International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7, Issue-6, March 2019

# An Effective Data Classification Method for Medical Dataset in terms of Accuracy and Time

**Sivakumar Venkataraman, Subitha Sivakumar**

***Abstract:*** ***Data classification plays a major role in organizing the optimal features for the most effective and efficient use. Feature Selection technique is one of the foremost methods to select the optimal features from the dataset. The classification accuracy and the processing time required to build the model are the two main keys in obtaining the effective data classification by using the feature subset methods and ranking methods. The work was tested on seven real time dataset (Breast Cancer, Breast Tissue, Contact Lenses, Dermatology, Hypothyroid, Iris and Liver Disorders) obtained from UCI Data repository. The results obtained from CFS Subset Attribute Evaluator, Correlation Attribute Evaluator, Gain Ration Attribute Evaluator, Info Gain Attribute Evaluator, OneR Attribute Evaluator, Principal Components Attribute Evaluator, ReliefF Attribute Evaluator, Symmetrical Uncertainty Attribute Evaluator and Wrapper Subset Attribute Evaluator were compared. Classification algorithms like Navis Bayes, Bayes Net, Multilayered Perception, Sequential Minimum Optimization, K Nearest Neighbours, Decision Tree, OneR, J48 and Random Tree are used to analyze the classification accuracy and processing time. Comparison are done with the results obtained by using the ranking methods and the results obtained bynot using the ranking method, to find whether the ranking methods are important in obtaining the classification accuracy and processing time.***

***Keywords: Data Classification, Feature Subset Methods, Ranking Methods, Supervised learning algorithms, classification accuracy***

## I. INTRODUCTION

Manually, people's extracted patterns from data for periods. Huge volume of data accumulated because of the advanced in Information Communication and Technology (ICT) and many computerized systems entered. In pattern recognition and data classification, data plays a leading and successful element. Positioning data as to utilize at the best level of effective and efficiency are referred to data classification.
Data Mining is an automated tool used to obtain the relevant patterns from the dataset by using the specified methods and algorithms, the where the results are used to analysis the present and historic data which helps in predicting the future development [15]. Classification is used to find the group of related features which describes the data classes by using the supervised algorithms or machine learning [7].

**Revised Manuscript Received on 30 March 2019.**
***** Correspondence Author
**Sivakumar Venkataraman,** Botho University, Gaborone, Botswana
**Subitha Sivakumar,** Annai Fathima College of Arts and Science, Madurai, India



The derived class is obtained by using a set of trained features. Feature ranking is a method in data preprocessing to obtain the most relevant features were to help the model simpler and more effective [9].Feature selection plays a big role in machine learning. Feature selection is a process of selecting relevant features from the original features by removing the irrelevant and redundant features where to obtain the most relevant features [13]. Most of the feature selector techniques applied in classifications are used with ranking methods. This article studies the effect of various ranking techniques on data classification accuracy by implementing the ranking methods with the classification algorithms and to compare the results to find the efficiency of ranking methods in classification algorithms. The table 1.1 shows the ranking methods used in the experiment.

| **Evaluators** | **Referred in article** |
|---|---|
| CFS Subset Attribute Evaluator | CFS |
| Correlation Attribute Evaluator | Cor.R |
| Gain Ratio Attribute Evaluator | GR |
| Info Gain Attribute Evaluator | IG |
| OneR Attribute Evaluator | OR |
| Principal Components Attribute Evaluator | PC |
| ReliefF Attribute Evaluator | RF |
| Symmetrical Uncertainty Attribute Evaluator | SU |
| Wrapper Subset Attribute Evaluator | WS |

**Table 1.1 Ranking methods**

### CFS (Correlation based Feature Selection) Subset Attribute Evaluator

CFS Subset Attribute Evaluator method calculates the weightage of each subset features using their own predictive abilities along with the degree of redundancy among them. The feature subsets with high correlated values are selected [6]. According to [11] states that the best feature subset has only the predictive features in the classes and uncorrelated features as not predictive classes. CFS checks the importance of the features along with the predictive ability of each features and the level of redundant [5]. The CFS zare combined and executed with the ranking methods to obtain the predictive features from the dataset.

### Correlation Attribute Evaluator

Correlation Attribute Evaluator weighs the features in relation to the target classes. According to [5], Pearson correlation approach has implemented to evaluate the relations among the features and the target classes.

The Correlation Attribute Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**Gain Ratio (GR) Attribute Evaluator**

Gain Ratio Attribute Evaluator is a method to evaluate the value of each features through checking the gain ratio with respect to the classes. Gain Ratio (GA) is derived from the method Information Gain (IG). Gain Ratio considers the root nodes and branch nodes into account while selecting the features [12]. The Gain Ratio Attribute Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**Information Gain (IG) Attribute Evaluator**

Information Gain (IG) is a ranking method which determines the value of each features by evaluating the Information Gain with respect to the target classes [9]. This method selects the test feature in each node for the decision tree where the features are having the maximum values. The IG Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**OneR Attribute Evaluator**

OneR alogorithm creates a rule for every single feature and then select the rule which is having the minimum error as "One Rule" [8]. OneR algorithm performs in ranking the features based on the error rating [1]. The OneR Attribute Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**Principal Components Attribute Evaluator**

According to [2] states that Principal Components Attribute are used to get a subset of features from the original dataset. By using Principal Components Attribute are used to reduce the high-volume dataset to low volume dataset by removing the redundant features and the noisy features from the original dataset. The Principal Components Attribute Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**Symmetrical Uncertainty Attribute Evaluator**

Symmetrical Uncertainty Attribute Evaluators are used to determine the strength of the feature selector among the attributes and target classes. The attributes determining the maximum values are selected as the top priority attributes from the dataset [3]. The Symmetrical Uncertainty Attribute Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

**Wrapper Subset Attribute Evaluator**

Wrapper Subset Attribute Evaluator are used to calculate the accuracies of the machine learning algorithms by using the cross-validation techniques and uses the machine algorithm as functions in evaluating the selected subset attributes [10]. This Evaluator are combined and executed with the ranking methods to obtain the predictive features from the dataset.

## II. METHODOLOGY

Data mining is a process of selecting the best attributes or patterns from the dataset which helps to get a valid information's. WEKA is a machine learning tool which includes several machine learning algorithms to perform the datamining tasks. WEKA can perform the data pre-processing, can evaluate the attributes by using the in-built evaluation methods and can compare the results obtained by the learning algorithms in terms of number of selected features, time to build the model and classification accuracies [14]. For this study, the WEKA tool is used to obtain the optimal features from the high dimensional dataset to extract the required knowledge.

For this study, seven dataset: Breast Cancer, Breast Tissue, Contact Lenses, Dermatology, Hypothyroid, Iris and Liver Disorders are used from UCI repository [4]. Table1.2 shows the dataset used with the number of instances and attributes in the original dataset. Table 1.3 shows the different types of classification algorithms used for this study.

| Dataset Name | Attributes | Instances |
|---|---|---|
| Breast Cancer | 10 | 286 |
| Breast Tissue | 10 | 699 |
| Contact Lenses | 5 | 24 |
| Dermatology | 35 | 366 |
| Hypothyroid | 30 | 3772 |
| Iris | 5 | 150 |
| Liver Disorders | 7 | 345 |

**Table 1.2 Dataset**

| Classifier | Algorithms | Referred as |
|---|---|---|
| Bayes | Navie Bayes | NB |
| | Bayes Net | BN |
| Functions | Multilayered Perception | MP |
| | Sequential Minimum Optimization | SMO |
| Lazy | K Nearest Neighbours | K-NN |
| | Locally Weighted Learning | LWL |
| Rules | Decision Tree | DT |
| | OneR | OR |
| Tree | J48 | J48 |
| | Random Tree | RT |

**Table 1.3 Classification Algorithms**

## III. RESULTS AND DISCUSSION

Using WEKA tools, in the first stage the different classification algorithms are executed on the seven dataset to find the number of selected attributes, the time taken to build the model and the classification accuracies. The results are noted and compared in the Table 1.4 and Table 1.5. In the second stage, the different classification algorithms combined with the ranking method are executed on the dataset to find the number of selected attributes, the time taken to build the model and the classification accuracy. The results are noted and compared in the Table 1.6 to Table1.21. The Average classification accuracy and processing time for classification algorithms are shown in Table1.22

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT |
|---|---|---|---|---|---|---|---|---|---|---|
| Breast Cancer | 71.67 | 72.02 | 64.68 | 69.58 | 72.37 | 72.37 | 73.42 | 65.73 | 75.52 | 66.78 |
| Breast Tissue | 95.99 | 97.13 | 95.27 | 96.99 | 95.13 | 90.27 | 95.27 | 92.7 | 94.56 | 94.57 |
| Contact Lenses | 70.83 | 70.83 | 70.83 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 |
| Dermatology | 97.26 | 97.54 | 96.17 | 95.35 | 94.53 | 87.7 | 86.88 | 49.72 | 93.98 | 87.43 |
| Hypothyroid | 95.28 | 98.59 | 94.16 | 93.61 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 97.11 |
| Iris | 96.00 | 92.66 | 97.33 | 96 | 95.33 | 93.33 | 92.66 | 92 | 96 | 92 |
| Liver Disorders | 55.36 | 56.23 | 71.59 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 68.69 | 67.82 |
| Lung Cancer | 78.12 | 78.12 | 65.62 | 65.62 | 68.75 | 75 | 71.87 | 87.5 | 78.12 | 75 |
| Primary Tumour | 50.14 | 46.9 | 38.34 | 46.9 | 39.23 | 39.93 | 39.82 | 27.43 | 39.82 | 37.16 |
| Average | 78.96 | 78.89 | 77.11 | 77.02 | 77.66 | 75.99 | 76.88 | 70.80 | 81.07 | 76.52 |

**Table 1.4 Classification accuracy of different Classification algorithm without Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT |
|---|---|---|---|---|---|---|---|---|---|---|
| Breast Cancer | 0.01 | 0 | 4.88 | 0.06 | 0 | 0 | 0.1 | 0 | 0.01 | 0.01 |
| Breast Tissue | 0.01 | 0.02 | 0.89 | 0.04 | 0 | 0 | 0.05 | 0 | 0.02 | 0.01 |
| Contact Lenses | 0.00 | 0 | 0.05 | 0.09 | 0 | 0 | 0.01 | 0 | 0 | 0 |
| Dermatology | 0.00 | 0 | 53.43 | 0.23 | 0 | 0 | 0.17 | 0.01 | 0.03 | 0.01 |
| Hypothyroid | 0.02 | 0.05 | 38.09 | 2.72 | 0 | 0 | 1.02 | 0.02 | 0.08 | 0.02 |
| Iris | 0.00 | 0.01 | 0.13 | 0.03 | 0 | 0 | 0.02 | 0 | 0.01 | 0 |
| Liver Disorders | 0.01 | 0.01 | 0.31 | 0.02 | 0 | 0 | 0.02 | 0 | 0.002 | 0.01 |
| Lung Cancer | 0.00 | 0.01 | 5.01 | 0.03 | 0 | 0 | 0.04 | 0 | 0 | 0 |
| Primary Tumour | 0.01 | 0.01 | 5.58 | 3.34 | 0 | 0 | 0.15 | 0 | 0.05 | 0.01 |
| Average | 0.01 | 0.01 | 12.04 | 0.73 | 0.00 | 0.00 | 0.18 | 0.00 | 0.02 | 0.01 |

**Table 1.5 Processing Time of different Classification algorithm without Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 71.67 | 73.42 | 70.97 | 70.62 | 72.37 | 72.02 | 75.52 | 65.73 | 75.52 | 67.83 | 71.57 |
| GR | 72.02 | 72.37 | 66.08 | 69.23 | 70.97 | 71.32 | 75.87 | 65.73 | 75.52 | 66.08 | 70.52 |
| IG | 72.37 | 72.37 | 67.13 | 68.53 | 71.67 | 71.67 | 75.87 | 65.03 | 75.52 | 66.08 | 70.62 |
| OR | 70.97 | 71.32 | 68.18 | 69.58 | 69.58 | 71.32 | 74.82 | 66.78 | 73.42 | 66.78 | 70.28 |
| PC | 72.72 | 69.23 | 70.27 | 70.62 | 59.09 | 69.58 | 69.23 | 68.53 | 72.02 | 67.13 | 68.84 |
| RF | 69.58 | 69.58 | 68.18 | 68.53 | 69.58 | 70.27 | 71.32 | 68.53 | 70.97 | 65.38 | 69.19 |
| SU | 72.02 | 72.37 | 68.18 | 69.23 | 70.97 | 71.32 | 75.87 | 65.03 | 75.52 | 64.68 | 70.52 |
| Classification Avg | 71.62 | 71.52 | 68.43 | 69.48 | 69.18 | 71.07 | 74.07 | 66.48 | 74.07 | 66.28 | |

**Table 1.6 Classification accuracy on selected features for Breast Cancer Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 95.99 | 97.13 | 95.56 | 96.85 | 95.27 | 90.27 | 94.13 | 91.41 | 94.56 | 93.41 | 94.46 |
| GR | 95.99 | 97.13 | 94.56 | 96.99 | 95.27 | 90.27 | 94.42 | 92.7 | 94.56 | 93.84 | 94.57 |
| IG | 95.99 | 97.13 | 95.42 | 96.99 | 95.27 | 90.27 | 95.27 | 92.27 | 94.56 | 93.56 | 94.67 |
| OR | 95.99 | 97.13 | 95.42 | 96.99 | 95.27 | 90.27 | 94.13 | 91.41 | 94.56 | 93.7 | 94.49 |
| PC | 94.99 | 95.27 | 95.99 | 96.85 | 96.28 | 96.85 | 96.13 | 96.99 | 96.56 | 95.85 | 96.18 |
| RF | 95.99 | 97.13 | 95.13 | 96.99 | 95.27 | 90.27 | 94.13 | 94.41 | 94.56 | 93.7 | 94.76 |
| SU | 95.99 | 97.13 | 95.42 | 96.85 | 95.27 | 90.27 | 95.27 | 92.7 | 94.56 | 94.7 | 94.82 |
| Classification Avg | 95.85 | 96.86 | 95.36 | 96.93 | 95.41 | 91.21 | 94.78 | 93.13 | 94.85 | 94.11 | |

**Table 1.7 Classification accuracy on selected features for Breast Tissue Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 70.83 | 70.83 | 70.83 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 | 73.33 |
| GR | 70.83 | 70.83 | 70.83 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 | 73.33 |
| IG | 70.83 | 70.83 | 70.83 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 | 73.33 |
| OR | 70.83 | 70.83 | 66.67 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 75 | 73.33 |
| PC | 58.33 | 66.67 | 70.83 | 54.16 | 50 | 66.67 | 58.33 | 62.5 | 70.83 | 66.66 | 62.5 |
| RF | 70.83 | 70.83 | 66.67 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 | 72.91 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SU | 70.83 | 70.83 | 70.83 | 70.83 | 79.16 | 70.83 | 75 | 70.83 | 83.33 | 70.83 | 73.33 |
| Classification Avg | 69.04 | 70.24 | 69.64 | 68.45 | 74.99 | 70.24 | 72.62 | 69.64 | 81.54 | 70.83 | |

**Table 1.8 Classification accuracy on selected features for Contact Lenses Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 97.26 | 97.54 | 96.44 | 95.35 | 94.53 | 87.97 | 97.7 | 49.72 | 93.98 | 83.6 | 89.41 |
| GR | 97.26 | 97.54 | 95.9 | 95.35 | 94.53 | 87.97 | 87.7 | 48.9 | 93.98 | 85.79 | 88.49 |
| IG | 97.26 | 97.54 | 96.17 | 95.35 | 94.53 | 87.97 | 87.43 | 50.27 | 93.98 | 90.43 | 89.09 |
| OR | 97.26 | 97.54 | 95.9 | 95.5 | 94.53 | 87.7 | 87.43 | 50.27 | 93.98 | 87.7 | 88.78 |
| PC | 95.35 | 93.98 | 95.62 | 96.99 | 90.16 | 46.72 | 88.52 | 72.67 | 93.71 | 70.76 | 84.45 |
| RF | 97.26 | 97.54 | 95.9 | 95.35 | 94.53 | 87.97 | 88.25 | 50.27 | 93.98 | 89.61 | 89.07 |
| SU | 97.26 | 97.54 | 96.17 | 95.35 | 94.53 | 87.97 | 87.7 | 49.72 | 93.98 | 85.79 | 88.6 |
| Classification Avg | 96.99 | 97.03 | 96.01 | 95.61 | 93.91 | 82.04 | 89.25 | 53.12 | 93.94 | 84.81 | |

**Table 1.9 Classification accuracy on selected features for Dermatology Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 95.28 | 98.59 | 94.64 | 93.58 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 97.95 | 96.21 |
| GR | 95.28 | 98.59 | 94 | 93.58 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 98.4 | 96.19 |
| IG | 95.28 | 98.59 | 93.74 | 93.61 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 97.66 | 96.09 |
| OR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PC | 82.58 | 93.37 | 96.68 | 93.39 | 92.52 | 95.52 | 93.84 | 92.47 | 93.31 | 90.82 | 92.45 |
| RF | 95.28 | 98.59 | 93.9 | 93.58 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 98.06 | 96.14 |
| SU | 95.28 | 98.59 | 93.76 | 93.58 | 91.51 | 95.38 | 99.33 | 96.23 | 99.57 | 98.01 | 96.12 |
| Classification Avg | 79.85 | 83.76 | 80.96 | 80.19 | 78.58 | 81.77 | 84.36 | 81.95 | 84.45 | 82.99 | |

**Table 1.10 Classification accuracy on selected features for Hypothyroid Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 96 | 92.66 | 97.33 | 96 | 95.33 | 93.33 | 92.66 | 92 | 99.57 | 97.95 | 95.28 |
| GR | 96 | 92.66 | 97.33 | 96 | 95.33 | 93.33 | 92.66 | 93.33 | 99.57 | 98.4 | 95.46 |
| IG | 96 | 92.66 | 98 | 96 | 95.33 | 93.33 | 94 | 93.33 | 99.57 | 97.66 | 95.59 |
| OR | 96 | 92.66 | 97.33 | 96 | 95.33 | 93.33 | 94 | 94.66 | 0 | 0 | 75.93 |
| PC | 88 | 86.66 | 89.33 | 88 | 84.66 | 86 | 86.66 | 91.33 | 93.31 | 90.82 | 88.48 |
| RF | 96 | 92.66 | 98 | 96 | 95.33 | 93.33 | 94.66 | 95.33 | 99.57 | 98.06 | 95.89 |
| SU | 96 | 92.66 | 97.33 | 96 | 95.33 | 93.33 | 92.66 | 93.33 | 99.57 | 98.01 | 95.42 |
| Classification Avg | 94.86 | 91.8 | 96.38 | 94.86 | 93.81 | 92.28 | 92.47 | 93.33 | 84.45 | 82.99 | |

**Table 1.11 Classification accuracy on selected features for Iris Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 55.36 | 56.23 | 69.85 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 66.69 | 64.63 | 60.58 |
| GR | 55.36 | 56.23 | 68.4 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 66.69 | 62.6 | 60.23 |
| IG | 55.36 | 56.23 | 69.56 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 68.69 | 62.31 | 60.52 |
| OR | 55.36 | 56.23 | 70.43 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 68.69 | 65.5 | 60.92 |
| PC | 54.78 | 57.97 | 61.15 | 57.97 | 57.1 | 57.68 | 57.97 | 53.33 | 55.65 | 52.75 | 56.64 |
| RF | 55.36 | 56.23 | 70.43 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 68.69 | 65.5 | 60.92 |
| SU | 55.36 | 56.23 | 68.4 | 58.26 | 62.89 | 59.13 | 57.68 | 55.07 | 68.69 | 62.6 | 60.43 |
| Classification Avg | 55.28 | 56.48 | 68.32 | 58.22 | 62.06 | 58.92 | 57.72 | 54.82 | 66.26 | 62.27 | |

**Table1.12 Classification accuracy on selected features for Liver Disorders Dataset with ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 83.2 | 83.77 | 85.09 | 83.07 | 84.44 | 81.28 | 84.57 | 74.43 | 87.6 | 82.31 | 82.98 |
| GR | 83.25 | 83.62 | 83.87 | 82.89 | 84.24 | 81.18 | 83.24 | 74.68 | 87.6 | 82.28 | 82.68 |
| IG | 83.3 | 83.62 | 84.41 | 82.8 | 84.34 | 81.23 | 83.51 | 74.72 | 87.89 | 82.65 | 82.85 |
| OR | 69.49 | 69.39 | 70.56 | 69.59 | 70.97 | 67.51 | 69.01 | 61.29 | 59.14 | 55.53 | 66.25 |
| PC | 78.11 | 80.45 | 82.84 | 79.71 | 75.69 | 74.15 | 78.67 | 76.83 | 82.2 | 76.4 | 78.5 |
| RF | 82.9 | 83.22 | 84.03 | 82.79 | 84.04 | 81.03 | 82.91 | 75.81 | 87.24 | 83.02 | 82.7 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SU | 83.25 | 83.62 | 84.3 | 82.87 | 84.24 | 81.18 | 83.36 | 74.7 | 87.89 | 82.09 | 82.75 |
| Classification Avg | 80.5 | 81.1 | 82.16 | 80.53 | 81.13 | 78.22 | 80.75 | 73.21 | 82.79 | 77.75 | |

**Table 1.13 Average classification accuracy with ranking methods**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.01 | 0.01 | 2.74 | 0.06 | 0.01 | 0.01 | 0.04 | 0.01 | 0.02 | 0.01 | 0.29 |
| GR | 0.01 | 0 | 4.25 | 0.06 | 0 | 0 | 0.03 | 0 | 0 | 0 | 0.44 |
| IG | 0 | 0.01 | 4.61 | 0.06 | 0.01 | 0 | 0.02 | 0 | 0.01 | 0 | 0.47 |
| OR | 0.05 | 0.04 | 2.53 | 0.07 | 0.04 | 0.04 | 0.06 | 0.03 | 0.04 | 0.04 | 0.29 |
| PC | 0.21 | 0.1 | 0.34 | 0.17 | 0.02 | 0.07 | 0.05 | 0.08 | 0.04 | 0.08 | 0.12 |
| RF | 0.04 | 0.05 | 2.93 | 0.09 | 0.04 | 0.04 | 0.06 | 0.04 | 0.04 | 0.05 | 0.34 |
| SU | 0 | 0 | 4.58 | 0.06 | 0 | 0.01 | 0.02 | 0 | 0 | 0.01 | 0.47 |
| Classification Avg | 0.05 | 0.03 | 3.14 | 0.08 | 0.02 | 0.02 | 0.04 | 0.02 | 0.02 | 0.03 | |

**Table 1.14 Processing Time of different Classification algorithm for Breast Cancer with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.04 | 0.01 | 0.87 | 0.05 | 0.01 | 0.01 | 0.06 | 0.01 | 0.03 | 0.02 | 0.11 |
| GR | 0.01 | 0.01 | 0.8 | 0.01 | 0.02 | 0.02 | 0.05 | 0.02 | 0.01 | 0.01 | 0.10 |
| IG | 0.01 | 0.01 | 0.81 | 0.02 | 0.01 | 0.01 | 0.06 | 0.01 | 0.02 | 0.01 | 0.10 |
| OR | 0.06 | 0.04 | 0.9 | 0.06 | 0.07 | 0.09 | 0.011 | 0.05 | 0.05 | 0.06 | 0.14 |
| PC | 0.04 | 0.02 | 0.63 | 0.03 | 0.04 | 0.12 | 0.07 | 0.03 | 0.02 | 0.03 | 0.10 |
| RF | 0.2 | 0.2 | 98 | 0.02 | 0.25 | 0.2 | 0.26 | 0.2 | 0.19 | 0.2 | 9.97 |
| SU | 0.01 | 0.01 | 0.87 | 0.03 | 0.01 | 0.01 | 0.05 | 0.01 | 0.02 | 0.01 | 0.10 |
| Classification Avg | 0.05 | 0.04 | 14.70 | 0.03 | 0.06 | 0.07 | 0.08 | 0.05 | 0.05 | 0.05 | |

**Table 1.15 Processing Time of different Classification algorithm for Breast Tissue with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.00 | 0.01 | 0.04 | 0.03 | 0.05 | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.02 |
| GR | 0.00 | 0.00 | 0.03 | 0.09 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 |
| IG | 0.01 | 0.00 | 0.03 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| OR | 0.02 | 0.02 | 0.04 | 0.03 | 0.02 | 0.02 | 0.01 | 0.02 | 0.02 | 0.02 | 0.02 |
| PC | 0.01 | 0.01 | 0.03 | 0.02 | 0.05 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.02 |
| RF | 0.01 | 0.00 | 0.05 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| SU | 0.00 | 0.00 | 0.03 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| Classification Avg | 0.01 | 0.01 | 0.04 | 0.03 | 0.02 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | |

**Table 1.16 Processing Time of different Classification algorithm for Contact Lenses with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.0 | 0.0 | 42.3 | 0.9 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 4.36 |
| GR | 0.02 | 0.01 | 40.72 | 0.38 | 0.01 | 0.01 | 0.17 | 0.01 | 0.02 | 0 | 4.14 |
| IG | 0.01 | 0.02 | 40.14 | 0.015 | 0.01 | 0.01 | 0.17 | 0 | 0.01 | 0.01 | 4.04 |
| OR | 0.13 | 0.11 | 40.74 | 0.024 | 0.13 | 0.11 | 0.28 | 0.1 | 0.1 | 0.1 | 4.18 |
| PC | 0.39 | 0.29 | 13.93 | 0.51 | 0.26 | 0.25 | 0.62 | 0.28 | 0.19 | 0.17 | 1.69 |
| RF | 0.2 | 0.18 | 40.93 | 0.04 | 0.18 | 0.18 | 0.4 | 0.18 | 0.19 | 0.18 | 4.27 |
| SU | 0 | 0.02 | 41.76 | 0.024 | 0.01 | 0.01 | 0.17 | 0.01 | 0.01 | 0.01 | 4.20 |
| Classification Avg | 0.11 | 0.09 | 37.22 | 0.27 | 0.09 | 0.09 | 0.29 | 0.09 | 0.08 | 0.07 | |

**Table 1.17 Processing Time of different Classification algorithm for Dermatology with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.09 | 0.07 | 36.94 | 2.99 | 0.03 | 0.04 | 1.16 | 0.05 | 0.14 | 0.05 | 4.16 |
| GR | 0.04 | 0.07 | 40.74 | 3.65 | 0.03 | 0.03 | 1.06 | 0.05 | 0.13 | 0.03 | 4.58 |
| IG | 0.04 | 0.04 | 41.55 | 2.73 | 0.04 | 0.03 | 1.11 | 0.06 | 0.06 | 0.04 | 4.57 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| PC | 0.02 | 0.19 | 25.19 | 1.15 | 0.16 | 0.14 | 0.92 | 0.17 | 0.41 | 0.2 | 2.86 |
| RF | 13.27 | 12.68 | 50.16 | 16.14 | 13.41 | 12.37 | 13.39 | 12.15 | 12.47 | 12.19 | 16.82 |
| SU | 0.03 | 0.05 | 39.38 | 3.34 | 0.04 | 0.04 | 1.06 | 0.04 | 0.06 | 0.04 | 4.41 |
| Classification Avg | 1.93 | 1.87 | 33.42 | 4.29 | 1.96 | 1.81 | 2.67 | 1.79 | 1.90 | 1.79 | |

**Table 1.18 Processing Time of different Classification algorithm for Hypothyroid with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.00 | 0.01 | 0.12 | 0.06 | 0.00 | 0.00 | 0.02 | 0.00 | 0.14 | 0.07 | 0.04 |
| GR | 0.01 | 0.00 | 0.11 | 0.02 | 0.01 | 0.00 | 0.01 | 0.00 | 0.14 | 0.04 | 0.03 |
| IG | 0.00 | 0.01 | 0.11 | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 | 0.05 | 0.03 | 0.02 |
| OR | 0.03 | 0.02 | 0.15 | 0.03 | 0.01 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.03 |
| PC | 0.05 | 0.01 | 0.08 | 0.07 | 0.00 | 0.01 | 0.01 | 0.01 | 0.33 | 0.19 | 0.08 |
| RF | 0.02 | 0.01 | 0.16 | 0.03 | 0.01 | 0.02 | 0.01 | 0.02 | 12.32 | 12.23 | 2.48 |
| SU | 0.00 | 0.00 | 0.11 | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 | 0.06 | 0.05 | 0.03 |
| Classification Avg | 0.02 | 0.01 | 0.12 | 0.04 | 0.00 | 0.01 | 0.01 | 0.01 | 1.86 | 1.80 | |

**Table 1.19 Processing Time of different Classification algorithm for Iris with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.02 | 0.01 | 0.32 | 0.03 | 0 | 0.01 | 0.09 | 0 | 0.02 | 0.02 | 0.05 |
| GR | 0.01 | 0.01 | 0.27 | 0.01 | 0.01 | 0 | 0.01 | 0 | 0.01 | 0.01 | 0.03 |
| IG | 0.01 | 0.01 | 0.29 | 0.02 | 0 | 0 | 0.02 | 0.01 | 0.01 | 0.01 | 0.04 |
| OR | 0.04 | 0.03 | 0.3 | 0.04 | 0.03 | 0.02 | 0.05 | 0.04 | 0.04 | 0.04 | 0.06 |
| PC | 0.06 | 0.01 | 0.36 | 0.02 | 0.02 | 0.01 | 0.18 | 0.02 | 0.03 | 0.03 | 0.07 |
| RF | 0.12 | 0.05 | 0.32 | 0.04 | 0.05 | 0.04 | 0.16 | 0.05 | 0.06 | 0.06 | 0.10 |
| SU | 0 | 0 | 0.28 | 0.04 | 0 | 0 | 0.01 | 0.01 | 0.02 | 0.01 | 0.04 |
| Classification Avg | 0.04 | 0.02 | 0.31 | 0.03 | 0.02 | 0.01 | 0.07 | 0.02 | 0.03 | 0.03 | |

**Table 1.20 Processing Time of different Classification algorithm for Liver Disorders with Ranking**

| Dataset/Algorithm | NB | BN | MP | SMO | KNN | LWL | DT | OR | J48 | RT | FS Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cor.R | 0.03 | 0.02 | 11.90 | 0.59 | 0.02 | 0.01 | 0.23 | 0.01 | 0.06 | 0.03 | 1.29 |
| GR | 0.01 | 0.01 | 12.42 | 0.60 | 0.01 | 0.01 | 0.19 | 0.01 | 0.04 | 0.01 | 1.33 |
| IG | 0.01 | 0.01 | 12.51 | 0.41 | 0.01 | 0.01 | 0.20 | 0.01 | 0.02 | 0.01 | 1.32 |
| OR | 0.05 | 0.04 | 6.38 | 0.04 | 0.04 | 0.04 | 0.06 | 0.04 | 0.04 | 0.04 | 0.68 |
| PC | 0.11 | 0.09 | 5.79 | 0.28 | 0.08 | 0.09 | 0.27 | 0.09 | 0.15 | 0.10 | 0.70 |
| RF | 1.98 | 1.88 | 27.51 | 2.34 | 1.99 | 1.84 | 2.04 | 1.81 | 3.61 | 3.56 | 4.86 |
| SU | 0.01 | 0.01 | 12.43 | 0.50 | 0.01 | 0.01 | 0.19 | 0.01 | 0.02 | 0.02 | 1.32 |
| Classification Avg | 0.31 | 0.30 | 12.71 | 0.68 | 0.31 | 0.29 | 0.45 | 0.28 | 0.56 | 0.54 | |

**Table 1.21 Average processing time with ranking methods**

| Classification Algorithms | Ranked Full-Set | |
|---|---|---|
| | Classification Avg | Processing Time Avg |
| Navie Bayes | 80.5 | 0.31 |
| Bayes Net | 81.1 | 0.3 |
| Multilayered Perception | 82.16 | 12.71 |
| Sequential Minimum Optimization | 80.53 | 0.68 |
| K Nearest Neighbours | 81.13 | 0.31 |
| Decision Tree | 80.75 | 0.45 |
| OneR | 73.21 | 0.28 |
| J48 | 82.79 | 0.56 |
| Random Tree | 77.75 | 0.54 |

**Table 1.22 Average classification accuracy and Processing time for classification algorithms**

## IV. CONCLUSION

The different dataset is experimented by using the ranking methods and by non-ranking methods with the classification algorithms to find whether the ranking methods improves the performance in data classification. The result analysis shows that the classification algorithms deliver different classification accuracies when using the ranking methods. Study shows that ranking method delivers the best result and improves the classification accuracy. Table 1.22 shows that J48 algorithm performs better than other classification algorithms in terms of classification accuracy with ranking method. The OneR algorithm is having the less classification accuracy when compared to the other algorithms for the selected dataset. Looking at the processing time the OneR and Bayes yields better result for the selected dataset. Multilayered Perception algorithm takes more time to build a model when compared to other state-of-art algorithms.

## REFERENCES

1. M. Aditi, and G. Anita, Performance Evaluation of Rule Based Classification Algorithms. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), October 2010, 3(10).
2. G. J. Andreas, G. N. Wilfried, D. A. Michael and E. F. Gerhard, On the Relationship Between Feature Selection and Classification Accuracy. JMLR: Workshop and Conference Proceedings, 2008, Volume 4, pp. 90-105.
3. S. Bharat, K. Nidhi and V. Om Prakash, A Feature Subset Selection Technique for High Dimensional Data using Symmetric Uncertainty. Journal of Data Analysis and Information Processing, 2014, 2(4).
4. D. Dua and E. Karra Taniskidou, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2017.
5. S. Gnanambal, M. Thangaraj, V. Meenatchi and V. Gayathri, Classification Algorithms with Attribute Selection: an evaluation study using WEKA. Advanced Networking and Applications, 2018, 9(6), pp. 3640-3644.
6. M. A. Hall, Correlation-based Feature Subset Selection for Machine Learning, Hamilton, Newland: The University of Waikato, 1998.
7. J. P. Hem, V. Singh and V. Nisha, An Efficient Classification Approach for Data Minin. International Journal of Machine Learning and Computing, August 2012, 2(4).
8. R. Hotle, Very simple classification rules perform well on most commonly used dataset. Machine Learning, 1993, 11(1), pp. 63-91.
9. S. Ilangovan, K. Antony and B. Appavu, An Empirical Study on Different Ranking Methods for Effective Data Classification. Journal of Modern Applied Statistical Methods, November 2015, 14(2), pp. 35-52.
10. A. J. B. Kadhim and K. Rusul, Data Reduction Techniques: A Comparative Study for Attribute Selection Methods. International Journal of Advanced Computer Science and Technology, 2018, 8(1), pp. 1-13.
11. P. R. Praveena, M. Valarmathi and S. Sivakumari, Gain Ratio Based Feature Selection method for privacy preservation. ICTACT Journal of soft computing, April 2011, 1(4).
12. B. Sunita and A. Jitender, Classification and Feature Selection Techniques in Data Mining. International Journal of Engineering Research & Technology (IJERT), August 2012, 1(6).
13. C. S. Trilok and J. Manoj, WEKA Approach for Comparative Study of Classification Algorithm. International Journal of Advanced Research in Computer and Communication Engineering, April 2013, 2(4).
14. Vishal and G. Saurabh, Data Mining tools and Techniques. International Journal of Scientific & Engineering Research, April 2014.