# An Effective Performance of Feature Selection with Classification of Data Mining Using SVM Algorithm

A.Veeraswamy and S. Appavu Alias Balamurugan

***Abstract---*** *This paper proposes one method of feature selection by using Support Vector Machine. The purpose of the proposed method is to reduce the computational complexity and increase the classification accuracy of the selected feature subsets. The dependence between two attributes (binary) is determined based on the probabilities of their joint values that contribute to Zero Probability and One Probability classification decisions. Then the two attributes are considered independent of each other, otherwise dependent, and one of them can be removed and thus the number of attributes is reduced. The process must be repeated on all combinations of attributes. The paper also evaluates the approach by comparing it with existing feature selection algorithms over 6 datasets from University of California, Irvine (UCI) machine learning databases. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most existing algorithms.*

***Keywords---*** *Feature Selection, Classification, Data Mining J48, K-Star, SMO*

## I. INTRODUCTION

DATA mining is a form of knowledge discovery essential for solving problems in a specific domain. Classification is a technique used for discovering classes of unknown data. As the world grows in complexity, overwhelming us with the data it generates, data mining becomes the only hope for elucidating the patterns that underlie it [1]. The manual process of data analysis becomes tedious as size of data grows and the number of dimensions increases, so the process of data analysis needs to be computerized. The term Knowledge Discovery from data(KDD) refers to the automated process of knowledge discovery from databases. The process of KDD is comprised of many steps namely data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation [1].A "feature" or "attribute" or "variable" refers to an aspect of the data. Usually before collecting data, features are specified or chosen. Features can be discrete, continuous, or nominal. Generally, features are characterized as: Relevant: These are features which have an influence on the output and their role cannot be assumed by the rest [1].Irrelevant: Irrelevant features are defined as those features not having any influence on the output, and whose values are generated at random for each example. Redundant: A redundancy exists whenever a feature can take the role of another (perhaps the simplest way to model redundancy).Problem of selecting some subset of a learning algorithms input variables upon which it should focus attention, while ignoring the rest. Feature selection is the process of selecting the best feature among all the features because all the features are not useful in constructing the clusters: some features may be redundant or irrelevant thus not contributing to the learning process [1].

Feature selection, a process of choosing a subset of features from the original ones, is frequently used as a preprocessing technique in data mining. It has proven effective in reducing dimensionality, improving mining efficiency, increasing mining accuracy, and enhancing result comprehensibility. Feature selection methods can broadly fall into the wrapper model and the filter model [1].The quality of the data is one such factor. if information is irrelevant or redundant, or the data is noisy and unreliable, then knowledge discovery during training is more difficult. Feature subset Selection is the process of identifying and removing as much of the irrelevant and redundant information as possible. Machine learning algorithms differ in the amount of emphasis they place on feature selection [2].

## II. PROPOSED WORK

In this paper, we introduce a novel approach for feature selection in high dimensional data using Support Vector Machine. Support Vector Machines [19] are basically binary classification algorithms. Support Vector Machines (SVM) are

*A.Veeraswamy, Research Scholar, Veltech Dr.RR & Dr.SR Technical University, Chennai, TamilNadu, India.*
*S. Appavu Alias Balamurugan, Professor & Research Coordinator, KLN College of Information Technology, Madurai, TamilNadu, India.*

*ISBN 978-93-82338-68-0*

a classification system derived from statistical learning theory. It has been applied successfully in fields such as text categorization, hand-written character recognition, image classification, bio sequences analysis, etc. The SVM separates the classes with a decision surface that maximizes the margin between the classes. The surface is often called the optimal hyper plane, and the data points closest to the hyper plane are called support vectors. The support vectors are the critical elements of the training set. The mechanism that defines the mapping process is called the kernel function [5].

The SVM can be adapted to become a nonlinear classifier through the use of nonlinear kernels. SVM can function as a multiclass classifier by combining several binary SVM classifiers. The output of SVM classification is the decision values of each pixel for each class, which are used for probability estimates. The probability values represent "true" probability in the sense that each probability falls in the range of 0 to 1, and the sum of these values for each pixel equals 1.Classification is then performed by selecting the highest probability. SVM includes a penalty parameter that allows a certain degree of misclassification, which is particularly important for no separable training sets. The penalty parameter controls the trade-off between allowing training errors and forcing rigid margins. It creates a soft margin that permits some misclassifications, such as it allows some training points on the wrong side of the hyper plane. Increasing the value of the penalty parameter increases the cost of misclassifying points and forces the creation of amore accurate model that may not generalize well [4].

The paper also evaluates the approach by comparing it with existing feature selection algorithms over 6 datasets from University of California, Irvine (UCI) machine learning databases [6]. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most existing algorithms.

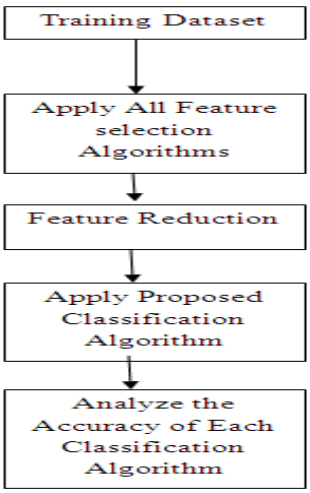


Fig 1: Data Flow Analysis Diagram for Feature Selection with Classification

## III. SYSTEM IMPLEMENTATION

The proposed algorithm is implemented using Java. The stepwise approach is as follows. The input to the system is given as an attribute-relation file format (ARFF) file. A table is created in Oracle using the name specified in @relation". The attributes specified under @attribute" and instances specified under @data" are retrieved from the ARFF ¯le and then they are added to the created table. This procedure is followed for providing the training set as well as test set. The created table acts as the dataset and is given as the input to the proposed algorithm.

The combination of attribute value should occur at least once in the dataset, because while finding the dependency between attribute values if a combination of attribute value did not occur once, then it will lead to alternate zeros resulting in zero probability and dependency that cannot be found. Thus, the above condition is checked before a combination of attribute value is given to the proposed method. The probabilities are calculated for the given input. Based on the probabilities, the dependent attributes are identified.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

The feature selection using SVM is applied to many datasets, and the performance evaluation is done. We presented the performance evaluation on 6 datasets. All together 8 datasets are selected from the UCI machine learning repository and the UCI knowledge discovery in databases (KDD) archive [6].

ISBN 978-93-82338-68-0

A summary of datasets is presented in Table 1. For each dataset, we run all seven feature-selection algorithms, SVM, consistency subset eval, Info Gain attribute eval, Gain Ratio attribute eval, OneR attribute eval, Chi Squared attribute eval, principal components, classifier subset eval, respectively, and record the number of selected features for each algorithm. We then apply SVM,Naïve Bayes, decision tree (J 48), K-Star ,Random Tree, OneR on the original dataset as well as each newly obtained dataset containing only the selected features from each algorithm and recorded the overall accuracy by 10 fold cross validation.

From Table 2, it is found that for all the datasets, some feature selection algorithms will select the attributes that are fewer than the number of attributes selected by the proposed method. However, when the selected attributes by the above specified existing feature selection methods are used for the classification. The main idea in the proposed method is finding the dependency between the attributes in deciding the class attributes value, and also the probabilities will decide the dependency between a set of attributes. Therefore, the proposed method removes the dependent attributes and identifies the perfect attributes which are sufficient for the classification of the datasets and also improve the classification accuracy.

Table 1: Details Description of Dataset used in the Experiment

| S.NO | Name of the Dataset | No. of Attributes | No. Of Instances | Selected Attributes on Proposed Algorithm |
|---|---|---|---|---|
| 1 | anneal | 39 | 898 | 5 |
| 2 | audiology | 70 | 226 | 24 |
| 3 | balance-scale | 5 | 625 | 3 |
| 4 | kr-vs-kp | 37 | 3196 | 2 |
| 5 | tic-tac-toe | 10 | 958 | 2 |
| 6 | vehicle | 19 | 846 | 4 |

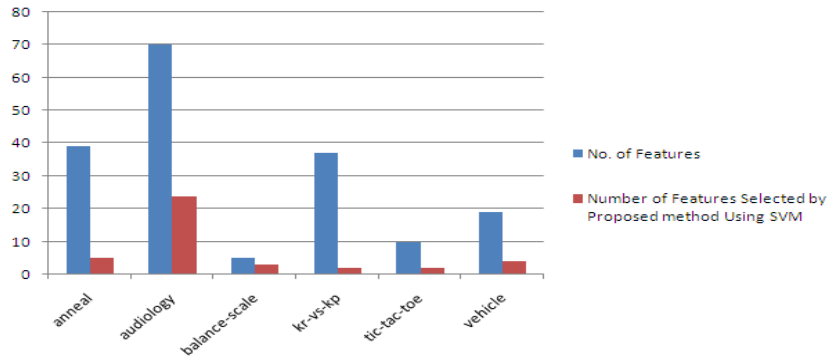


Fig 2: Number of Features Selected by the Proposed Method

Table 2: Number of Selected Features for each Feature Selection Algorithm

| S.NO | Name of the Dataset | Cfs | Chi | Gain Ratio | Info Gain | One Attribute | PCA | Constituent Subset Value | SMO |
|---|---|---|---|---|---|---|---|---|---|
| 1 | anneal | 7 | 38 | 38 | 38 | 38 | 37 | 8 | 5 |
| 2 | audiology | 6 | 69 | 69 | 69 | 69 | 63 | 13 | 24 |
| 3 | balance-scale | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 3 |
| 4 | kr-vs-kp | 3 | 36 | 36 | 36 | 36 | 31 | 6 | 2 |
| 5 | tic-tac-toe | 1 | 9 | 9 | 9 | 9 | 16 | 8 | 2 |
| 6 | vehicle | 11 | 18 | 18 | 18 | 18 | 7 | 18 | 4 |

*ISBN 978-93-82338-68-0*

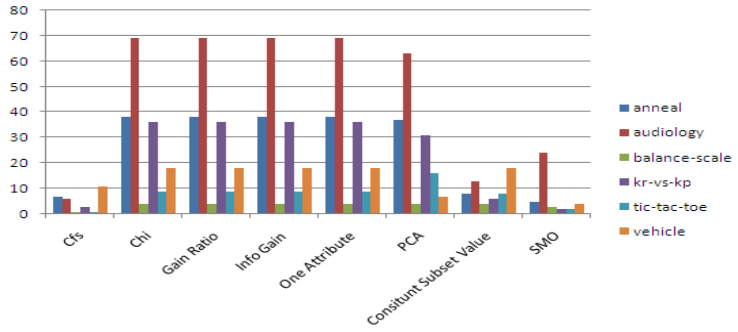


Fig 3: No. Of Features selected by each Feature Selection Algorithm

Table 3: Accuracy of each Classification Algorithm Applying for each Dataset

| S.NO | Name of the Dataset | K-Star | ONE-R | J48 | SMO | Naïve bayes | Random Tree |
|---|---|---|---|---|---|---|---|
| 1 | anneal | 95.66 | 83.63 | 97.22 | 97.33 | 85.97 | 96.43 |
| 2 | audiology | 70.80 | 46.46 | 69.47 | 81.86 | 68.14 | 59.29 |
| 3 | balance-scale | 63.52 | 56.32 | 63.52 | 87.52 | 63.52 | 77.76 |
| 4 | kr-vs-kp | 90.43 | 66.45 | 90.43 | 95.56 | 90.43 | 88.45 |
| 5 | tic-tac-toe | 69.94 | 69.93 | 69.94 | 98.32 | 69.94 | 73.48 |
| 6 | vehicle | 66.43 | 51.53 | 68.32 | 74.34 | 48.46 | 64.30 |
| Average Accuracy | | **76.13** | **62.39** | **76.48** | **89.15** | **71.08** | **76.62** |

Table 3 Specifies the Accuracy of all Classification Algorithms like K-Star, OneR, J48,SVM, Naïve Bayes and Random Tree. These Accuracy are Compared by SVM.

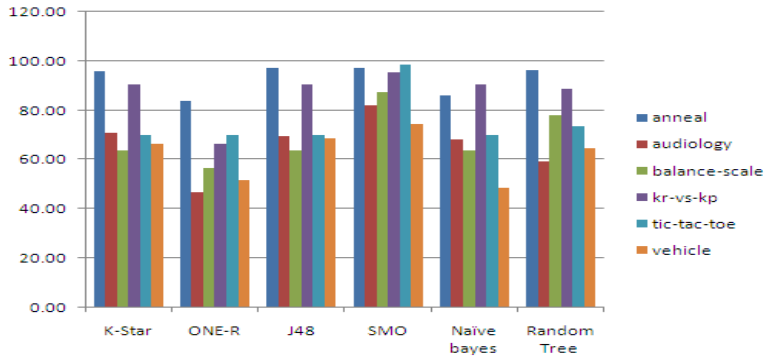


Fig 4: Accuracy of each Classification algorithm applying for each Dataset

*ISBN 978-93-82338-68-0*

Table 4: Applying Each Feature Selection Algorithm with K-Star and Compared With Proposed Algorithm

| S.NO | Name of the Dataset | Cfs | Chi | Gain Ratio | Info Gain | One Attribute | PCA | Constituent Subset Value | K-Star | SMO |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | anneal | 88.08 | 98.99 | 95.76 | 98.66 | 98.66 | 96.1 | 88.08 | 95.66 | 97.33 |
| 2 | audiology | 53.53 | 99.55 | 99.55 | 99.55 | 99.55 | 99.11 | 50.01 | 70.80 | 81.86 |
| 3 | balance-scale | 74.72 | 45.76 | 45.76 | 45.76 | 88.48 | 45.76 | 45.76 | 63.52 | 87.52 |
| 4 | kr-vs-kp | 70.91 | 72.62 | 72.62 | 72.62 | 72.62 | 92.45 | 70.52 | 90.43 | 95.56 |
| 5 | tic-tac-toe | 82.46 | 17.84 | 17.84 | 17.84 | 17.84 | 96.86 | 39.03 | 69.94 | 98.32 |
| 6 | vehicle | 66.78 | 25.65 | 25.65 | 25.65 | 25.65 | 68.55 | 25.65 | 66.43 | 74.34 |
| Average | | 72.75 | 60.07 | 59.53 | 60.01 | 67.13 | 83.14 | 53.18 | 76.13 | 89.15 |

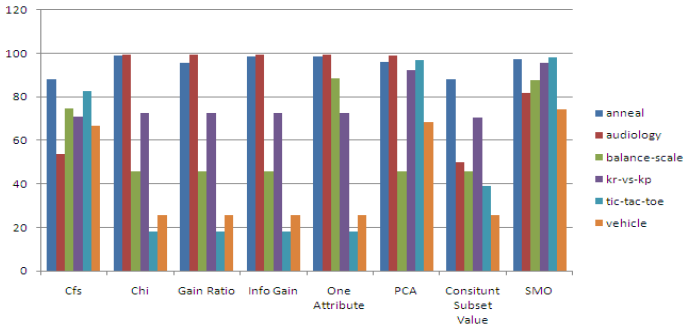


Fig 5: Performance Analysis Graph by using the Proposed Algorithm SVM

## V. CONCLUSION

This paper proposes a feature selection algorithm based on SVM. The algorithm can remove redundancy from the original dataset. The main idea provided is to find the dependent attributes and remove the redundant ones among them. The technology to obtain the dependency needed is based on SVM. A new attribute reduction algorithm of using SVM is implemented and evaluated through extensive experiments via comparison with related attribute reduction algorithms. In this paper, we consider the task of feature selection and investigate the performance of nine feature selection algorithms.

Our findings can be summarized as follows:

1) In feature selection approach, we have shown that SVM is a promising approach for automatic feature selection. It outperforms most existing algorithms in terms of number of selected features, classification accuracy. Well-established algorithms, such as InfoGain attribute eval, GainRatio attribute eval, OneR attribute eval, ChiSquared attribute eval, principal components are also more complex than SVM feature Selection,SVM based feature selection runs very efficiently on large datasets, which makes it very attractive for feature selection in high dimensional data.
2) We have implemented a new feature selector using SVM and found that it performs better thanthe popular and computationally expensive traditional algorithms.
3) We compared the performance of a number of algorithms on the UCI machine learning repository datasets.

## REFERENCES

[1] Classification and Feature Selection Techniques in Data Mining, Sunita Beniwal*, Jitender Arora International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 6, August – 2012 ISSN: 2278-0181.
[2] Feature Selection for Discrete and Numeric Class Machine Learning, Mark A. Hall, University of Waikato, Hamilton, New Zealand.
[3] Effective and Efficient Feature Selection for Large-scale Data Using Bayes Theorem, Subramanian Appavu Alias Balamurugan, Ramasamy Rajaram, International Journal of Automation and Computing, February 2009, 62-71.
[4] C.W. Hsu, C.C. Chang and C.J. Lin, "A practical guide to support vector classification", http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, 2003.
[5] V. N. Vapnik, *Statistical Learning Theory,* Wiley New York., 1998
[6] C. L. Blake, C. J. Merz. UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, USA, [Online],Available: http://www.ics.uci.edu/ ~ mlearn, 1998.