Jestr

Journal of Engineering Science and Technology Review 10 (6) (2017) 38 - 43

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance

Sai Prasad Potharaju* and M.Sreedevi

Dept of CSE , K L University, Guntur, Andhra Pradesh, India

Received 23 October 2017; Accepted 7 December 2017

Abstract

Feature Selection (FS) is an imperative issue in data mining and machine learning. It is an inevitable task to shorter the number of features presented in the initial data set for better classification result, minimized computation time, and reduced memory consumption. In this article, a novel framework using Correlation Coefficient (CCE) and Symmetrical Uncertainty (SU) for selecting the subset of feature is proposed. The selected features are congregated into finite number of clusters by grading their CCE and comparing the SU values. In each cluster, a feature with maximum SU value is retained while remaining features in the same cluster are ignored. The proposed framework was examined with Ten(10) real time benchmark data sets. Experimental outcomes show that the proposed method is outruns than majority of conventional feature selection methods(Information Gain, Chi-Square, Gain Ratio, ReliefF) in accuracy. This method is tested using Tree Based, Rule Based, Lazy, and Naive Bayes learners.

Keywords: Feature Selection, Correlation Coefficient, Classification, Machine Learning, Symmetrical Uncertainty

1. Introduction

Data Mining is a popular research area in every domain including education, financial, health, security, marketing, etc. Classification, regression, clustering, and association rule mining (ARM) are the well known data mining techniques for different purposes. It is an extensive analytic technique to get the more insights of data which will be useful for better decision making.Data Mining stages includes Data Collection, Preprocessing, Data Mining, Interpretation, and visualization. Data can be collected from diversity of sources. Collected data has to be cleaned for better results, as it includes noisy, imbalanced labels, missing values, missing labels, and high dimensional features in it. After preprocessing, data mining techniques will be applied to create the model , then results will be interpreted and represented in different forms.

This research focused on classification task of data mining and high dimensional issue of preprocessing. A data set with more number of features/variables is called high dimensional data set. If a data set has more number of features, it may create different problem to the learning model. Firstly, all the features in the initial data set may not be useful.

Few features may be repeated and noisy. These repeated features does not contribute any thing instead it may confuse the learning model and reduce the efficacy of model. High dimensional data set requires more processing power and memory. This issue can be addressed by feature selection techniques.

The fundamental target of FS is to choose the most dominant features and ignore worthless features[1]. Noisy and duplicated features are unnecessary features, which need to reduced in FS process. If a classification accuracy can't be increased by inclusion of a feature, we can say, the feature is unnecessary. But, a critical query is how to get these unnecessary (redundant and noisy) features ?. There have been some existing studies available in order to give the solution to this query.

Filter and Wrapper are two commonly used FS modes. Filter mode of FS measures the worth of each feature and gives the grade to each feature, there by top 'K' worthily features can be selected for model creation[2]. Information Gain(IG), Chi-Square(Chi), Gain Ratio , and ReliefF (Rel)[3] are few filter based methods.In another direction, wrapper FS method is usually time taking approach, because it requires to combine some learning algorithms for selecting the best features[4]. During this process, features with lesser the accuracy by learning algorithms will be ignored from the initial data set. In this research, we focus on selecting best features that accelerate the classification accuracy using CCE and SU by congregating the features into several clusters.

Euclidean distance is commonly used metric in the clustering analysis as a similarity measurement. But, to find out the dependency between two random features, CCE could be more useful. In the research, instead of popular euclidean distance, we considered CCE to measure the dependency between two variables. As per the mutual information theory, if two variables are mutually dependent, we can consider only one among them for classification as they share common properties and gives almost same result.

How to form the cluster of features and select the best feature from each cluster is discussed in the next section. The proposed framework is tested with Tree Based(J48, Simple cart(sc), Rule Based(Jrip, Ridor), Lazy (IBK), and

*E-mail address: psaiprasadcse@gmail.com
ISSN: 1791-2377
doi:10.25103/jestr.106.06

Bayes (NB) learners over Ten (10) data sets. In second section, brief literature review and related work is discussed. In third section, methodology of proposed framework with example is discussed. In section four experimental procedure is illustrated . Results with discussion is given in fifth section. Concluding remarks with future suggestions are drafted in the final section.

2. Related Study

Filter and Wrapper methods of FS have been applied by many researchers for classification problems to select candidate subset(reduced feature set) that increase the performance of classifiers[5][6]. In addition to these two methods, embedded FS approach also have been applied for the classification problems as part of modeling process. SVM-RFE[7], LASSO[8], Random-Forest[9] are popular embedded FS techniques. Filter and Wrapper method is applied to solve the protein disordered region prediction problem which has 440 features[10]. First information gain and F-Score is applied, then wrapper method is applied to get the improved classification result.

In this current research, we compared the proposed technique with some of the existing feature selection methods (IG, Chi, GR, Rel). The concept of IG is based on the information theory which analyze the association between features and classes for discarding the duplicated features and the most unrelated features to the class[11]. In literature, there is another FS method proposed by Peng et,.al [12] which is also based on mutual information. This concept is based on MaxDep[13], It calculates the subset statistical dependency with the target class. It choose 'n' features that combinely have the highest dependency on the target class.

Maximum Correlation Information-Recursive Feature Elimination(MCI-RFE) method is proposed for integrated high dimensional protein data[14]. In that method, an importance of each feature is evaluated by maximizing the correlation information(MCI). Then MCI is combined with recursive feature elimination(RFE) to form an optimal feature set. This MCI RFE is highly competitive with Random Forest, ReliefF-RFE, and SVM-RFE.

To deal with complex problems associated with huge number of features in pattern recognition,FS has become an alternative task for many researchers to get the high performance[15]. Correlation-Based Selection (CFS) method is used in the literature for different purposes. CFS is used to forecast the electricity demand in Australia with linear regression, tree based models, and neural networks prediction algorithms using two years of time series load data[16]. FAST algorithm was proposed by authors used the concept of Correlation Coefficient and Symmetrical Uncertainty to get the best subset. FAST is a clustering based algorithm, works in two phases.In first phase graph theory clustering method is applied to form the feature into clusters, in the second stage prims algorithm is applied to select most representative features [17].

This current research also mainly based on two statistical measurement used in the FAST in different way. Those are: Correlation Coefficient(CCE) and Symmetrical Uncertainty(SU). CCE is used in this present work to know the relationship(Weight) between two variables and form the cluster of features. SU is used to decide the threshold value of weight and also to select the best feature in each cluster. The procedure to measure the weight of feature discussed in methodology section. Out of 'n' observations, CCE of two random variables X and Y can be derived as below equation

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

If 'r' value is close to 1, we can say there is a strong dependency between X and Y. If 'r' value is 0, then we can say there is no relationship between X and Y. In our research, we considered positive 'r' value to measure the weight of the feature.

Symmetrical Uncertainty(SU) can defined as below equation

Symmetrical Uncertainty (SU) =
2*IG/(H(F1)+H(F2))
IG is Information Gain
H(F1) is Entropy of F1
H(F2) is Entropy of F2

In the next section proposed methodology is discussed with an example

Algorithm:

1. Find out the Symmetric Uncertainty (SU) value of each feature, such that all features will be in descending order of its SU value.
2. Choose the middle feature's SU value as Threshold (T).
3. Generate the Correlation Coefficient Symmetrical matrix ($CCE(X_i,Y_i)$) of initial data set .
4. Transform the above matrix to weighted binary matrix (WB) as per the below steps

```
for(i=1 to n)
  for(j=1 to n)
     if(CCE(Xi,Yi)>T)
     WB(Xi,Yi)=1
     else
     WB(Xi,Yi)=0
   End
 End
```

5. Calculate the total weight of each feature W(F) as per below steps.

```
for(i=1 to n)
   for(j=1 to n)
   W(Fi)=      WB(Xi,Yi)
    End
End
```

6. Group the features which are having same weight(W(F))

$Cluster_i=\{F_{i1}, F_{i2},...F_{ik}\}$ /* i is the cluster id, increment i by 1 until all features are formed */

7. Choose the best feature (feature which has maximum SU value) from each cluster and form the final candidate subset

```
for(i=1 to last cluster)
   Fi= MAX SU(cluster i)
   Candidate Feature set (CFS)<- Fi
End
```

3. Methodology

The objective of proposed methodology is to find out the candidate feature set which can boosts-up the classification performance. Proposed methodology is, as per the following algorithm.

According to the above algorithm, here an example to form the candidate feature set is given below.

Example

Assume there are ten features (a, b, c, d, e, f, g, h ,i, j) in a sample data set .

1. SU value of each feature is given in below Table.1

Table 1. SU value of sample data set features

SU	Rank	Fid
.19	10	j
.19	8	h
.19	7	g
.18	9	i
.15	**2**	**b**
.09	1	a
.07	4	d
.06	3	c
.06	5	e
.02	6	f

2. Threshold (T) =.15 , as 'b' is the middle feature . If top/bottom feature SU values is considered as T, number of clusters to be formed will be varied . To optimize the number of clusters, middle feature SU value is considered.

3. Correlation Coefficient Symmetrical matrix ($CCE(X_i,Y_i)$) of the data set is given in below Table 2.

Table 2. Correlation Coefficient Symmetrical matrix ($CCE(X_i,Y_i)$)

Feature Id	a	b	c	d	e	f	g	h	i	j
a	1	-0.08	-0.03	-0.15	-0.16	-0.05	-0.11	0.31	-0.28	0.29
b	-0.08	1	0.05	0.09	-0.11	-0.04	-0.13	-0.28	0.21	-0.37
c	-0.03	0.05	1	-0.07	0.05	-0.01	0.27	-0.1	0.12	-0.07
d	-0.15	0.09	-0.07	1	0.29	0.01	0.09	-0.23	0.29	-0.31
e	-0.16	-0.11	0.05	0.29	1	0.12	0.23	-0.12	0.56	-0.27
f	-0.05	-0.04	-0.01	0.01	0.12	1	0.01	0.04	0.03	-0.03
g	-0.11	-0.13	0.27	0.09	0.23	0.01	1	0.05	0.27	-0.14
h	0.31	-0.28	-0.1	-0.23	-0.12	0.04	0.05	1	-0.43	0.46
i	-0.28	0.21	0.12	0.29	0.56	0.03	0.27	-0.43	1	-0.47
j	0.29	-0.37	-0.07	-0.31	-0.27	-0.03	-0.14	0.46	-0.47	1

4. Transformed Weighted binary matrix of above matrix and weight of each feature is given below Table 3.

Table 3. Weighted binary matrix

Feature Id	a	b	c	d	e	f	g	h	i	j	Weight	Feature
a	1	0	0	0	0	0	0	1	0	1	3	a
b	0	1	0	0	0	0	0	0	1	0	2	b
c	0	0	1	0	0	0	1	0	0	0	2	c
d	0	0	0	1	1	0	0	0	1	0	3	d
e	0	0	0	1	1	0	1	0	1	0	4	e
f	0	0	0	0	0	1	0	0	0	0	1	f
g	0	0	1	0	1	0	1	0	1	0	4	g
h	1	0	0	0	0	0	0	1	0	1	3	h
i	0	1	0	1	1	0	1	0	1	0	5	i
j	1	0	0	0	0	0	0	1	0	1	3	j

5. Form the clusters and select the best feature in each cluster. For doing this, initially all the features are sorted as per its weight. Sorted list of feature and weight is given in below Table 4.

Table 4. Sorted list of feature and weight

Cluster Id	Weight	FID	Selected Feature From each cluster
1	**1**	**f**	**f**
2	**2**	**b**	**b (As SU value of 'b' is maximum than other features in cluster)**
	2	**c**	
	3	**a**	**j**
3	**3**	**d**	**(As SU value of 'j' is maximum than other features in cluster)**
	3	**h**	
	3	**j**	
4	**4**	**d**	**d (As SU value of 'd' is maximum than other features in cluster)**
	4	**e**	
	5	**e**	**i**
5	**5**	**i**	**As SU value of 'i' is maximum than other features in cluster**

6. Form the final candidate feature set (CFS)

CFS= {f, b, j, d, i}

4. Experiment

To examine the proposed framework, ten (10) real-time benchmark data sets are taken into consideration. The list of data sets and their brief description is given in Table 5.

Table 5. Data sets description

Data set ID	Name of the Data Set	# Instances	# Features	# Class
1	Ionosphere	351	34	2
2	Dermatology	366	34	6
3	Biodegradation	1055	41	2
4	Cardiotocography	2126	22	3
5	Lung Cancer	33	56	3
6	Libras Movement	360	90	15
7	Connectionist Bench(Sonar)	208	60	2
8	Spambase	4601	57	2
9	Breast Cancer(WDBC)	569	30	2
10	Musk (V 2)	476	166	2

The framework is tested using open source machine learning tool WEKA. We used 10-fold cross validation for all the data sets. After applying this method, number of features are minimized. The number of features generated as a result of proposed method for each data set is given in Table 6.

Table 6. Features formed by proposed method

Data set ID	Name of the Data Set	# Features in Original Data set	#Features formed by Proposed Method (S)
1	Ionosphere	34	13
2	Dermatology	34	13
3	Biodegradation	41	23
4	Cardiotocography	22	12
5	Lung Cancer	56	15
6	Libras Movement	90	21
7	Connectionist Bench(Sonar)	60	28
8	Spambase	57	16
9	Breast Cancer(WDBC)	30	15
10	Musk (V 2)	166	54

To measure the strength of the proposed method, Top 'S' number of features derived by existing methods are selected. For calculating correlation coefficient value between the features of every data set, R statistical programming is used. Symmetrical Uncertainty and performance of classifiers with the selected features is measured using WEKA.

5. Results and Discussion

In this section, classification performance of proposed and existing methods using Jrip, Ridor, J48, Simple Cart, IBk, Naive
Bayes classifiers with different data sets are presented with brief discussion.

Table 7. Result Analysis

Dataset ID :1, Name: Ionosphere

	Jrip	**Ridor**	**J48**	**SC**	**NB**	**IBK**	**Avg**
IG	91.45	*89.74*	*91.73*	*87.46*	*84.9*	*86.89*	*88.70*
Chi	90.31	90.88	*91.16*	*88.88*	*88.03*	*88.03*	*89.55*
GR	90.02	*89.45*	*90.59*	*88.6*	*86.03*	90.31	*89.17*
Rel	91.16	91.16	94.01	91.45	90.02	89.17	91.16
Proposed	88.6	**90.31**	**92.02**	**90.02**	**89.17**	**88.88**	**89.83**

Dataset ID :2, Name: **Dermatology**

	Jrip	*Ridor*	*J48*	*SC*	*NB*	*IBK*	*Avg*
IG	*79.23*	*82.24*	*80.87*	*80.32*	*83.33*	*82.51*	*81.42*
Chi	*83.6*	*83.33*	*83.06*	*83.33*	*84.15*	*85.24*	*83.79*
GR	*83.6*	*83.33*	*83.06*	*83.33*	*84.15*	*85.24*	*83.79*
Rel	*76.77*	*79.23*	*77.86*	*77.32*	*81.42*	*81.14*	*78.96*
Proposed	**89.07**	**91.25**	**91.8**	**89.89**	**94.53**	**92.34**	**91.48**

Dataset ID :3, Name: Biodegradation

	Jrip	**Ridor**	**J48**	**SC**	**IBK**	**NB**	**Avg**
IG	82.27	81.51	*83.79*	83.79	*82.08*	*73.64*	81.18
Chi	82.18	81.32	*83.69*	*83.31*	*82.27*	*73.45*	*81.04*
GR	*81.61*	81.51	*83.69*	*82.18*	83.31	*74.02*	*81.05*
Rel	82.18	81.42	84.26	*83.22*	83.12	75.16	81.56
Proposed	**81.99**	80	**83.79**	**83.5**	**82.84**	**74.5**	**81.10**

Dataset ID :4, Name: Cardiotocography

	Jrip	**Ridor**	**J48**	**SC**	**NB**	**IBK**	**Avg**
IG	98.82	*98.4*	*98.58*	98.63	*88.33*	97.83	*96.77*
Chi	98.91	*98.11*	98.82	*98.54*	*89.46*	97.78	96.94
GR	*98.44*	*98.11*	98.82	*98.49*	90.21	97.69	96.96
Rel	*98.73*	*98.44*	*98.63*	98.63	90.54	*96.94*	96.99
Proposed	**98.73**	**98.49**	**98.63**	**98.54**	**89.93**	**97.22**	**96.92**
Dataset ID :5, Name: Lung Cancer							
	Jrip	Ridor	J48	SC	NB	IBK	Avg
IG	59.37	*53.12*	*59.37*	62.5	*65.62*	56.25	*59.37*
Chi	*53.12*	59.37	62.5	62.5	*62.5*	62.5	*60.42*
GR	*53.12*	59.37	62.5	62.5	*62.5*	62.5	*60.42*
Rel	*53.12*	68.75	*56.25*	56.25	*71.87*	68.75	62.50
Proposed	**56.25**	**56.25**	**59.37**	**68.75**	**71.87**	50	**60.42**
Dataset ID :6, Name: Libras Movement							
	Jrip	Ridor	J48	SC	NB	IBK	Avg
IG	*44.16*	*46.38*	*56.66*	*55.55*	*44.16*	*73.05*	*53.33*
Chi	*45.55*	*48.88*	*55.83*	*54.16*	*44.16*	*72.22*	*53.47*
GR	*43.61*	*47.77*	*57.5*	*52.5*	*41.38*	*72.22*	*52.50*
Rel	*48.61*	*49.72*	*60*	*58.88*	*44.16*	*76.38*	*56.29*
Proposed	**51.94**	**57.22**	**65.83**	**60.55**	**60.83**	**84.44**	**63.47**
Dataset ID :7, Name: Connectionist Bench(Sonar)							
	Jrip	**Ridor**	**J48**	**SC**	**NB**	**IBK**	**Avg**
IG	*77.4*	*72.11*	74.51	72.59	70.19	*87.98*	*75.80*
Chi	*77.4*	*72.11*	74.51	72.59	70.19	*87.98*	*75.80*
GR	*77.4*	*72.11*	74.51	72.59	70.19	*87.98*	*75.80*
REL	75.96	*76.44*	74.03	73.55	69.23	*87.5*	*76.12*
Proposed	**81.25**	**74.03**	**76.92**	**78.36**	**72.59**	85.09	**78.04**
Dataset ID :8, Name: Spambase							
	Jrip	**Ridor**	**J48**	**SC**	**NB**	**IBK**	**Avg**
IG	91.98	91.15	92.91	91.87	88.24	89.48	90.94
Chi	91.54	91	93.02	91.76	86.06	89.89	90.55
GR	*90.06*	89	*90.61*	*90.28*	*70.68*	*88.39*	*86.50*
Rel	*86.3*	*85.41*	*87.37*	*87.58*	*68.31*	*85.98*	*83.49*
Proposed	**90.52**	**90.15**	**91.48**	**91.45**	**75.94**	**88.93**	**88.08**
Dataset ID :9, Name: Breast Cancer(WDBC)							
	Jrip	**Ridor**	**J48**	**SC**	**NB**	**IBK**	**Avg**
IG	*91.91*	*92.26*	*92.79*	*92.26*	*92.44*	*92.97*	*92.44*
Chi	*92*	*92.26*	*92.79*	*92.26*	*92.44*	*92.97*	*92.45*
GR	*92*	*92.26*	*92.79*	*92.26*	*92.44*	*92.97*	*92.45*
Rel	93.84	*94.55*	*93.67*	*92.97*	94.55	96.13	94.29
Proposed	**93.49**	**94.9**	**94.37**	**93.84**	**93.32**	**95.07**	**94.17**
Dataset ID :10, Name: Musk (V 2)							
	Jrip	Ridor	J48	SC	Ibk	NB	Avg
IG	78.99	*72.89*	83.61	79.41	*84.87*	75.84	79.27
Chi	*74.36*	73.31	82.35	80.25	85.71	*76.05*	*78.67*
GR	*74.78*	75.42	*80.67*	77.94	*80.61*	*67.43*	*76.14*
Rel	76.26	74.78	82.35	81.09	*82.77*	*72.05*	*78.22*
Proposed	**75.84**	**73.52**	**81.09**	**81.72**	**85.5**	**74.78**	**78.74**

With Ionosphere data set, proposed method outruns than existing IG and GR using Ridor. Using J48, Simple cart and NB our method has recorded the best performance than IG, Chi, GR, but displayed less accuracy than existing ReliefF method. Using Instance based learner(IBK), proposed method performed better than IG, and Chi. Overall Average performance of the method has exhibited good accuracy than IG, Chi, GR, but not than Relief. With Dermatology and Libras Movement data sets ,proposed method has recorded the best performance than all existing methods using the all classifiers. With Connectionist Bench(Sonar) data set, this method performed well than all existing methods using all the classifiers except IBK. Our method has given the best performance than almost all existing methods with breast cancer data set. In the similar fashion, performance of proposed method can be interpreted on remaining data sets. The average competence of our method with the existing methods is given in below Table 8 With Win(W), Draw (D), and Loss(L)

From the above statistics, proposed method performed better than existing IG , Chi on 70 % of the data sets . ReliefF, performed better on 50 % of the data sets than proposed method. GR, performed better on 80 % of the data sets than proposed method.

Table 8. Average competence of proposed method with existing methods

	Data set ids			Out of 10 Data sets		
Technique	**Win**	**Draw**	**Loss**	**Win%**	**Draw %**	**Loss %**
IG	1,2,4,5,6,7,9	nil	3,8,10	70	nil	30
Chi	1,2,3,6,7,9,10	5	4,8	70	10	20
Gr	1,2,3,6,7,8,9,10	5	4	80	10	10
Rel	2,6,7,8,10	nil	1,3,4,5,9	50	nil	50

6. Conclusion

In this article, we have proposed a feature selection framework to reduce the data set dimensionality by selecting the best features in it to boosts-up the classification performance. For this research, two statistical approaches namely correlation coefficient and Symmetrical Uncertainty are considered to select the best features. Our proposed method was compared with four existing filter based methods namely, information gain(IG), Chi- Square (Chi), Grain Ratio (GR), and ReliefF. For testing the our proposed method, six different classifiers Jrip, Ridor, J48, Simple cart, Naive Bayes, IBk are applied on Ten different real time data sets. After thorough comparison analysis, our method displayed better results than existing IG and GR on 7 data sets, also performed better on 8 data sets. It is also competing with ReliefF method. The same technique can be implemented using Hadoop framework ,which is our future work.

References

1. Liao, S.H., Chu, P.H. and Hsiao, P.Y., 2012. Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications*, *39*(12), pp.11303-11311.
2. Kamal, N.A.M., Bakar, A.A. and Zainudin, S., 2015, August. Filter-wrapper approach to feature selection of GPCR protein. In *Electrical Engineering and Informatics (ICEEI), 2015 International Conference on* (pp. 693-698). IEEE.
3. Chatcharaporn, K.O.M.K.I.D., Kittidachanupap, N.A.R.O.D.O.M., Kerdprasop, K.I.T.T.I.S.A.K. and KERDPRASOP, N., 2012. Comparison of feature selection and classification algorithms for restaurant dataset classification. In *Proceedings of the 11th Conference on Latest Advances in Systems Science & Computational Intelligence* (pp. 129-134).
4. Chandrashekar, G. and Sahin, F., 2014. A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), pp.16-28.
5. Lin, K.C., Zhang, K.Y., Huang, Y.H., Hung, J.C. and Yen, N., 2016. Feature selection based on an improved cat swarm optimization algorithm for big data classification. *The Journal of Supercomputing*, *72*(8), pp.3210-3221.
6. Panthong, R. and Srivihok, A., 2015. Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. *Procedia Computer Science*, *72*, pp.162-169.
7. Maldonado, S. and Weber, R., 2011, November. Embedded Feature Selection for Support Vector Machines: State-of-the-Art and Future Challenges. In *CIARP* (pp. 304-311).
8. Fonti, V. and Belitser, E., 2017. Feature Selection using LASSO. https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf
9. Lebedev, A.V., Westman, E., Van Westen, G.J.P., Kramberger, M.G., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M. and Vellas, B., 2014. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, *6*, pp.115-125.
10. Hsu, H.H., Hsieh, C.W. and Lu, M.D., 2008, November. A Hybrid Feature Selection Mechanism. In *Intelligent Systems Design and Applications, 2008. ISDA'08. Eighth International Conference on* (Vol. 2, pp. 271-276). IEEE.
11. Potharaju, S.P. and Sreedevi, M., 2017.A Novel Cluster of Feature Selection Method Based on Information Gain.IJCTA, 10(14),pp.9-16
12. Peng, H., Long, F. and Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, *27*(8), pp.1226-1238.
13. Ding, C. and Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, *3*(02), pp.185-205.
14. Yuan, M., Yang, Z., Huang, G. and Ji, G., 2017. Feature selection by maximizing correlation information for integrated high-dimensional protein data. *Pattern Recognition Letters*, *92*, pp.17-24.
15. Partila, P., Voznak, M. and Tovarek, J., 2015. Pattern recognition methods and features selection for speech emotion recognition system. *The Scientific World Journal*, *2015*.
16. Koprinska, I., Rana, M. and Agelidis, V.G., 2015. Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, *82*, pp.29-40.
17. Song, Q., Ni, J. and Wang, G., 2013. A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE transactions on knowledge and data engineering*, *25*(1), pp.1-14.