

Comparison of Feature Ranking Methods for Effective Data Classification

¹S. Ilangovan, ²Vincent Antonykumar and ³S. Appavu Balamurugan

¹Department of Information Technology, K.L.N. College of Engineering, Madurai, Tamilnadu, India

²Department of Information Technology,
PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India

³Department of Information Technology,
K.L.N. College of Information Technology, Madurai, Tamilnadu, India

Abstract: Ranking is the process of selecting the attributes based on their values over different criteria in relation to other attributes for a given set of attributes. In data mining, ranking is an important pre-processing technique to select the most relevant attributes in order to produce the accurate and compatible results. In this study, we considered eight well known ranking methods and these methods are applied on ten real world datasets. The quality of feature rankings obtained by these methods is evaluated using eight learning algorithms in terms of classification accuracy and running time. The experiment results give quite different results for different ranking methods and significant differences are found in some cases, but there is no single best ranking method that works best for all data and all classifiers.

Key words: Ranking methods • Feature selection • Classification accuracy • Data mining

INTRODUCTION

As computer power grows and data collection technologies advance, a plethora of data is generated in almost every field where computers are used [1]. In this digital world, information is being created at ever increasing rate and this will lead to increase in size of data. Such a huge data will amplify the challenges in protecting and managing data. Handling of such a massive data has become critical for success. But the business depends on fast and reliable access. Data mining is a process of discovering knowledge from these huge volumes of data. However, these data are usually associated with a high level of noise and redundant data. Discovering useful knowledge and patterns from such a huge data is a challenging task. All features may not be important for some problems, only a small subset of features is usually relevant. A high dimensional dataset with irrelevant, redundant and noise features making the knowledge discovery process worthless [2],[3]. To deal with such problems dimensionality of the feature space has to be reduced first. This may be done by selecting a subset of relevant features from the total number of features, or by

ranking these features and selecting the most important ones. Feature selection and ranking are the two long existing methods that deal with these problems. The objective of these methods are to improve the prediction performance by lowering computational complexity with reduced data [4]. In this paper, eight important ranking methods such as Relief, GainRatio, Information Gain, OneR, Symmetrical Uncertainty, Chi-squared, SVM and Filter are selected and the quality of these methods are evaluated. Significant differences are found in some cases and the experimental results shows that there is no single best ranking method that works best for all data and all classifiers. The paper is organized as follows, section 2, reviews about the ranking methods, Section 3 describes the datasets used in experiments and fourth section discusses about the experimental results. Finally, in Section 5, the most interesting conclusions are summarized.

Literature Review: Ranking is one of the attribute selection method used in the pre-processing phase in data mining to select the most relevant attributes which allow models of classification simpler and easy to understand.

It is a very important and a central task for information retrieval, such as web search engines, recommendation systems and advertisement systems. There are various ranking methods available for feature selection many of such feature ranking methods have been proposed in the literature [5], [6], [7], [8].

RELIEF is one such a very popular instance based feature ranking method used in binary classification. This method is a simple, fast and effective approach to attribute weighting. The output of the Relief algorithm is a weight between -1 and 1 for each attribute, with more positive weights indicating more predictive attributes. This method does not depend on heuristics which is proposed by Kira and Rendell in 1992 [9]. Relief [10] has proved to be a successful feature selector but when handling a large dataset, it is computationally expensive. It is noise-tolerant, robust to feature interactions, but low numbers of training instances may fool the method. Another ranking method called Information gain [11] is used to determine which attribute in a given set of training feature vectors is most useful for discriminating between the classes to be learned. This measure is based on Shannon's information theory. It tells how important a given attribute of the feature vectors is, which is used to ordering of attributes in the nodes of a decision tree. The Information gain method is biased toward tests with many outcomes. The modification of the IG is called as GainRatio [GR] that reduces its bias. GR takes number and size of branches into account when choosing an attribute. It corrects the information gain by taking the intrinsic information of a split into account. Intrinsic information is entropy of distribution of instances into branches. Value of attribute decreases as intrinsic information gets larger [12]. The Gain Ratio is the non-symmetrical measure that is introduced to compensate for the bias of the IG [13]. GainRatio is given by:

$$G(R) = IG/H(X)$$

OneR, short for "One Rule", is a simple and accurate ranking method that generates one rule for each attribute in the data and then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, it constructs a frequency table for each predictor against the target. It has been shown that OneR produces rules only slightly less accurate than state-of-the-art classification algorithms while producing rules that are simple for humans to interpret. This method is proposed by Holte [14]. Symmetric uncertainty means how much

information is partaged between X and Y relatively at all information contained in both X and Y. This is one of normalized form of Mutual Information which is introduced by Witten and Frank in 2005 [15]. Its defined as below:

$$SU = 2 \left(\frac{IG}{H(Y) + H(X)} \right)$$

Since this method is symmetric in nature, it reduces the number of comparisons required and it is not influenced by multivalued attributes as that is in the case of information gain and its values are normalized. Another most commonly used and easiest ranking method is Chi square [16]. It evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. More specifically in feature selection we use it to test whether the occurrence of a specific term and the occurrence of a specific class are independent. Thus we estimate the following quantity for each term and we rank them by their score:

$$X^2(D, t, c) = \sum_{e_t \in \{1,0\}} \sum_{e_c \in \{1,0\}} \frac{(N_{e_t e_c} - E_{e_t e_c})^2}{E_{e_t e_c}}$$

High scores on x^2 indicate that the null hypothesis (H_0) of independence should be rejected and thus that the occurrence of the term and class are dependent. If they are dependent then we select the feature for the text classification. Support Vector Machine (SVM) is another promising method and the main idea behind this algorithm is to map input vectors into a feature space of higher dimension, construct a linear decision surface and then optimize that hyper plane for generalization. The algorithm of ranking SVM [17] was published by Thorsten Joachims in 2003. SVMs are used for classification, regression and ranking and are used for related tasks such as information retrieval and optical character recognition. Filter method [18] is used to obtain the reduced or ranked data and not just output the selected or ranked attributes. Each attribute is evaluated individually and the results are in either ranked list of attributes or subset of attributes.

Proposed Work

Weka Workbench: Weka is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka stands for "Waikato Environment for Knowledge Analysis" and it is a free offering from University of Waikato, New Zealand. It is available under the GNU General Public License.

Table 1: Datasets used in the Experiment

Datasets	Attributes	Instances
Diabetes	9	768
Segment-challenge	20	1500
Soybean	36	683
Vote	17	435
Ionosphere	35	351
Dermatology	35	366
Lung cancer	57	32
Wine	14	178
Hepatitis	20	155
Vehicle	19	846

This tool has a user-friendly interface and it incorporates numerous options to develop and evaluate machine learning models [19]. The algorithms can either be applied directly to a dataset or called from your own Java code. It also supports several standard data mining tasks, more specifically, data reprocessing, clustering, classification, regression, visualization and feature selection. Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query.

Datasets Used in Experiments: To compare different feature ranking methods on dataset for which the importance of feature is known, ten real datasets were used from the UCI repository of machine learning databases [20]. A list of datasets used in our study is listed in the Table 1. This table contains ten datasets along with their characteristics, number of attribute. A brief summary of datasets is described in Table 1. From these datasets the diabetes dataset has 9 attributes and 768 records which are obtained from diabetes patients. The next segment-challenge data contains 20 attributes with 1500 instances. Similarly soybean, vote and Ionosphere datasets have 36, 17, 35 attribute and 683,435,351 instances respectively. The dermatology dataset contains 366 instances with 35 attributes, 33 of which are linear valued and one of them is nominal. The Lung cancer dataset has 57 attributes and 32 instances and it described 3 types of pathological lung cancers. The wine data has 14, 178 attributes and records. All attributes are continuous. Hepatitis dataset has 20 attributes and 155 instances and vehicle data has 19 attributes with 846 records.

RESULT AND DISCUSSION

In this study, we experimentally evaluate the effectiveness of the different ranking methods. All this ranking techniques are evaluated in terms of predictive

accuracy and running time for a classifier on selected features. The evaluation is done for all eight ranking methods, for each dataset we run all the ranking methods and select the half of the top ranked attribute for classification. For the validation of the ranking methods we have tested the classification accuracy against different classifier like NB, J48, SMO, JRIP, Decision Table, Random forest, MLP and Kstar. For this experimental evaluation, we have used a machine learning software tool Weka. All experiments were performed on Intel core i3 CPU running at 3.4 GHz and 4 GB RAM and the results are reported in the following sections. Table 2 shows the classification accuracy and running time by the classifiers against ten UCI datasets without ranking. The average performance of all the classifiers in terms of classification accuracy and running time is represented in Figure 1 and 2. From Table 2, we observed that RandomForest performs better than other classifiers and it is also observed that the classifiers like MLP, J48 and SMO also perform equally well as compared with the RandomForest except Naïvebayes. Naivebayes shows poor performance in terms of classification accuracy. In terms of speed, the classifiers like Kstar, NB, J48 and JRIP take less time and Random forest, Decisiontable, SMO takes moderate time to build the model. But the MLP classifier takes much time to build the model with highest classification accuracy. The NB takes less time to build the model with lowest classification performance. From these observations, the RandomForest classifier performs well in terms of accuracy and processing time among the all other classification methods against the datasets without ranking.

In the second study, for each of the ten data sets, we applied all the ranking methods and retained half of the top attributes from the original dataset. The Table 3 shows the number of attributes selected by ranking methods in each dataset. For each of the selected data sets, we built models using different classifiers. The 10 fold cross validation is performed and the results obtained from the ranked datasets have been tabulated in Table 4 and 5. For ease of comparison and to make the accuracy percentages meaningful, the accuracy percentages were rounded to the nearest integer. These tables give the performance of all the eight ranking methods against the eight classifiers in terms of accuracy and speed.

From the experimental results, this table itself advocated that a ranked reduced subset improves the classification accuracy for some of the classifiers and most of the classifiers perform relatively constant.

Table 2: Classification accuracy and running Time of different Classification algorithm for full datasets without ranking

S.No.	Dataset	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
		%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
1.	Diabetes	76.3	0.01	73.8	0.01	77.3	0.06	76.0	0.09	71.2	0.17	73.8	0.16	75.3	0.66	69.1	0
2.	Segment	81.1	0.02	95.7	0.04	91.9	0.28	93.7	0.14	87.4	0.16	96.9	0.13	96.7	5.31	96.6	0
3.	Soybean	92.97	0	91.5	0.03	93.85	1.12	91.94	0.09	84.33	0.24	92.09	0.19	93.41	31.8	87.99	0
4.	Vote	90.11	0	96.32	0	96.09	0.01	95.4	0.02	94.94	0.15	95.63	0.15	94.71	0.84	93.33	0
5.	ionosphere	82.62	0.01	91.45	0.08	88.6	0.08	89.74	0.06	89.45	0.05	92.87	0.04	91.16	2.27	84.61	0
6.	Dermatology	97.3	0.01	94	0.03	95.4	0.21	86.9	0.07	86.7	0.19	94.8	0.14	96.2	23.7	96.7	0
7.	Lung cancer	50.0	0	50.0	0	40.6	0.02	43.8	0	62.5	0.02	56.2	0.01	37.5	5.75	62.5	0
8.	Wine	97.2	0	93.8	0.01	98.3	0.02	91.6	0.1	88.8	0.01	97.2	0.01	97.2	0.29	97.8	0
9.	Hepatitis	84.5	0	83.9	0.05	85.2	0.01	78	0.01	76.1	0.04	80	0.09	80	0.36	83.9	0
10.	Vehicle	44.8	0.01	72.5	0.04	74.3	0.05	68.6	0.06	65.7	0.07	77	0.07	81.7	2.09	66.2	0
Classification Average		79.7	0.01	84.3	0.03	84.2	0.19	81.6	0.06	80.7	0.11	85.7	0.10	84.4	7.301	83.8	0

Table 3: Number of features selected by ranking methods

Datasets	Features	Selected Features
Diabetes	9	4
Segment-challenge	20	10
Soybean	36	18
Vote	17	8
Ionosphere	35	17
Dermatology	35	17
Lung cancer	57	23
Wine	14	7
Hepatitis	20	10
Vehicle	19	10
Total Features	262	124

Table 4: Average performance of ranking methods for all datasets

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	MLP	Kstar	Avg.
Relief	80.81	84.17	83.61	82.42	80.69	85.35	84.81	83.57	83.2
GainRatio	78.86	82.42	82.16	80.80	79.51	83.51	83.71	81.65	81.6
InfoGain	80.75	84.48	84.17	84.04	80.40	84.43	84.95	83.77	83.5
OneR	78.87	82.37	82.19	82.74	80.26	83.22	84.14	82.36	82.0
SU	80.72	84.48	84.49	82.57	80.70	85.16	84.77	83.58	83.3
Chi-squared	78.93	83.06	81.87	81.52	80.50	82.31	84.23	83.33	82.0
SVM	81.52	85.86	86.03	84.93	81.02	86.97	87.87	85.16	84.9
Filter	80.75	84.43	84.16	84.39	80.40	84.36	85.95	83.63	83.5
Average Classification	80.2	83.9	83.6	82.9	80.4	84.4	85.1	83.4	83.0

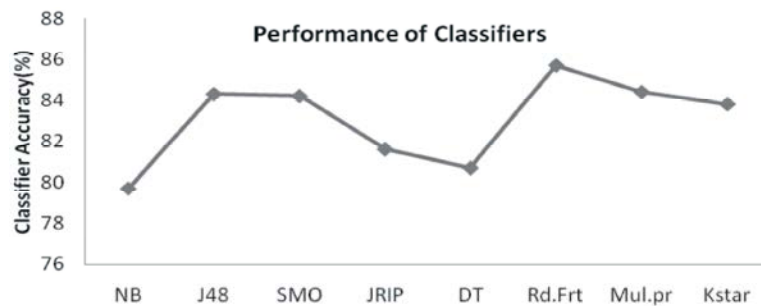


Fig. 1: Average classifier accuracy for all datasets

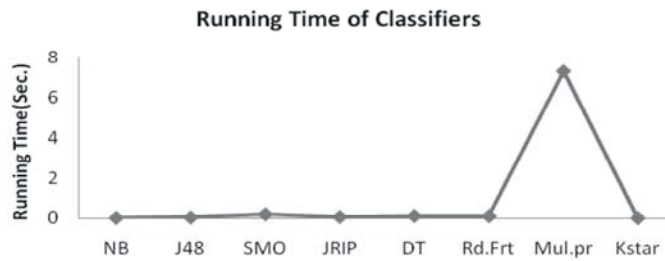


Fig. 2: Average running time of classifiers for all datasets

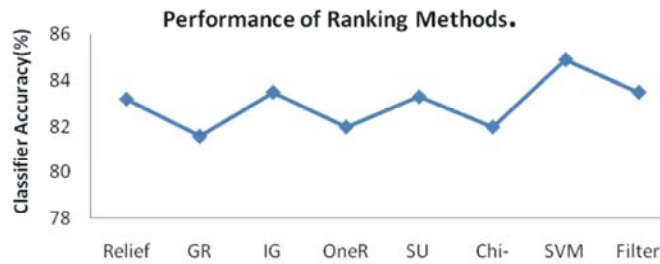


Fig. 3: Average performance of ranking methods with ranked datasets

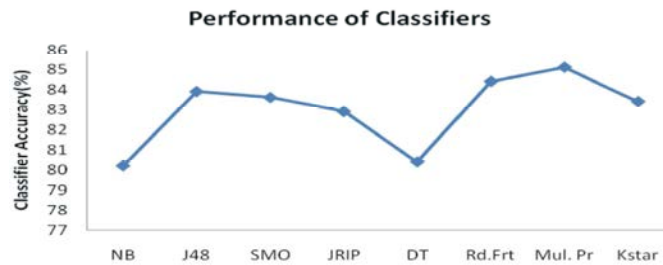


Fig. 4: Average performance of classifiers with ranked datasets

The average classification accuracy and running time of all datasets for each ranking methods against the classification algorithms is shown in Figures 3 and 4. Ranking methods with different classification methods gives different accuracy. Hence selection of ranking method is an important task for improving the classification accuracy. In this experiment, the Support Vector Machine ranking method does not run against soybean, Dermatology and Lung cancer data sets. This is because these data sets are containing a multiclass data set a discrete values. But from the average accuracy over all datasets, SVM scores 84.9 percentages. This method also scores a maximum average classification accuracy of 85.1 percentages against MLB classifier. The methods like InfoGain, Filter, SU and Relief which do not score the maximum accuracy for any classifies, but perform relatively better by scoring an average accuracy of between 83.2 to 83.5 percentages. But GainRatio, OneR and Chi-squared methods show a very poor performance in terms of classification accuracy,

which score between 81 to 82 percentage only. Hence, from the statistics obtained from our work, we observe that SVM ranking method outperforms other well known ranking methods and also the reduced data sets yields encouraging results in terms of classification accuracy.

After determining the performance of ranking methods with various classification algorithms, the processing time for the selected ranked subset evaluated and values tabulated in Table 5. This table gives the average running time of each ranking methods with respect to the classifiers. From these measurements, again SVM method is proved to be the best ranking method and it performs better than any other ranking methods by taking only an average of 0.3 seconds to build the model. This method performs extremely well against SMO and MLB classifiers. The other ranking techniques like OneR, Filter, Chi-squared InfoGain and GainRatio are also done well and are close second algorithm to SVM, which takes an average running time of between 0.9 to 1.0 seconds.



Fig. 5: Average running time of ranking methods



Fig. 6: Average running time of Classifiers

Table 5: Average running time of ranking methods for all the datasets

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	MLP	Kstar	Average
Relief	0.01	0.03	0.78	0.09	0.12	6.65	7.16	0.00	1.9
GainRatio	0.00	0.02	0.48	0.08	0.08	0.06	7.50	0.00	1.0
InfoGain	0.00	0.02	0.51	0.07	0.10	0.09	7.57	0.00	1.0
OneR	0.00	0.01	0.51	0.07	0.10	0.06	6.41	0.00	0.9
SU	0.00	0.02	0.67	0.08	0.10	0.10	7.60	0.00	1.1
Chi-squared	0.00	0.01	0.52	0.08	0.08	0.07	7.35	0.00	1.0
SVM	0.00	0.02	0.22	0.07	0.06	0.08	2.04	0.00	0.3
Filter	0.00	0.01	0.42	0.07	0.08	0.08	7.53	0.00	1.0
Average Time	0.00	0.02	0.51	0.08	0.09	0.90	6.65	0.00	1.03

Table 6: Average performance of ranking methods

Ranking Method	Classification accuracy (%)	Processing Speed (Sec)
SVM	84.9	0.3
InfoGain	83.5	1.0
Filter	83.5	1.0
SU	83.3	1.1
Relief	83.2	1.9
OneR	82.0	0.9
Chi-squared	82.0	1.0
GainRatio	81.6	1.0

This is because; these methods do not perform up to the level of SVM against SMO and MLP classifiers. It is also observed that all algorithms except Relief is perform the same way when compared with processing time with reduced number of features. But the Relief method takes much time of 1.9seconds than the other methods, since this method uses statistical method and avoid heuristic search. From classifiers point of view, the Multilayer perceptron classifier takes more time to build the model when compared with other classification algorithms. The Naïve bayes classifier and Kstar learn very rapidly for the given dataset.

CONCLUSION

It is difficult to arrive at a conclusion with various ranking methods applied on ten datasets and performing classification against eight classification algorithms. Therefore, for this purpose, we computed average performance for rankings from the various individual ranks by summing the ranks obtained across all data sets. From the experimental results following observations can be made.

- It is evident that the application of rank algorithms on the data sets and then building models with the resultant reduced data sets yield faster and more accurate models than models built without ranking.
- The SVM ranking method takes very minimum processing time of 0.3 seconds to process the data and provide the highest classification accuracy of 84.9 percentages than other ranking methods.
- SVM method performs exceptionally well against multilayer perceptron and SMO classifiers.

- The ranking methods like IG, Filter, SU and Relief perform well in terms of classification accuracy with a moderate processing time between 0.9 to 1.1 seconds.
- The performance of ranking methods like GainRatio, OneR, Chi-Squared is poorer than other selected ranking methods in terms of classification accuracy.
- The Relief method provides better classification accuracy of 83.2 percentages with highest running time of 1.9 seconds.

REFERENCES

1. Feature selection for knowledge discovery and data mining by Huanliu and Hiroshi Motoda.
2. Dash and Liu, 1997. "Feature selection for classification Intelligent Data Analysis",.
3. Langley, 1994. "Selection of relevant features in machine learning",.
4. Luis Carlos Molina, Lluís Belanche and Angela Nebot, in: IEEE 2002. International Conference on Feature Selection Algorithms: A Survey and Experimental Evaluation, pp: 306-313.
5. Chi, J., 1993. "Entropy based feature evaluation and selection technique", Proceedings of 4th Australian Conf. on Neural Networks (ACNN'93), pp: 181-196.
6. Lopez de Mantaras R., 1991. "A Distance-Based Attribute Selecting Measure for Decision Tree Induction", Machine Learning, 6: 81-92.
7. Setion R. and H. Liu, 1996. "Improving back propagation learning with feature selection". Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks and Complex Problem-Solving Technologies, 6: 129-139.
8. Shridhar, D.V., B. Bartlett and E.R.C. Seagrave, 1998. "Information theoretic subset selection." Computers in Chemical Engineering, 22: 613-626.
9. Kira, Kenji and Rendell, Larry, 1992. The Feature Selection Problem: Traditional Methods and a New Algorithm. AAAI- 92 Proceedings.
10. Lijuan, Zhang and Zhoujun Li, 2009. An optimization of ReliefF for classification in large datasets, Data and Knowledge Engineering.
11. Jasmina Novaković, Perica Strbac and Dusan Bulatović, 1998. Toward optimal feature selection using ranking methods and classification algorithms, Yugoslav Journal of Operations Research, 1: 119-135.
12. Han, J. and M. Kamber, 2001. "Data Mining Concepts and Techniques", Morgan Kaufmann.
13. Hall, A., M.A. and L.A. Smith, "Practical feature subset selection for machine learning", Proceedings of the 21st Australian Computer Science Conference.
14. Holte, R.C., 1995. "Very simple classification rules perform well on most commonly used datasets", The twelfth International conference on Machine Learning.
15. Ian H. Witten and Eibe Frank Data, 2005. Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2nd edition.
16. Jasmina Novaković, Perica Strbac and Dusan Bulatović, 2003. Toward optimal feature selection using ranking methods and classification algorithms, Yugoslav Journal of Operations Research.
17. Joachims, T., "Optimizing Search Engines using Click through Data", Proceedings of the ACM Conference on Knowledge Discovery and Data Mining.
18. Mladenic, D. and M. Grobelnik, 1999. Feature selection for unbalanced class distribution and Naïve Bayes. ICML1.
19. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann and Ian H. Witten, 2009. "The WEKA Data Mining Software: An Update", SIGKDD Explorations, 11(1): 10-18.
20. UCI Repository: <http://archive.ics.uci.edu/ml>.

ANEXURE I:

Table 6: Classification accuracy on selected features for Diabetes dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	75.4	0	74.3	0.01	76.4	0.25	74.1	0.04	73.0	0.02	73.4	0.07	74.7	1.05	69.0	0
GainRatio	75.5	0	74.9	0.01	76.2	0.15	75.9	0.04	72.4	0.02	72.0	0.07	76.3	0.97	71.4	0
InfoGain	75.4	0	74.3	0.01	76.0	0.04	75.1	0.07	72.1	0.02	72.0	0.08	77.2	0.96	71.6	0
OneR	75.5	0	74.9	0.01	76.2	0.03	76.2	0.04	72.4	0.02	72.6	0.07	76.0	0.95	71.4	0
SU	75.4	0	74.3	0.01	76.0	0.04	75.1	0.06	72.1	0.02	72.0	0.07	77.2	0.96	71.6	0
Chi-squared	75.4	0	74.3	0.01	76.0	0.04	74.9	0.03	71.6	0.02	71.2	0.07	76.7	0.96	71.6	0
SVM	77.2	0	74.9	0.01	76.8	0.04	74.2	0.04	72.7	0.02	72.4	0.07	75.1	1.09	71.9	0
Filter	75.4	0	74.3	0.01	76.0	0.03	75.1	0.06	72.1	0.02	72.0	0.07	77.2	1.07	71.6	0
Average	75.7	0.0	74.5	0.0	76.2	0.1	75.1	0.0	72.3	0.0	72.2	0.1	76.3	1.0	71.3	0.0

Table 7: Classification accuracy on selected features for segment-challenge dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	73.3	0.01	94.6	0.06	83.1	1.33	93.8	0.49	87.0	0.21	96.2	0.17	95.6	8.37	96.9	0
GainRatio	66.4	0.01	89.2	0.17	77.4	1.62	86.6	0.53	82.8	0.19	90.6	0.22	86.3	8.49	92.1	0
InfoGain	76.9	0.01	94.8	0.06	89.6	1.37	93.9	0.37	87.0	0.21	96.2	0.18	85.3	8.28	97.1	0
OneR	75.0	0.01	94.9	0.06	87.6	0.54	93.6	0.43	87.0	0.22	96.4	0.17	95.5	8.35	97.0	0
SU	76.9	0.01	94.9	0.05	89.6	1.37	93.2	0.4	87.0	0.23	96.8	0.18	95.5	8.31	97.1	0
Chi-squared	66.4	0.01	89.2	0.06	77.6	0.58	88.0	0.51	95.6	0.2	82.8	0.22	88.9	8.33	95.1	0
SVM	82.0	0.01	94.6	0.05	90.7	1.37	93.4	0.33	88.2	0.21	96.7	0.17	96.0	8.36	95.1	0
Filter	76.9	0.01	94.8	0.05	89.6	0.54	93.9	0.37	87.0	0.21	96.2	0.18	95.3	8.37	95.7	0
Average	74.2	0.0	93.4	0.1	85.7	1.1	92.1	0.4	87.7	0.2	94.0	0.2	92.3	8.4	95.8	0.0

Table 8: Classification accuracy on selected features for soybean dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	89.5	0	88.6	0.01	92.8	5.59	87.8	0.07	80.1	0.32	89.0	0.19	92.1	48.22	88.3	0
GainRatio	85.8	0	85.2	0.01	86.2	2.79	84.9	0.05	82.7	0.3	87.4	0.14	87.4	34.22	86.1	0
InfoGain	89.9	0	88.3	0.01	93.0	3.45	88.7	0.07	80.1	0.47	86.8	0.35	93.3	52.93	88.9	0
OneR	83.6	0	85.4	0.01	87.1	4.22	84.8	0.06	83.9	0.5	86.5	0.18	87.3	41.14	86.4	0
SU	89.8	0	90.3	0.01	93.4	4.67	89.8	0.07	82.4	0.31	88.3	0.31	93.6	52.87	90.5	0
Chi-squared	89.2	0	89.8	0.01	93.9	4.3	89.6	0.07	81.3	0.31	91.4	0.19	93.7	50.6	90.0	0
Filter	89.9	0	88.3	0.01	93.0	3.35	89.9	0.08	80.1	0.32	86.8	0.2	93.3	52.31	88.9	0
Average	88.2	0.0	88.0	0.0	91.3	4.1	87.9	0.1	81.5	4.9	88.0	0.2	91.5	47.5	88.4	0.0

Table 9: Classification accuracy on selected features for vote dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	93.56	0	96.09	0	95.6	0.03	95.9	0.01	95.48	0.03	94.5	0.03	93.8	1.03	94.9	0
GainRatio	91.3	0	95.2	0	95.6	0.03	95.6	0	95.6	0.03	94.5	0.03	95.2	1.04	92.9	0
InfoGain	91.3	0	95.2	0	95.6	0.04	95.6	0	95.6	0.03	94.5	0.03	95.2	1.04	92.9	0
OneR	90.6	0	94.7	0	95.6	0.03	95.4	0.01	95.4	0.04	95.2	0.03	94.0	1.03	92.9	0
SU	91.3	0	95.2	0	95.6	0.04	95.6	0	95.6	0.03	94.1	0.04	95.2	1.04	92.9	0
Chi-squared	91.3	0	95.2	0	95.6	0.02	95.6	0	95.6	0.03	93.6	0.03	94.0	1.06	92.9	0
SVM	91.5	0	96.3	0	95.9	0.04	96.3	0.01	94.7	0.04	95.9	0.03	94.9	1.18	94.0	0
Filter	91.3	0	95.2	0	95.6	0.04	95.6	0	95.6	0.03	94.3	0.03	95.2	1.1	92.9	0
Average	91.5	0.0	95.4	0.0	95.6	0.03	95.7	0.0	95.4	0.0	94.6	0.0	94.7	1.1	93.3	0.0

Table 10: Classification accuracy on selected features for ionosphere dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	86.3	0	92.9	0.02	87.7	0.05	90.9	0.05	89.5	0.08	93.2	0.06	90.9	2.09	84.6	0
GainRatio	87.5	0.01	90.3	0.02	87.7	0.05	91.7	0.04	89.5	0.08	93.4	0.06	92.6	20.5	85.2	0
InfoGain	88.0	0	92.0	0.03	87.7	0.05	90.9	0.06	89.5	0.09	93.4	0.06	94.0	2.11	86.6	0
OneR	88.0	0.01	92.0	0.02	87.7	0.06	90.9	0.06	89.5	0.08	93.4	0.07	91.5	2.12	84.6	0
SU	88.0	0	92.0	0.02	87.7	0.04	90.9	0.04	89.5	0.09	93.4	0.07	92.0	2.01	86.3	0
Chi-squared	88.0	0	92.0	0.02	87.7	0.05	90.9	0.06	89.5	0.09	93.4	0.06	94.6	1.99	86.6	0
SVM	88.0	0	92.0	0.03	87.7	0.05	90.9	0.04	89.5	0.08	93.4	0.18	91.1	2.14	87.2	0
Filter	88.0	0	92.0	0.03	87.7	0.05	90.9	0.06	89.5	0.08	93.4	0.07	94.0	2.1	86.6	0
Average	87.7	0.0	91.9	0.0	87.7	0.1	91.0	0.1	89.5	0.1	93.4	0.1	92.6	4.4	86.0	0.0

Table 11: Classification accuracy on selected features for Dermatology dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	96.1	0.01	90.44	0.02	94.80	0.18	88.25	0.06	85.79	0.13	92.89	0.1	94.80	7.35	93.44	0
GainRatio	85.80	0.01	81.69	0	85.52	0.08	81.69	0.06	78.14	0.03	80.87	0.03	84.43	6.72	82.24	0
InfoGain	91.80	0	87.98	0.03	92.89	0.09	89.90	0.06	86.34	0.13	90.43	0.11	92.62	7.24	90.71	0
OneR	91.80	0	87.98	0	92.90	0.09	90.16	0.02	86.06	0.03	91.26	0.01	92.90	7.47	90.71	0
SU	91.80	0.01	87.97	0.02	92.90	0.26	89.34	0.05	86.34	0.14	91.53	0.11	92.07	7.60	90.71	0
Chi-squared	87.98	0	81.14	0.01	84.15	0.09	82.78	0.01	78.14	0.04	84.42	0.02	86.88	7.40	87.15	0
Filter	91.80	0	87.98	0.02	92.89	0.09	89.89	0.05	86.34	0.12	90.43	0.11	92.62	7.33	90.71	0
Average	91.0	0.0	86.5	0.0	90.9	0.1	87.4	0.0	83.9	0.1	88.8	0.1	90.9	7.3	89.4	0.0

Table 12: Classification accuracy on selected features for Lung cancer dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	75	0	59.37	0.02	62.5	0.06	53.13	0.01	62.5	0.08	65.63	65.63	56.25	1.76	59.37	0
GainRatio	71.87	0	62.5	0	68.7	0.01	53.13	0	65.63	0.01	65.63	0.01	65.62	1.63	59.37	0
InfoGain	68.75	0	65.62	0	59.37	0.02	65.62	0	65.62	0	62.5	0.01	59.37	1.63	56.25	0
OneR	62.5	0	50	0	46.87	0.02	56.25	0	59.37	0.01	43.75	0.01	53.12	1.67	46.87	0
SU	68.75	0	62.5	0	62.5	0.01	50	0	65.62	0	62.5	0.01	53.12	1.62	53.12	0
Chi-squared	65.62	0	62.5	0	56.25	0.02	53.12	0	65.62	0.01	56.25	0.01	59.37	1.71	56.25	0
Filter	68.75	0	65.62	0	59.37	0.02	65.62	0	65.62	0.01	62.5	0.01	59.37	1.63	56.25	0
Average	79.6	0.0	75.6	0.0	79.5	0.1	76.5	0.0	73.4	0.1	77.7	0.1	79.5	6.4	78.2	0.0

Table 13: Classification accuracy on selected features for Wine dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	96.0	0	94.9	0.02	96.6	0.03	94.4	0.03	88.8	0.07	96.06	0.05	98.9	0.31	96.6	0
GainRatio	96.06	0	94.9	0	96.6	0.02	94.9	0	88.8	0.06	96.6	0.01	98.9	0.16	96.6	0
InfoGain	96.06	0	94.9	0	96.6	0.01	92.13	0.01	88.8	0.01	96.6	0.01	98.9	0.16	96.6	0
OneR	95.5	0	94.9	0	97.7	0.02	93.2	0.01	88.7	0	97.2	0.01	96.06	0.17	96.0	0
SU	96.0	0	94.9	0	96.6	0.02	93.2	0.01	88.76	0.01	96.6	0.02	98.3	0.21	96.6	0
Chi-squared	96.0	0	94.9	0	96.6	0.01	93.3	0	88.7	0.01	96.6	0.01	98.9	0.17	96.6	0
SVM	96.6	0	94.3	0	96.0	0.01	96.0	0.01	89.3	0.01	97.2	0.01	97.1	0.22	97.1	0
Filter	96.0	0	94.4	0	96.6	0.02	94.4	0	88.8	0.01	96.06	0.01	98.87	0.16	96.6	0
Average	96.0	0.0	94.8	0.0	96.7	0.0	93.9	0.0	88.8	0.0	96.6	0.0	98.2	0.2	96.6	0.0

Table 14: Classification accuracy on selected features for Hepatitis dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	82.9	0	79.35	0.01	82.6	0.05	81.9	0.02	81.3	0.06	81.9	0.08	78.7	0.23	83.22	0
GainRatio	83.22	0	83.22	0.01	83.9	0.01	80	0.03	74.8	0.01	83.22	0.01	80.6	0.13	81.3	0
InfoGain	85.8	0	83.22	0	83.9	0.01	81.3	0.01	74.8	0.01	81.3	0.01	81.3	0.13	84.5	0
OneR	82.58	0	80.64	0	83.2	0.01	78.7	0	76.12	0.01	82.6	0.01	81.9	0.13	85.16	0
SU	85.8	0	83.22	0	83.9	0.01	80	0	74.83	0.01	82.6	0.01	80.6	0.13	84.5	0
Chi-squared	85.8	0	83.22	0	83.9	0.01	78.7	0.01	74.8	0.01	81.3	0.01	76.8	0.14	84.5	0
SVM	87.0	0	80	0	83.9	0.01	78	0.01	70.3	0.01	80	0.02	83.8	0.13	80.6	0
Filter	85.8	0	83.22	0	83.9	0.01	81.3	0.01	74.8	0.01	81.3	0.01	81.3	0.13	84.5	0
Avg.	84.9	0.0	82.0	0.0	83.7	0.0	80.0	0.0	75.2	0.0	81.8	0.0	80.6	0.1	83.5	0.0

Table 15: Classification accuracy on selected features for Vehicle dataset

Ranking Method	NB		J48		SMO		JRIP		DT		Rd.Frt		Mul.pr		Kstar	
	%	S	%	S	%	S	%	S	%	S	%	S	%	S	%	S
Relief	39.9	0.03	71.15	0.08	64.0	0.21	64.0	0.16	63.47	0.17	70.7	0.15	72.3	1.22	69.38	0
GainRatio	45.15	0	67.13	0.02	63.7	0.04	63.59	0.05	64.77	0.03	70.9	0.05	69.7	1.1	69.3	0
InfoGain	43.61	0	68.43	0.02	67.0	0.06	67.25	0.06	64.18	0.03	70.56	0.05	72.34	1.17	72.57	0
OneR	43.6	0	68.32	0.02	67.0	0.06	68.2	0.05	64.18	0.04	73.28	0.05	73.16	1.1	72.57	0
SU	43.4	0.02	69.50	0.09	66.66	0.2	68.55	0.17	64.89	0.16	73.75	0.15	70.09	1.2	72.45	0
Chi-squared	43.61	0	68.32	0.02	67.02	0.07	68.32	0.07	64.18	0.04	72.10	0.05	72.45	1.12	72.57	0
SVM	48.34	0	68.91	0.02	71.21	0.05	65.72	0.07	62.41	0.03	73.16	0.06	77.06	1.15	70.21	0
Filter	43.61	0	68.43	0.01	66.9	0.06	67.25	0.07	64.18	0.03	70.56	0.06	72.3	1.13	72.57	0
Average	43.9	0.0	68.8	0.0	66.7	0.1	66.6	0.1	64.0	0.1	71.9	0.1	72.4	1.1	71.5	0.0