Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No.04 (2023)

Feature Selection using Functional Dependency (FSFD)

[1*]**Dr. Vimalkumar B.Vaghela,** [2]**Kalpesh Patel,** [3]**Tushar A. Champaneria**

[1,2]*Assistant Professor, Department of Computer Engineering, L. D. College of Engineering, Ahmedabad, Gujarat, India*

[3]*Assistant Professor, Department of Computer Engineering, Government Engineering College, Modasa, Gujarat, India*

Abstract:- While performing data mining we come across large number of attributes. The more the number of attributes more time is required to process it .Also it might lead to incorrect and unexpected outcomes ,as the number of attributes increases ,they might become irrelevant to each other. Before performing any data mining task, preprocessing needs to be done to remove the unwanted attributes. Feature selection is one of the dimensionality reduction techniques which can be used to complete the data mining task efficiently. This paper throws light upon the importance of Feature selection technique in selecting the attributes of interest for the successful completion of data mining task and puts forth a new algorithm for Feature selection FSFD .i.e. Feature Selection using Functional dependency.

Keywords*: Feature selection, Data Mining, Filter, Wrapper.*

1. Introduction

Data mining is analyzing the data to solve the real time problems. It can also be used to enhance the business decisions. Ex:-Predicting the risk for granting the loan, guessing the next year's sale, finding the frequent item sets bought by the customers. Mining can be performed in different ways using Classification, Clustering, Association, Prediction, etc. depending on the application**.**

1.1 Classification

Classification is predicting the categorical label and is a supervised machine learning procedure. It mainly comprises of two stages. First, a classifier is built using the training data set .The tuples in the training set contains the class labels. Second, the trained classifier is applied on testing data set, indirection to forecast the label for the unknown data.There are numerouskindsof classifiers Neural Network, Decision Tree, Naive Bayes Classifier, If-Then Rules.If largeamount of attributes are present in the training data set it affects the processing speed of classifier and might lead to incorrect results. In direction to improve the precision of classifier Feature selection needs to be performed.

1.2 Feature selection

Feature selection is a Dimensionality reduction technique. It selects the useful features and eliminates the redundant oneseg:-age, birth date.Thus,it provides the optimal set of features. The optimal set of features consists of relevant and non redundant features. Large data sets contain many features but only someof they are useful for performing the data mining task. There are two Frameworks for Feature selection one is Classical Framework and the new one. The Classical Framework of Feature selection is of four steps Generation of subset, Evaluation, stopping criteria and validation. The New Framework is of two steps that are; it can be done using Relevance analysis and Redundancy analysis. Therefore, Relevance analysis is done to find the relevant feature set. The features are classified in to four categories as strongly relevant, weakly relevant non-redundant, weakly relevant redundant and irrelevant features based on relevancy.

Optimal subset=strongly relevant features+ weakly relevant (non-redundant) features

As mentioned above the optimal set contains weakly relevant non-redundant features, Redundancy Analysis is needed to identify which subset of weakly relevant features is non-redundant. Therefore redundancy analysis is an important step in feature selection. Correlation between attributes can be used for redundancy analysis. Example:-Pearson coefficient, Symmetric Uncertainty. Feature Selection can be done in two ways using Filter model, Wrapper model.

Filter model uses the characteristics of training data to perform attributes selection.eg: Entropy, distance. It does not use any feedback from learning algorithm. It is computationally cheaper than Wrapper model. It is suitable for large data sets. Filter model can be implemented in two ways Feature weighting approach, Subset search method. Feature weighting approach assigns weights or ranks to individual features using measures such as Information gain, Distance, Consistency, Classifier error rate, Dependency.

In subset search method optimal subset of features is found out using different search strategies. Example:-exhaustive, heuristic, random search.Exhaustive search means checking all the subsets. Heuristic search can be Forward selection, backward selection, etc. Time complexity of subset selection approach is high. Therefore, it is less suitable for High data sets.

Feature weighting approach removes only irrelevant features as this approach selects all the features above particular threshold and the redundant features likely have same ranking therefore they gets selected.Whereas, subset search method removes irrelevant and redundant features.Drawback of Filter model is it does not consider the effect of the selected features on the performance of induction algorithm leading to less accurate results than Wrapper.

In Wrapper model performance of learning algorithm is used to select the attributes. It is useful for selecting the best subset. It selects the subset using subset search method. Consider, if it uses forward search strategy for selecting the subset then it verifies the result using the classifier. In the above case if the accuracy increases by selecting a particular attribute then it keeps the attribute otherwise eliminates it. It gives more accurate results compared to Filter model.Drawback of Wrapper model is,it is computationally expensive and therefore, is less suitable for high dimensional data.

The combination of above two models can be implemented as Hybrid model.

Feature selection not only helps in reducing the size of the data, processing time of the data ,removing noise from it but also increasing accuracy, simplicity and understanding of the data.

In our paper we will be using Filter model. Also, we have proposed a new algorithm for Feature selection where the redundant attributes is removed using the concept of Functional dependency.

2. Related work

M Dash and H Liu in 1997(Feature Selection for classification) explains the Feature selection process, compared various algorithms and gave guidelines for goodfeature selection methods. JasminaNovaković, PericaStrbac, DusanBulatović in Toward Optimal Feature Selection UsingRanking Methods And Classification Algorithm compared different feature ranking methods using various classifiers and concluded that the Ranking method to be used depends on the classifier to be used.JasminaNovakovic inThe Impact of Feature Selectionon the Accuracy of Naïve Bayes Classifieranalysed the impact of various ranking measures for Naïve Bayesian Classifier.Vijay Kumar Verma and Pradeep Sharma in Data Dependencies Mining In Database by Removing Equivalent Attributes proposed a new algorithm DM_EC i.e.(dependency mining using Equivalent Candidates) for removing redundant attributes and data from a database. It uses Functional dependency for removing redundant attributes. Lei Yu &Huan Liu in Efficient Feature Selection via Analysis ofRelevance and Redundancyhighlighted the importance of Redundancy analysis in Feature Selection and the relevant concepts like eg:Correlation measures (Pearson coefficient, Entropy based), Markov’s blanket, Predominant features.Also proposed FCBF algorithm which uses symmetric uncertainty measure for correlation based feature selection. BarisSenliol, GokhanGulgezen, Lei Yu and ZehraCataltepeput forth FCBF# which uses different searches strategy than FCBF to find the optimal set of attributes. Compared the results with mRMR algorithm which subset search strategy for Feature selection and found that FCBF# gives equivalent results with the mRMR algorithm for smaller data sets.Ponsa and Antonio L´opez in Feature Selection Based on a New Formulation of the Minimal-Redundancy-Maximal-Relevance Criterion has suggested a modification to the minimal Redundancy maximal relevance algorithm so as to improve its performance.John, George H., Ron Kohavi, and Karl Pfleger in Irrelevant Features and the Subset Selection Problem modified the subset search methods i.e.Forward selection and Backward elimination to perform both addition and deletion for implementing Wrapper model. Selection of Relevant Features for Multi-Relational Naive Bayesian Classifier implements Hybrid model where Filter is used to select minimum redundant features and Wrapper is used to select maximium relevant features. They have used info distance for relevance analysis and Pearson coefficient for redundancy analysis.

3. Feature Selection Using Functional Dependency

Filter model is suitable for high dimensional data. There exist many algorithms for implementing the filter model, Relief uses Euclidean distance and calculates nearest hit and nearest miss, which is then used to find out the relevant features but it does not take care of redundancy. The Hybrid model implemented in previous paper had used Pearson coefficient for redundancy analysis but it cannot handle nominal values.It is also known linear coefficient.The Linear correlation cannot always be used as the features might not be linearly correlated every time. Therefore, a measure which can handle all types of data is needed.In this paper we have proposed a new algorithm FSFD which implements the Filter model for Feature Selection in two steps .i.e. Relevance and Redundancy analysis using the concepts of Information Gain, Symmetric Uncertainty and Functional dependency.

3.1 Information Gain

To find the relevant attributes Information gain can be used. Information Gain can be defined, usingconcept of entropy. Consider a variable X, which takes N values $\{S_i\}_{i=1 \text{ to } N}$

$P(S_k)$ be the probability when $X=S_k$.Then the information obtained when $X=S_k$is :

$$\mathrm{I(X)} = \log\left(\frac{1}{\mathrm{P(X)}}\right) = -\log(\mathrm{P(X)}) \qquad (1)$$

The entropy is a measure of unpredictability. It is the expectation of information.The entropy of X can be calculated as below:-

$$\mathrm{E(X)} = -\sum_{i=1}^{N} \mathrm{P}(x_i) \log_2 \mathrm{P}(x_i) = \sum_{i=1}^{N} \mathrm{P}(x_i)\, \mathrm{I}(x_i) \qquad (2)$$

The Entropy of variable X after observing the value of Y can be calculated as below:

$$\mathrm{E(X|Y)} = -\sum \mathrm{P}(y_i) \sum \mathrm{P}(x_i|y_i) \log_2(\mathrm{P}(x_i|y_i)) \qquad (3)$$

$$\text{Where}\mathrm{P}\left(x_i|y_j\right) = P(x_i, y_j)/P(y_j) \qquad (4)$$

If the observed values of X in the training data set S are partitioned according to the values of a second feature Y, and the entropy of X with respect to the partitions induced by Y is less than the entropy of X prior to partitioning, then there is a relationship between features X and Y. Given the entropy is a criterion of impurity in a training set S, we can define a measure reflecting additional information about X provided by Y that represents theamount by which the entropy of X decreases .This measure is known as Information Gain or Mutual Information.Below is the equation for calculating the Information gain.

$$\mathrm{IG(X, Y)} = \mathrm{E(X)} - \mathrm{E(X|Y)} \qquad (5)$$

But Information gain is biased towards attributes with more number of distinct values. So, it can be normalized using Symmetric uncertainty.

3.2 Symmetric Uncertainty

$$\mathrm{SU}(X, Y) = 2\left[\frac{\mathrm{IG}(X,Y)}{\mathrm{E}(X)+\mathrm{E}(Y)}\right] \qquad (6)$$

For Normalizing the value we are taking $SU(F_i, C)$. It normalizes the values in the range [0, 1]. A value of SU = 1 means one feature completely predictsthe other, and SU = 0 indicates, that X and Y are independent.

3.3 Threshold

Using above SU value, Mean, variance, standard deviation is calculated.Threshold value is calculated as:

Threshold= mean + (0.6*standard_deviation) (7)

All the SU values above this Threshold are selected as Relevant attributes.

3.4 Funtional dependency

The concept of Functional dependency can be used to remove redundant attributes. i. e if a->b means whenever a repeats if b repeats and b→a then they are redundant attributes.

To Find, whether a→b, we calculate the distinct values of "a". For each distinct value of "a" number of distinct values of "b" are calculated.If this number is greater than one means whenever "a" repeats "b" has more than one value and therefore the condition a->b fails. i.e a is not redundant to b. On the otherhand, if for each distinct value of "a","b" has only 1 distinct value indicates that "a" is redundant to "b". In this case same procedure is used to verify whether b→a is also true or not. If this also gets satisfied then only we can say that "a" and "b" are functionally dependent otherwise not.

Functional Dependency supports all types of data and is independent of class therefore it can handle any number of class values. Thus, itsatisfies the requirement of the good filter as specified in[1].

Below is the proposed algorithm:-

input: *S*(*F*1,*F*2, ...,*FN*,*C*) // a training data set

□□// a predefined threshold

output: *Sbest*// a selected subset

begin

```
for i= 1 to N do begin
calculate SUi,cfor Fi;
if (SUi,c>□)
append Fi to S0 list ;
end;
order S0 list in descending SUi,c value;
for k =1 to selected do begin
for  m = k+1 to selected do begin
S1 list= Get the distinct values for feature Fk;
Count = 0;
while S1 list!= null
Nk=Getnextelement (S1);
count distinct values of FmwhereFk=Nk;
if (count>1)
flag = 0;//feature is not redundant
break;
else
Flag = 1;
if(flag == 1)
```

S_2 list = Get the distinct values for feature *Fm*;
Count1= 0;
while S_2 list*!= null*
Nm = Getnextelement(S_2);
count distinct values of FkwhereFk=Nm;
if (*count1 >1*)
remove_ind = 0;//feature is not redundant
break;
else
remove_ind = 1;//remove F_m from S_0*list*
end for;
end for;
Sbest = S_0*list* ;
end;

Figure 1: Algorithm for feature selection using functional dependency (FSFD)

4. Experimental Work and Result

We performed the testing on 5 data sets from UCI repository. The data sets are Promoters, Splice, Chess, Ionosphere, Sonar. We purposefully introduced redundant fields in them and those were identified by the algorithm. Also we have compared the results of Symmetric uncertainty with Functional dependency for those 5 data sets.

Tabel 1: Feature selection using FD

Datasets	***Full Set of Features***	***Feature Selection using FD***
Promoters	58	08
Splice	60	07
Chess	36	03
Ionosphere	32	07
Sonar	60	12
Average	**49.2**	**7.4**

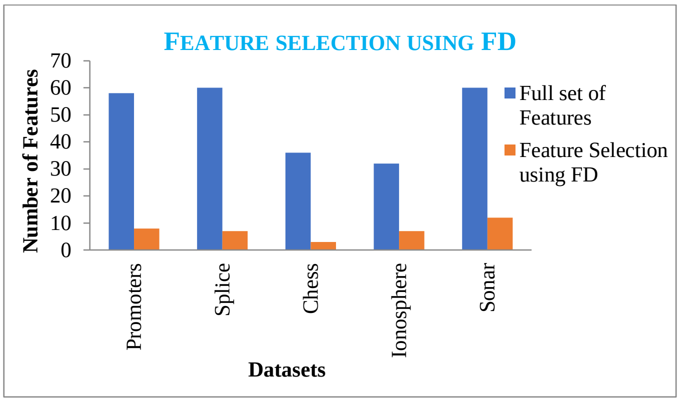

Figure 2: Feature Selection using functional dependency (FSFD)

Table 1 shows how our proposed method reduces the number of features from the original datasets. In promoters dataset contains the 58 features and when we apply the feature selection using the functional dependency it will reduce to 08 features. It is also important to measure the accuracy of the dataset with different classifiers. In table 2 we have shown the accuracy comparisons.

Table 2: Accuracy comparision using FSFD as a decision tree as a classifier

Datasets	***Decision Tree as a Classifier (Accuracy %)***	
	Full Set of Features	***Feature Selection using FD***
Promoters	73.47	74.52
Splice	91.35	94.02
Chess	91.43	91.68
Ionosphere	84.61	87.18
Sonar	73.43	75.36
Average	**82.86**	**84.57**

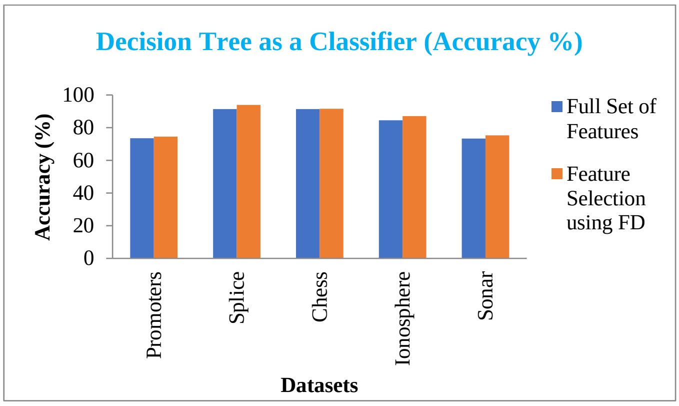

Figure 3: Accuracy Comparison using FSFD as a Decision Tree as a Classifier

Table 2 shows that accuracy improves after applying the feature selection using the propose algorithm based on the functional dependency.

TABLE 3: ACCURACY COMPARISION USING FSFD AS A NAÏVE BAYESIAN AS A CLASSIFIER

Datasets	**Naïve Bayesian as a Classifier (Accuracy %)**	
	Full Set of Features	***Feature Selection using FD***
Promoters	85.93	95.28
Splice	95.36	94.02
Chess	87.89	90.43
Ionosphere	86.35	89.46
Sonar	76.82	79.72

Average	**86.47**	**89.79**

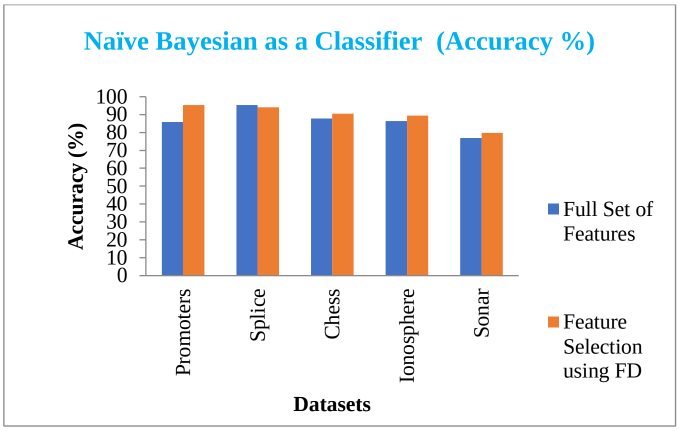

Figure 4: Accuracy Comparison using FSFD as a Naïve Bayesian as a Classifier

Table 3 shows the accuracy comparison using the Naïve Bayesian as a classifier. Only in the splice dataset there is reduction in the accuracy on minor basis but the overall average of the accuracy improvement is good on remaining four datasets.

TABLE 4: ACCURACY COMPARISION BETWEEN FCFB ALGORITHM AND OUR PROPOSED ALGORITHM FSFD (DECISITION TREE AS A CLASSIFIER)

Datasets	***Decision Tree as a Classifier (Accuracy %)***	
	FCFB	***FSFD***
Promoters	74.53	75.32
Splice	94.01	95.10
Chess	90.42	90.42
Ionosphere	86.04	87.18
Sonar	72.95	75.36
Average	**83.59**	**84.68**

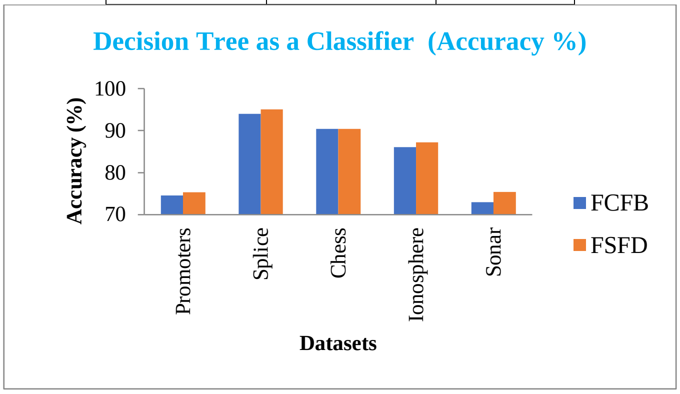

Figure 5: Accuracy comparision between FCFB algorithm and our proposed algorithm FSFD (Decisition Tree as a classifier)

Table 4 shows the accuracy comparison between the well-known algorithms FCFB with our proposed algorithm FSFD. We have performed the comparison using the Decision Tree as a classifier. Only in the chess dataset the algorithm perform the same as the FCFB and there is accuracy improvement on remaining four datasets and also the average accuracy improves using the FSFD approach.

TABLE 5: ACCURACY COMPARISION BETWEEN FCFB ALGORITHM AND OUR PROPOSED ALGORITHM FSFD (NAÏVE BAYESIAN AS A CLASSIFIER)

Datasets	***Naïve Bayesian as a Classifier (Accuracy %)***	
	FCFB	***FSFD***
Promoters	92.43	95.28
Splice	94.17	94.87
Chess	89.24	90.43
Ionosphere	84.90	89.46
Sonar	77.78	79.71
Average	**87.70**	**89.95**

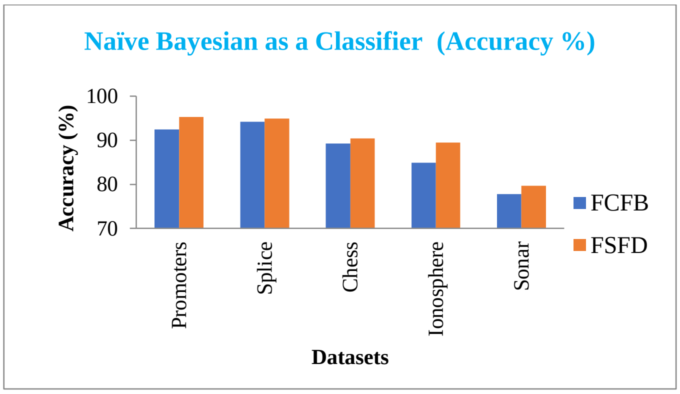

Figure 6: Accuracy comparision between FCFB algorithm and our proposed algorithm FSFD (Naïve Bayesian as a classifier)

Table 5 shows the accuracy comparison between the well-known algorithms FCFB with our proposed algorithm FSFD. We have performed the comparison using the Naïve Bayesian as a classifier. We have achieved the better performance in all the dataset and also the overall average accuracy improves.

5. Conclusion

After comparing the results of Functional dependency with Symmetric Uncertainty we found that Functional dependency gives results equivalent or better than Symmetric Uncertainty.

Refrences

[1] Dash, Manoranjan, and Huan Liu. "Feature selection for classification." Intelligent data analysis 1.3 (1997): 131-156.

[2] Novaković, Jasmina, Perica ŠTRBAC, and Dušan Bulatović. "Toward optimal feature selection using ranking methods and classification algorithms." *Yugoslav Journal of Operations Research ISSN: 0354-0243 EISSN: 2334-6043* 21.1 (2011).

[3] Novakovic, Jasmina. "The Impact of Feature Selection on the Accuracy of Naïve Bayes Classifier." *18th Telecommunications forum TELFOR*. 2010.

[4] John, George H., Ron Kohavi, and Karl Pfleger. "Irrelevant Features and the Subset Selection Problem." *ICML*. Vol. 94. 1994.

[5] Yu, Lei, and Huan Liu. "Efficient feature selection via analysis of relevance and redundancy." *The Journal of Machine Learning Research* 5 (2004): 1205-1224.

[6] Vaghela, V. B., Amit Ganatra, and Amit Thakkar. "Boost a weak learner to a strong learner using ensemble system approach." *Advance Computing Conference, 2009. IACC 2009. IEEE International*. IEEE, 2009.

[7] Senliol, Baris, et al. "Fast Correlation Based Filter (FCBF) with a different search strategy." *International Symposium on Computer and Information Sciences (ISCIS 2008)*. 2008.

[8] Vimalkumar B. Vaghela, Dr. Kalpesh H. Vandra, Dr. Nilesh K. Modi, *"Multi-Relational Classification Using Inductive Logic Programming"*, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 3, May - 2012 ISSN: 2278-0181.

[9] A. Ouardighi, A. Akadi, and D. Aboutajdine (2007), "Feature Selection on Supervised Classification Using Wilk's Lambda Statistic," *Proceeding of 2007Computational Intelligence and Intelligent Informatics (ISCIII'07)*, pp. 51~55, March 2007.

[10] Vimalkumar B. Vaghela, Kalpesh H. Vandra, Nilesh K. Modi, *"Analysis and Comparative Study of Classifiers for Relational Data Mining"*, International Journal of Computer Applications (0975 – 8887) Volume 55– No.7, October 2012.

[11] B. Hu, H. Liu, J. He, and X. Du (2008), "FARS: A Multi-relational Feature and Relation Selection Approach for Efficient Classification," *Advanced Data Mining and Applications: 4th International Conference, Lecture Notes in Computer Science*, vol. 5139, pp. 73~86, 2008.

[12] Vaghela, Vimalkumar Bhupatbhai. "MR2 Based Feature Selection for Multi-Relational Naïve Bayesian Classifier." *International Journal of Applied Research in Computer Science and Information Technology* 2.1 (2013).

[13] Gaertner, T., Flach, P., Kowalczyk (2002),. "Multi-instance kernels", *In Proceedings of 19th International Conf. on Machine Learning, pp.179-186*, 2002.

[14] Vaghela, Vimalkumar B., Kalpesh H. Vandra, and Nilesh K. Modi. "MR-MNBC: MaxRel based feature selection for the multi-relational Naïve Bayesian Classifier." *Engineering (NUiCONE), 2013 Nirma University International Conference on*. IEEE, 2013.

[15] M. Hall (2000), "Feature Selection for Discrete and Numeric Class Machine Learning," *Proceeding of Seventeenth International conference on Machine Learning*, San Francisco, California, pp. 359~366, 2000.

[16] Vaghela, V. B., K. H. Vandra, and N. K. Modi. "Information Theory Based Feature Selection for Multi-Relational Naïve Bayesian Classifier." *J Data Mining Genomics Proteomics* 5.155 (2014): 2153-0602.

[17] S. Muggleton and C. Feng (1990), "Efficient induction of logic programs," *Proceedings of the 1st Conference on Algorithmic Learning Theory*, Tokyo, Japan, pp. 368~381, 1990.

[18] Xu GM, Yang BR, Qin YQ (2008), "New multi relational naïve Bayesian classifier", *Systems Engineering and Electronics, vol. 30, No.4, pp 655-655*, 2008.

[19] Y. Tadeuchi, R. Oshima, K. Nishida, K. Yamauchi, and T. Omari (2007), "Quick Online feature selection method for regression - A feature selection method inspired by human behavior," *Proceeding of the IEEE International Conference on Systems, Man and Cybernetics*, Canada, October 2007.

[20] Vaghela, Vimalkumar B., Kalpesh H. Vandra, and Nilesh K. Modi. "Entropy Theory Based Feature Selection for Multi-Relational Naïve Bayesian Classifier." *Journal of International Technology and Information Management (JITIM) (1543-5962) 13-26 Volume 23, Number 1, 2014.*

First A. Author Vimalkumar B. Vaghela, is an Assistant Professor in Department of Computer Engineering at L. D. College of Engineering, Ahmedabad, Gujarat, India. He had done his Ph.D. in Relational Data Mining. He has published more than 35 research papers in international journals and conferences. He had published a book titled "Ensemble Classifier in Data Mining". He has more than 18 years of teaching experience; his biography was Biography has been published in the Who's Who in Science and Engineering – 11thEdition, 2010-11, 2013-14, 2016-17. Written a Book with title "Operating System" in Dreamtech Publisher. Senior Professional Members of Association for Computing Machinery (ACM), Life time Professional Member of Computer Society of India (CSI) & Professional Member of International Association of Engineers (IAENG).

Second Author Kalpesh M Patel, is an Assistant Professor in Department of Computer Engineering at L D College of Engineering, Ahmedabad. He has more than 15 years of teaching experience at engineering colleges.

Third Author Tushar A. Champaneria, is an Assistant Professor in Department of Computer Engineering at Government Engineering College, Modasa. He is pursing Ph.D. in area of IoT and He has published more than 30 research papers in international journals and conferences. He had published two patents. He has 3 years of industry experience and more than 13 years of teaching experience.