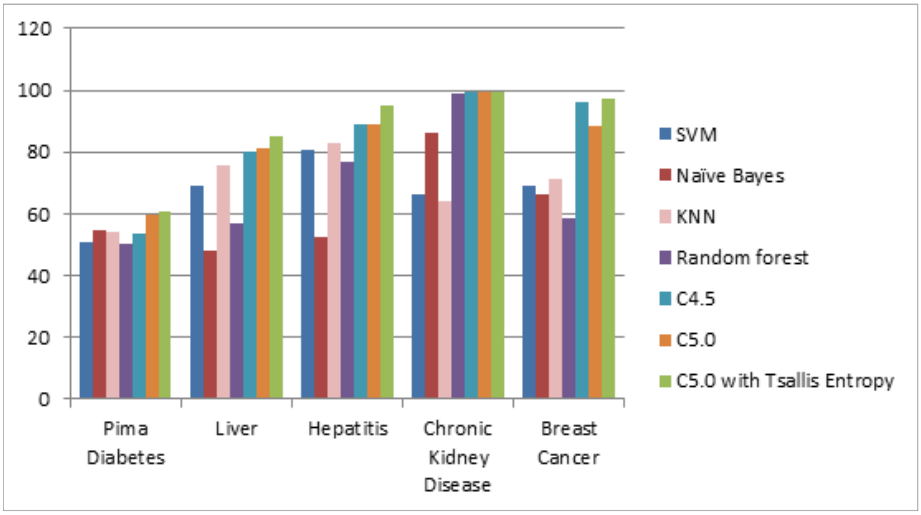

Figure 2; Comparison of different Classifiers with Proposed Method of C5.0 with Tsallis Entropy

From the above graph it is inferred that for medical dataset, the proposed algorithm C5.0 with Tsallis entropy and Association function provides more accuracy by considering top n features. Hence feature selection is considered as a important part in computing the accuracy of the classifier.

On an average, for the various algorithm of feature selection the time complexity is in the order of $O(KMN)$, where N denotes the number of instances from the dataset with M features and top K selected features. Time complexity of modified C5.0 algorithm will be in the order of $O(hn(kN + NlogN)$, where h is the height of the tree; 'N' to be the total number of observations present in the dataset, 'n' is the number of features in the dataset and k is the number of class labels. Then the total time taken for the proposed method will be in the order of $O(KMN + hK(kN + NlogN))$.

Conclusion

The best subset of features arrived from the various feature selection algorithms are applied to the modified C5.o algorithm using Tsallis entropy works well for sample Medical dataset. The combinational approach of hybrid mining algorithms helps in achieving high accuracy for prediction model. Thus the obtained result has higher accuracy of prediction when compared to the traditional methodology. Thus, the decision tree generated will help to understand the characteristics of dataset in a better way.

References

Alzubi, R, Ramzan, N, Alzoubi, H & Amira, A 2018, 'A hybrid feature selection method for complex diseases SNPs', IEEE Access, vol. 6, pp. 1292-1301.

Ayar, M & Sabamoniri, S 2018, 'An ECG based feature selection and heartbeat classification model using a hybrid heuristic algorithm', Informatics in Medicine Unlocked, vol. 13, pp. 167-175.

Bahassine, S, Madani, A, Al-Sarem, M & Kissi, M 2018, 'Feature selection using an improved Chi-square for arabic text classification', Journal of King Saud University-Computer and Information Sciences, in Press.

Bikku, T, Nandham, SR & Akepogu, AR 2018, 'A contemporary feature selection and classification framework for imbalanced biomedical datasets', Egyptian Informatics Journal, vol. 19, no. 3, pp. 191-198.

Chormunge, S & Jena, S 2018, 'Correlation based feature selection with clustering for high dimensional data', Journal of Electrical Systems and Information Technology, vol. 5, no. 3, pp. 542-549.

Gangurde, HD & Metre, KV 2015, 'Mining of high dimensional data using feature selection', International Journal of Computer Science and Mobile Computing, vol. 4, no. 6, pp. 901-906.

Harb, HM & Desuky, AS 2014, 'Feature selection on classification of medical datasets based on particle swarm optimization', International Journal of Computer Applications, vol. 104, no. 5, pp. 14-17.

Hoque, N, Ahmed, HA & Bhattacharyya, DK 2018, 'A fuzzy mutual information-based feature selection method for classification', Fuzzy Information and Engineering, vol. 8, no. 3, pp. 355-384.

Li, F, Piao, M, Piao, Y, Li, M & Ryu, KH 2014, 'A new direction of cancer classification: Positive effect of low ranking MicroRNAs', Osong Public Health and Research Perspectives, vol. 5, no. 5, pp. 279-285.

Pavya, K & Srinivasan, B 2017, 'Feature selection algorithms to improve thyroid disease diagnosis', in: Proceedings of International Conference on Innovations in Green Energy and Healthcare Technologies,

Coimbatore, pp. 1-5.

Peter, TJ & Somasundaram, K 2015, 'Study and development of novel feature selection framework for heart disease prediction', International Journal of Scientific and Research Publications, vol. 2, pp. 1-7.

Sasikala, Balamurugan, SA & Geetha, S 2016, 'Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical dataset', Applied Computing and Informatics, vol. 12, no. 2, pp. 117-127.

Subanya, B & Rajalaxmi, R 2015, 'A novel feature selection algorithm for heart disease classification', International Journal of Computational Intelligence and Informatics, vol. 4, no. 2, pp. 117-124.

Uma K.V & Appavu alias Balamurugan S 2020. 'C5.0 Decision Tree Model using Tsallis Entropy and Association Function for General and Medical Dataset'. Intelligent Automation and Soft Computing, vol. 26, No 1, pp 61-70.

Ibrahim, S., & Rababah, A. (2022). Decomposition of Fourth-Order Euler-Type Linear Time-Varying Differential System into Cascaded Two Second-Order Euler Commutative Pairs. Complexity, 2022.