

Instituto de Ciências Matemáticas e de Computação

ISSN - 0103-2569

**The *MLC++* Wrapper for Feature Subset Selection
Using Decision Tree, Production Rule,
Instance Based and Statistical Inducers:
Some Experimental Results**

**José Augusto Baranauskas
Maria Carolina Monard/ILTC**

Nº 87

RELATÓRIOS TÉCNICOS DO ICMC

São Carlos
Julho/1999

The $M\mathcal{L}\mathcal{C}++$ Wrapper for Feature Subset Selection Using Decision Tree, Production Rule, Instance Based and Statistical Inducers: Some Experimental Results*

José Augusto Baranauskas
Maria Carolina Monard/ILTC

University of São Paulo
Institute of Mathematics and Computer Sciences
Department of Computer Science and Statistics
Laboratory of Computational Intelligence
P.O. Box 668, 13560-970 - São Carlos, SP, Brazil
e-mail: {jaugusto, mcmonard}@icmc.sc.usp.br

Abstract

A common problem that learning algorithms (inducers) should face is where to focus their attention. In the feature selection problem, a inducer is faced with the problem of selecting relevant features while ignoring irrelevant ones. Methods for feature selection that have been developed can be grouped into three classes: those that *embed* the selection within the basic induction algorithm, those that use feature selection to *filter* features during a pre-processing step ignoring the induction algorithm, and those that treat feature selection as a *wrapper* around the induction process, using the induction algorithm as a black box. This work describes in details several results from experiments using the $M\mathcal{L}\mathcal{C}++$ wrapper approach with five inducers in several datasets, its computational cost (time to run), accuracy gain and some basic statistical tests comparing standard inducers (no feature selection) with inducers using wrapper selected features.

Keywords: Feature Selection; Wrapper; Machine Learning; Data Mining.

July 1999

*Work partially supported by National Research Councils — CNPq and FINEP, Brazil

Contents

1	Introduction	1
2	The Feature Subset Selection Problem	1
2.1	Feature Selection as Heuristic Search	2
2.2	Approaches for Feature Subset Selection	2
2.2.1	Embedded	3
2.2.2	Filters	3
2.2.3	Wrappers	4
3	Inducers	5
3.1	Data Format	6
3.2	C4.5	6
3.3	CN2	6
3.4	Naive Bayes	7
3.5	Instance Based	7
3.6	Table Majority	7
4	Datasets	7
4.1	General Description	8
4.2	Datasets Summary	9
5	Experimental Setup	10
6	Experimental Results	11
6.1	Summary Tables Description	11
6.2	Bupa	12
6.3	Pima	13
6.4	Breast Cancer	14
6.5	Hungaria	15
6.6	Crx	17
6.7	Letter	18
6.8	Hepatitis	20
6.9	Anneal	21
6.10	Sonar	23
6.11	Genetics	26
6.12	DNA	28
7	Results Comparison	32
7.1	Number of Selected Features	32
7.2	Time for Selecting Features	33
7.3	Comparing No FSS, Forward and Backward Wrapper FSS	34
8	Conclusions	39

List of Figures

1	Features Subsets State Space	3
2	Feature Filter Model	4

3	Wrapper Approach	5
4	Datasets Dimensionality	10
5	<i>C4.5</i> Absolute Difference in Standard Deviations of Accuracies	36
6	<i>CN2</i> Absolute Difference in Standard Deviations of Accuracies	37
7	IB Absolute Difference in Standard Deviations of Accuracies	37
8	NB Absolute Difference in Standard Deviations of Accuracies	38
9	TM Absolute Difference in Standard Deviations of Accuracies	38

List of Tables

1	Feature-Value or Spreadsheet Format	6
2	Datasets Summary Descriptions	10
3	Bupa Feature Description	12
4	Bupa Time for Selecting Features	12
5	Bupa Wrapper Selected Features	12
6	Bupa Accuracies	13
7	Pima Feature Description	13
8	Pima Time for Selecting Features	13
9	Pima Wrapper Selected Features	14
10	Pima Accuracies	14
11	Breast Cancer Feature Description	14
12	Breast Cancer Time for Selecting Features	15
13	Breast Cancer Wrapper Selected Features	15
14	Breast Cancer Accuracies	15
15	Hungaria Feature Description	16
16	Hungaria Time for Selecting Features	16
17	Hungaria Wrapper Selected Features	16
18	Hungaria Accuracies	17
19	Crx Feature Description	17
20	Crx Time for Selecting Features	17
21	Crx Wrapper Selected Features	18
22	Crx Accuracies	18
23	Letter Feature Description	18
24	Letter Wrapper Selected Features	19
25	Letter Wrapper Selected Features	19
26	Letter Accuracies	19
27	Hepatitis Feature Description	20
28	Hepatitis Time for Selecting Features	20
29	Hepatitis Wrapper Selected Features	20
30	Hepatitis Accuracies	21
31	Anneal Feature Description	22
32	Anneal Time for Selecting Features	22
33	Anneal Wrapper Selected Features	22
34	Anneal Accuracies	23
35	Sonar Feature Description	24
36	Sonar Time for Selecting Features	24
37	Sonar Wrapper Selected Features	25
38	Sonar Accuracies	25
39	Genetics Feature Description	26

40	Genetics Time for Selecting Features	27
41	Genetics Wrapper Selected Features	28
42	Genetics Accuracies	28
43	DNA Feature Description	28
44	DNA Time for Selecting Features	29
45	DNA Wrapper Selected Features	31
46	DNA Accuracies	31
47	Number of Selected Features	32
48	Proportion of Selected Features	32
49	Time (in seconds) for Selecting Features	33
50	Time for Running Ten-Fold Cross-Validation and Ten-Fold Stratified Cross-Validation Using all Features	33
51	Absolute Difference in Standard Deviations of Accuracies	35
52	Improved Accuracies at Significance Level	36

1 Introduction

Data Mining — DM — is an interdisciplinary field bringing together techniques from Machine Learning, statistics, pattern recognition, databases and visualization to address the issue of extracting high-level knowledge from low-level data in large databases.

When using Machine Learning — ML — techniques for DM, where the number of records (instances) is very large, usually several representative samples from the database are taken and presented to a ML algorithm. Afterwards the knowledge extracted from those samples by ML algorithms is combined in some way (Fayyad et al., 1996a).

One important issue to be considered when using ML in DM is database records dimensionality reduction by reducing the number of records attributes (*i.e.* deleting columns in tables in the database literature or features in the Machine Learning literature). The data subset resulting from these deletions maintains the same number of instances but only a subset of features with predictive performance comparable to the full set of features. This is known as the Feature Subset Selection — FSS — problem where one of the central issues is the selection of relevant features and/or the elimination of irrelevant ones.

There are several reasons for doing Feature Subset Selection. One of them is that they improve accuracy since many ML algorithms degrade in performance when are given too many features. Another reason is that FSS improves comprehensibility, which is the ability for humans to understand the data and classification rules induced by symbolic ML algorithms. Finally, FSS can reduce measurement cost since in some domains measuring features may be expensive.

There are, basically, three main approaches that have been pursued in ML for FSS: embedded, filter and wrapper. In the first case, the feature selection process is embedded within the basic induction algorithm. The filter approach uses just the data whereas the wrapper approach uses the induction algorithm as a black box.

As induction algorithms differ considerably in their emphasis on focusing on relevant features, it is considered that the wrapper method should provide a better estimate of accuracy than filter methods since wrapper methods use the same induction algorithm that will be used on the feature subset selected, *i.e.* they run a search using the inducer itself to determine which attributes in the database are useful for learning. On the other hand, the computational cost of wrapper methods can be very high since they have to call the induction algorithm for each feature set considered.

In this work we concentrate on the wrapper approach. Specifically, we use the wrapper method implemented in $\mathcal{MLC}++$ where the search is conducted in the space of subsets with add/delete operators using best-first search, and the heuristic for the search is the estimated prediction accuracy using cross-validation (Kohavi & Sommerfield, 1995).

This work is organized as follows. Section 2 describes the feature selection problem and Section 3 briefly describes each of the classification algorithms (inducers) used as black box to the wrapper approach for FSS. Section 4 gives some background of the datasets used in the experiments and Section 5 shows the experimental setup used. Section 6 describes the results and Section 7 reports analysis and comparison of results. Finally, conclusions are given in Section 8.

2 The Feature Subset Selection Problem

Supervised learning is the process of automatically creating a classification model from a set of instances (records or examples) called the *training set* which belong to a set of classes. There are two aspects to be considered in this process: which features to use in describing the concept and how to combine those features. Once a model is created, it can be used to automatically predict the class of other unclassified records.

In other words, in supervised learning, a set of n training examples is given to an inducer. Each example \mathbf{X} is an element of the set:

$$F_1 \times F_2 \times \dots \times F_m$$

where F_j is the domain of the j th feature. Training examples are tuples (\mathbf{X}, Y) where Y is the label, output or class. The Y values are typically drawn from a discrete set of classes $\{1, \dots, K\}$ in the case of *classification* or from the real values in the case of *regression*. In this work we will refer to classification. Given a set of training examples, the learning algorithm (*inducer*) outputs a *classifier* such that, given a new instance, it accurately predicts the label Y .

One of the central problems in supervised learning is the selection of useful features. Although most learning methods attempt to either select features or assign them degree of importance, both theoretical analysis and experimental studies indicate that many algorithms scale poorly to domains with large numbers of irrelevant features. For example, the number of training cases needed for simple nearest neighbor to reach a given level of accuracy appears to grow exponentially with the number of irrelevant features, independent of the target concept. Even methods for inducing univariate decision trees, which explicitly select some attributes in favor of others, exhibit this behavior for some target concepts. And some techniques, like the Naive Bayes classifier, are robust with respect to irrelevant features but can be very sensitive to domains with correlated features, even if the features are relevant. This can be explained by the assumption of this sort of techniques related to independence among features. This suggests the need for additional methods to select a useful subset of features when many are available (Langley, 1996).

2.1 Feature Selection as Heuristic Search

FSS can be considered as a state space search, with each node (state) in the search space specifying a subset of possible features — see Figure 1 where four features are considered and a dark circle indicates the presence of that feature. It should be observed that there is a natural partial ordering on this space, with each of a state's children including one more feature than its parent.

Related to the value (or *goodness*) of a node, several strategies can be used, such as the estimated prediction accuracy. The operators are commonly add/delete feature.

Having a partial ordering on the FSS space, there are two immediate approaches that determine the starting point as well as the direction of the search:

1. *forward selection*
2. *backward selection*

In forward selection the starting point is an empty set of features and features are added successively. In backward selection the starting point contains all features and features are removed successively. Several other variations can be used such as selecting any other starting point and move outward from it.

Other important issue is the organization of the search. An exhaustive search of the space is impractical when many features are present since there exist 2^m possible subsets (states) of m attributes. More realistic methods, such as hill-climbing or best-first search may be used to traverse the space.

2.2 Approaches for Feature Subset Selection

Approaches for feature selection that have been developed can be grouped into three classes: those that *embed* the selection within the basic induction algorithm, those that use feature selection to *filter* features during a pre-processing step ignoring the induction algorithm, and those that treat feature selection as a *wrapper* around the induction process, using the induction algorithm as a black box (Blum & Langley, 1997). It follows a brief description of these approaches.

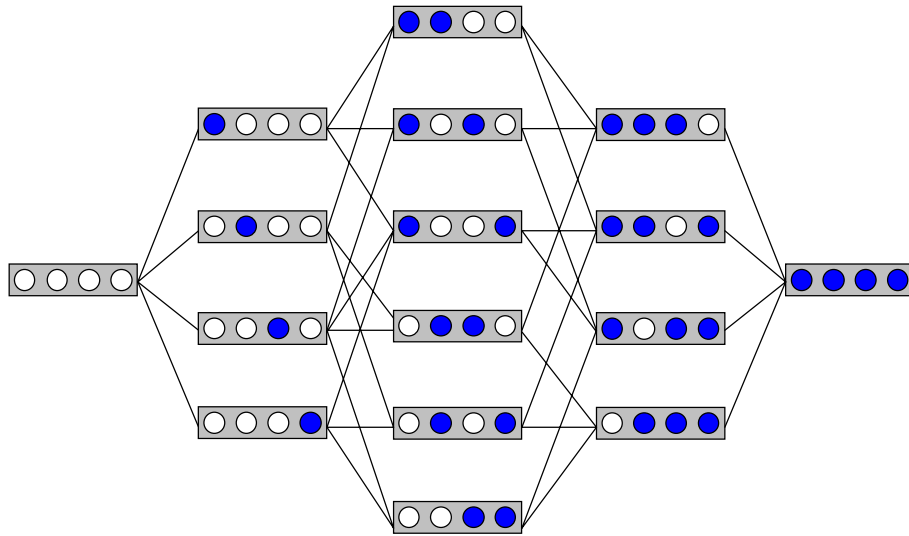


Figure 1: Features Subsets State Space

2.2.1 Embedded

Some inducers are capable of doing their own feature selection, in a dynamic way, while searching for a concept description. In fact, FSS is an integral part of these inducers.

For example, recursive partitioning methods for induction, such as decision trees, carry out a greedy search through the space of decision trees. At each step they use an evaluation function to select the feature that has the best ability to discriminate among the classes. They partition the training set based on this feature and repeat the process on each subset, extending the tree downward until no further discrimination is possible. This method is used by the $\mathcal{C}4.5$ inducer (Quinlan, 1988).

Separate-and-conquer methods for learning decision lists embed FSS in a similar way. These techniques use an evaluation function to select a feature that helps distinguish a class C from others, then add the resulting test to a single conjunctive rule for C . They repeat this process until the rule excludes all members of other classes, then remove the members of C that the rule covers and repeat the process on the remaining training cases. This method is used by the $\mathcal{CN}2$ inducer (Clark & Boswell, 1991).

Partition and separate-and-conquer methods explicitly select features that appear more relevant for inclusion in a branch or a rule.

2.2.2 Filters

In the FSS filter model, the features are filtered independent of the induction algorithm. In this model the FSS is done as a preprocessing step, totally ignoring the effects of the selected features subset on the performance of the induction algorithm — see Figure 2.

For example, a simple decision tree algorithm can be used as a FSS filter to select features in large feature space for other inducers that take more time to search their solution space. The set of features selected by the tree are the output of the filter FSS process and the tree itself is discarded. The remaining unused features are then deleted from the training set, reducing its dimension, and this training set can be used by any other inducer. Still, features that are good for decision trees are not necessarily useful for other family of algorithms that may have an entirely different inductive bias.

Besides decision trees, there are other approaches for filtering using this model, and several algorithms have been proposed, such as FOCUS (Almuallim & Dietterich, 1991) and Relief (Kira & Rendell, 1998; Kira

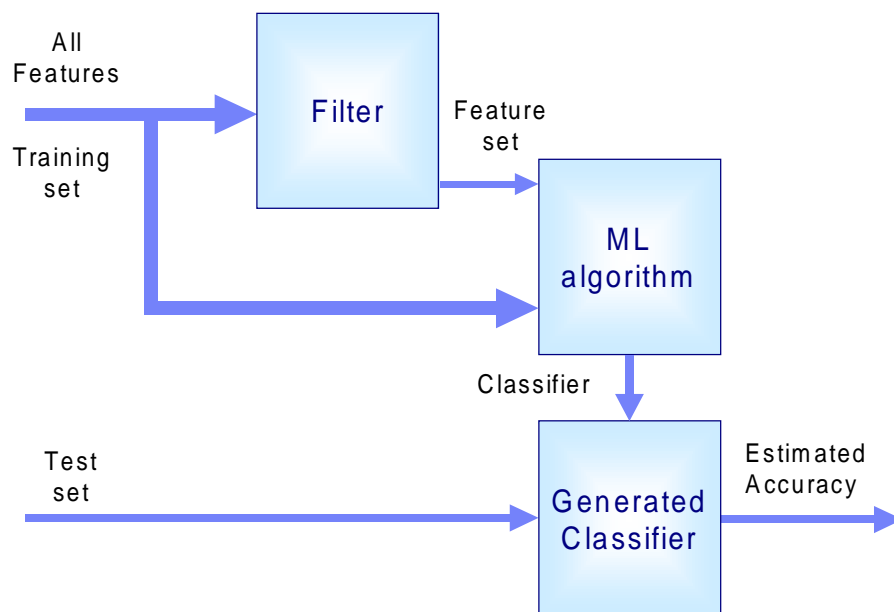


Figure 2: Feature Filter Model

& Rendell, 1992).

The main disadvantage of the filter approach is that it totally ignores the effects of the selected feature subset on the performance of the induction algorithm. It is claimed that the filter model should be replaced with the wrapper model that takes the induction algorithm itself into account (Kohavi, 1997).

2.2.3 Wrappers

In contrast with filter methods, wrapper methods generate a feature subset as candidate, run the induction algorithm with only those features on the training set, and use the accuracy of the resulting description to evaluate the feature subset. This process is repeated for each feature subset candidate until the criterion for halting the search is satisfied.

The argument in favor of the wrapper approach is that the same inducer that will use the feature subset should provide a better estimate accuracy for itself than the accuracy provided by another method that may have a different inductive bias. On the other hand, the wrapper approach can be very expensive since the induction algorithm should be called for each feature set considered.

In this work we use the FSS wrapper method implemented in $MCC++$ which is a library of $C++$ classes and tools developed at Stanford University (Kohavi et al., 1994). $MCC++$ provides general machine learning algorithms as well as a wide variety of tools that can be used by end users. Some algorithms support visual output of the classifiers and may generate output for Silicon Graphic's MineSetTM product.

The general idea behind the wrapper approach is shown in Figure 3 (Kohavi, 1997). The feature subset selection algorithm exists as a wrapper around the induction algorithm and it is responsible for conducting the search for a good subset. The goodness of a candidate subset is evaluated using the induction algorithm itself as a black box. The goal of the search is to find the state with the highest evaluation, using an heuristic function to guide it. Ten-fold cross-validation to estimate prediction accuracy is the default method used by the $MCC++$ wrapper around the induction algorithm to evaluate states¹.

¹Five-fold is reported in (Kohavi, 1997).

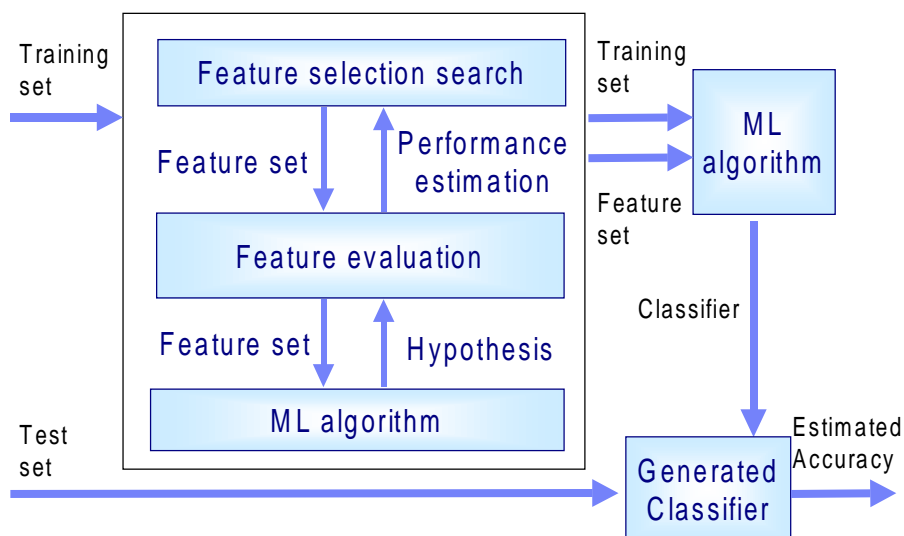


Figure 3: Wrapper Approach

The search of the $MCC++$ FSS wrapper is conducted in the space of subsets with add/delete operators using as search method hill-climbing or best-first, and forward or backward as the direction of the search. The heuristic for the search is the estimated prediction accuracy using cross-validation (Kohavi & Sommerfield, 1995). In this work the search method used is best-first, which is a more robust method than hill climbing, using forward and backward selection as the direction of the search and five different induction algorithms as a basis for comparison.

3 Inducers

The following five inducers found in the $MCC++$ library (Kohavi et al., 1996) have been used as black box inducers to the $MCC++$ wrapper algorithm for FSS:

1. $C4.5$ which learn propositional decision trees;
2. $CN2$ which learn propositional production rules;
3. IB, instance-based learning;
4. Naive Bayes — NB —, statistics learning and
5. Table Majority — TM —, a simple instance-based learning.

All of them are well known in the ML community and represent three different learning approaches. NB is a very simple statistical inducer, IB and TM are lazy inducers and $C4.5$ and $CN2$ are typical eager inducers.

Purely lazy algorithms store all their training data (by default there is not embedded FSS in this kind of algorithms) and reply to information request by combining their stored training data, discarding the constructed answer and any intermediat result. In contrast, eager learning algorithms greedily compile the training data into an intentional concept description, such as a rule set or decision tree, discarding the data after this process (Aha, 1997).

This lazy/eager distinction exhibits many interesting tradeoffs. For example, while lazy algorithms have lower computational costs than eager algorithms during training, they typically have greater storage requirements and often have higher computational costs when answering requests.

An important point to consider about any inducer is how it handles “unknown” data values. From the five inducers considered in this work, only $\mathcal{C}4.5$ and $\mathcal{CN}2$ are capable of treating unknown attribute values — see Sections 3.2 and 3.3. Naive Bayes skips them while IB and Table Majority deal with unknown values as any other value or, in other words, they do not treat unknown values. The next sections describe the data format used as input to inducers as well as some basic background for each of them.

3.1 Data Format

As explained before, in supervised Machine Learning an inducer is generally presented with a set of training instances. Each instance is described typically by a vector of feature values and a class label denoted by (\mathbf{X}, Y) . This common format is the feature-value (either attribute-value or spreadsheet) format. Table 1 illustrates this organization where a row i refers to the i -th example or instance \mathbf{X}_i and column entries x_{ij} refer to the individual value of the j -th feature f_j of instance i . The column rotated *class* refers to the label or classification of that instance.

f_1	f_2	\dots	f_m	<i>class</i>
x_{11}	x_{12}	\dots	x_{1m}	y_1
x_{21}	x_{22}	\dots	x_{2m}	y_2
\dots	\dots	\dots	\dots	\dots
x_{n1}	x_{n2}	\dots	x_{nm}	y_n

Table 1: Feature-Value or Spreadsheet Format

The datasets file formats that $\mathcal{MLC}++$ recognizes by default are the *data*, *test* and *names* files. The *data* and *test* files contain labelled instances, one per line, of the training and test set respectively. The *names* file defines the scheme that allows parsing these two previous files. It describes the name and domain for each attribute and for the label. Anyway, the accuracy of the classifier output by the inducer is measured on unseen data (the test set). More details can be found in (Kohavi et al., 1994; Felix et al., 1998).

3.2 $\mathcal{C}4.5$

$\mathcal{C}4.5$ (Quinlan, 1988) is a member of a more general Machine Learning inducers family named Top Down Induction of Decision Trees — TDIDTs. $\mathcal{C}4.5$ is a learning algorithm that creates decision trees to represent classification rules. A node in a decision tree represents a test on a particular attribute.

Building a decision tree proceeds as follows (Quinlan, 1986): using the training set, an attribute is chosen to split it according to attribute’s value. For each subset, another attribute is chosen to split each one. This continues as long as each subset contains a mix of instances belonging to different classes. Once a uniform subset — *i.e.* all instances in that subset belong to the same class — has been obtained, a leaf node is created and labelled with the same name of the respective class.

When a new instance should be classified, beginning from the root of the induced tree, $\mathcal{C}4.5$ tests and branches each node with the respective feature until it reaches one leaf. The class prediction of this instance is assigned as the class of that leaf. If no rule is satisfied, the default rule assigns the most common (majority) class to the new example.

3.3 $\mathcal{CN}2$

The $\mathcal{CN}2$ (Clark & Niblett, 1987; Clark & Niblett, 1989; Clark & Boswell, 1991) is a Machine Learning algorithm that induces ‘*if* <complex> *then* <class>’ rules in domains where there might be noise. Each

<complex> is a disjunction of conjunctions.

For unknown feature values, $\mathcal{CN}2$ use the method of simply replacing unknown values with the most commonly occurring value if the feature is nominal. For continuous features, the midvalue of the most commonly occurring sub-range replaces the unknown value.

To classify a new instance using induced unordered rules (default $\mathcal{CN}2$ rule generation), all rules are tried and those which fire are collected. If more than one class is predicted by fired rules, the method used is to tag each rule with the distribution of covered examples among classes and then to sum these distributions to find the most probable class. For instance, consider the three rules:

if	head=square	and	hold=gun	then	class=enemy	covers	[15,1]
if	size=tall	and	flies=no	then	class=friend	covers	[1,10]
if	look=angry			then	class=enemy	covers	[20,0]

Here the two classes are [enemy,friend] and [15,1] denotes that the rule covers 15 training instances of enemy and 1 of friend. Given a new instance of a robot which has square head, carries a gun, tall, non-flying and angry, all three rules are fired. $\mathcal{CN}2$ resolve this clash by summing the covered instances [36,11] and then predicting the most common class in the sum — enemy.

3.4 Naive Bayes

The Naive Bayes inducer (Langley et al., 1992) computes conditional probabilities of the classes given the instance and picks the class with the highest posterior probability. Attributes are assumed to be *independent*, an assumption that is unlikely to be true in real domains, but the algorithm is very robust to violations of this assumption.

The probabilities for nominal features are estimated by count. The probability for zero counts is $1/2n$ for n instances. The probabilities for continuous features are estimated by assuming a normal distribution for each feature and class. Unknown values in the test instance are skipped, which is equivalent to marginalizing over them.

3.5 Instance Based

IB is an instance-based inducer (Aha, 1997; Aha, 1992), which is a lazy learning algorithm. The general idea is to delay compilation of the training set, storing instances. Predictions are generated at classification time according to an implicit Voronoi partition of the instance space.

3.6 Table Majority

Table Majority is one of the simplest lazy inducers. A TM inducer simply passes the training set to the TM classifier for its internal structure. Given an instance, a TM classifier searches for all matching instances in the table. If no matching instances are found, the majority class of the table is returned as the class label.

4 Datasets

Experiments were conducted on several real world domains. Most datasets are from the UCI Irvine Repository (Blake et al., 1998) but DNA and Genetics are from the StatLog Project (Taylor et al., 1994). Besides coming from different domains, the datasets chosen have different dimensionality: the number of instances and the number of features varies over a large range. To assist comparisons, the datasets chosen also have different type of attributes. They involve continuous attributes, either alone or in combination with nominal attributes, as well as unknown values. The next Section gives basic datasets descriptions and Section 4.2 summarizes datasets characteristics.

4.1 General Description

It follows some descriptions about the datasets used in this work. For datasets Anneal, Hepatitis and Hungaria a more detailed or specific description was not available.

Anneal This dataset was donated by David Sterling and Wray Buntine.

Breast-cancer This dataset was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The problem is to predict whether a tissue sample taken from a patient's breast is malignant or benign. Tissue samples consist of visually assessed nuclear features of fine needle aspirates taken from patient's breast. Each sample was assigned a 9-dimensional vector. Each component is in the range 1 to 10, with 1 referring to a normal state and 10 to a most abnormal one. Malignancy is determined by taking a tissue sample from patient's breast and performing a biopsy on it. A benign diagnosis is confirmed by biopsy or by periodic examination, depending on the patient's choice.

Bupa This dataset was contributed by R. S. Forsyth to the UCI repository. The problem is to predict whether or not a male patient has liver disorders based on various blood tests and the amount of alcohol consumption.

CRX This dataset concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data.

DNA The domain of this dataset (Taylor et al., 1994) is drawn from the field of molecular biology. Splice junctions are points on a DNA sequence at which *superfluous* DNA is removed during protein creation. The task is to recognize exon/intron boundaries, referred to as **EI** sites; intron/exon boundaries, referred to as **IE** sites; or neither. The **IE** borders are referred as *acceptors* and the **EI** borders are *donors*. The instances were taken from GenBank 64.1 (genbank.bio.net). The features provide a window of 60 nucleotides, each represented as 3 binary indicator features that represent the value A, C, G, or T, thus giving 180 binary features. The classification is the middle point of the window, thus providing 30 nucleotides at each side of the junction.

Genetics Genetics dataset is an alternative codification for the DNA dataset. The features provide a window of 60 nucleotides, each represented as values A, C, G, or T.

Hepatitis This dataset is for predicting life expectation of patients with hepatitis.

Hungaria This dataset is for diagnosing heart diseases.

Letter The objective is to identify each of a large number of black-and-white rectangular pixel display as one of the 26 capital letters in the English alphabet. The character images were based on 20 different fonts and each letter within these 20 fonts was randomly distorted to produce a file of 20,000 unique stimuli. Each stimulus was converted into 16 primitive numerical attributes (statistical moments and edge counts) which were then scaled to fit into a range of integer values from 0 through 15.

Pima This dataset was donated by V. Sigillito, Applied Physics Laboratory, Johns Hopkins University and it is available at the UCI repository. This dataset is also a subset of a larger database maintained by the National Institute of Diabetes and Digestive and Kidney Diseases.

All patients are females at least 21 years old of Pima Indian heritage living near Phoenix, Arizona, USA. The problem is to predict whether a patient would test positive for diabetes according to World Health Organization (WHO) criteria — *i.e.* if the 2-hour post-load plasma glucose is at least 200 mg/dl at any survey examination or if found during routine medical care — given a number of physiological measurements and medical test results.

Sonar This is the dataset used by Gorman and Sejnowski in their study of the classification of sonar signals using a neural network (Gorman & Sejnowski, 1988). The task is to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock.

The dataset contains 111 patterns obtained by bouncing sonar signals off a metal cylinder at various angles and under various conditions and it also contains 97 patterns obtained from rocks under similar conditions. The transmitted sonar signal is a frequency-modulated chirp, rising in frequency. The dataset contains signals obtained from a variety of different aspect angles, spanning 90 degrees for the cylinder and 180 degrees for the rock.

Each pattern is a set of 60 numbers in the range 0.0 to 1.0. Each number represents the energy within a particular frequency band, integrated over a certain period of time. The integration aperture for higher frequencies occur later in time, since these frequencies are transmitted later during the chirp.

The label associated with each record contains the letter **R** if the object is a rock and **M** if it is a mine (metal cylinder). The numbers in the labels are in increasing order of aspect angle, but they do not encode the angle directly.

4.2 Datasets Summary

Table 2 summarizes the datasets employed in this study. It shows, for each dataset, the number of instances (#Instances), duplicate (appearing more than once) or conflicting (same attribute-value but different class) instances, number of features (#Features) continuous and nominal, class distribution, the majority error and if the dataset have at least one missing value.

Datasets are presented in ascending order of the number of features, as will be in the remaining tables and graphs. Figure 4 shows datasets dimensionality, *i.e.* number of features and number of instances of each dataset. Observe that due to large variation, the number of instances in Figure 4 is represented as $\log_{10}(\#Instances)$.

Dataset	# Instances	Duplicate or conflicting	# Features (cont., nom.)	Class	Class %	Majority Error	Missing Values
bupa	345	4	6 (6,0)	1	42.03%	42.03%	N
				2	57.97%	on value 2	
pima	769	1	8 (8,0)	0	65.02%	34.98%	N
				1	34.98%	on value 0	
breast-cancer	699	8	10 (10,0)	2	65.52%	34.48%	Y
				4	34.48%	on value 2	
hungaria	294	1	13 (13,0)	presence	36.05%	36.05%	Y
				absence	63.95%	on value absence	
crx	690	0	15 (6,9)	yes	44.49%	44.49%	Y
				no	55.51%	on value no	
letter	15000	846	16 (16,0)	A	3.89%	95.92% on value T	Y
				B	3.95%		
				C	3.77%		
				D	3.93%		
				E	3.85%		
				F	3.87%		
				G	3.77%		
				H	3.71%		
				I	3.67%		
				J	3.76%		
				K	3.75%		
				L	3.71%		
				M	4.03%		
				N	3.90%		
				O	3.81%		
				P	3.97%		
Q	3.77%						
R	3.67%						
S	3.67%						

continued on next page

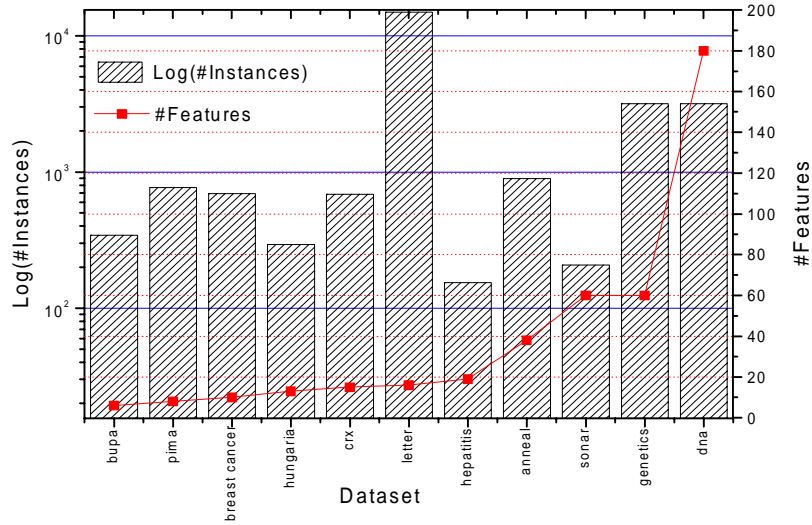


Figure 4: Datasets Dimensionality

continued from previous page

Dataset	# Instances	Duplicate or conflicting	# Features (cont.,nom.)	Class	Class %	Majority Error	Missing Values
hepatitis	155	0	19 (6,13)	T	4.08%	20.65% on value live	Y
				U	3.99%		
				V	3.97%		
				W	3.90%		
				X	4.01%		
				Y	4.02%		
				Z	3.60%		
anneal	898	12	38 (6,32)	1	0.89%	23.83% on value 3	Y
				2	11.02%		
				3	76.17%		
				5	7.46%		
				U	4.45%		
sonar	208	0	60 (60,0)	M	53.37%	46.63% on value M	N
				R	46.63%		
genetics	3190	185	60 (0,60)	N	51.88%	48.12% on value N	N
				EI	24.04%		
				IE	24.08%		
dna	3186	185	180 (0,180)	1	24.07%	48.09% on value 3	N
				2	24.01%		
				3	51.91%		

Table 2: Datasets Summary Descriptions

5 Experimental Setup

We performed a serie of experiments using the five algorithms, as well as the datasets described in Section 3 and 4 respectively. It should be observed that the original data has not been pre-processed in any way, for example by removing or replacing missing values or transforming nominal to numerical attributes.

First of all, it is important to state that a two independent steps were performed in the experiments:

- The first step run the wrapper approach using one of the five inducers as black box, *i.e.* the black box wrapper inducer;
- Using the selected features obtained in step 1, the accuracy was computed for each one of the five inducers, *i.e.* the accuracy estimator inducer.

A more detailed explanation is given as follows. First of all, the five inducers — $\mathcal{C}4.5$, $\mathcal{CN}2$, IB, NB and TM — were applied through the wrapper as black box inducers to every original dataset. Wrapper inducers were run with default setting for all parameters; no attempt was made to tune any inducer.

Afterwards we ran the $\mathcal{M}\mathcal{L}\mathcal{C}++$ wrapper with each algorithm over every dataset using best-first search and compound operators. We evaluated the wrapper on both, forward selection, where the initial state is the empty set of features, as well as backward selection where the initial state is the full set of features, obtaining for each algorithm, dataset and forward or backward selection the best features selected by the $\mathcal{M}\mathcal{L}\mathcal{C}++$ wrapper. It is expected that forward should be less expensive than backward selection, since building classifiers when there are few features in the training data should be computationally faster. Still, in theory, as pointed out in (Kohavi, 1997), going backward from the full set of features may capture interactive features more easily although the method is extremely expensive.

Finally, using for each case — besides all features — the selected features, each inducer was applied to the correspondent reduced dataset and accuracy was measured using ten-fold cross-validation² as well as ten-fold stratified cross-validation³. Several results are presented for each algorithm on each data set. Accuracies are presented as mean and standard deviation (proportion) of the ten folds. When appropriate, log values are also presented and should be always considered as \log_{10} .

6 Experimental Results

6.1 Summary Tables Description

In what follows, results are grouped by dataset. For each dataset, four tables (results not available are indicated with “N/A”) are presented from Section 6.2 through 6.12 where:

- The first table describes each feature in the dataset: feature number (features numbering starts at zero), feature name and type (continuous or nominal). For nominal features, the maximum possible number of values (as described in the *names* file) and the actual number of values (the ones really found in the dataset through the $\mathcal{M}\mathcal{L}\mathcal{C}++$ *info* utility) are shown. It should be observed that if a number of actual nominal values is greater than the possible number of values indicates that there are missing values for that specific attribute. The reverse is not true.
- The second table describes wrapper selected features. To indicate the experiment we are referring to, the notation $FSS(method, inducer)$ is used, where:
 - $method \in \{wf, wb\}$ indicating if wrapper forward (wf) or backward selection (wb) of features has been used;
 - $inducer \in \{\mathcal{C}4.5, \mathcal{CN}2, IB, NB, TM\}$ indicating the algorithm that has been wrapped around.

²A ten-fold cross-validation is performed by dividing the data into ten mutually exclusive subsets (folds) of cases of approximately equal size. The inducer is trained and tested 10 times; each time tested on a fold and trained on the dataset minus the fold. The cross-validation estimate of accuracy is the average of the estimated accuracies from the 10 folds.

³Similar to 10-fold cross-validation but the folds are stratified so that they contain approximately the same proportion of labels as the original dataset.

For each $FSS(method,inducer)$ considered this table shows the features subset selected, the number of features in the selected subset ($\#F$), proportion of selected features ($\%F$) as well as the time taken by this wrapper method to obtain the features subset. Time (in seconds) is related to a standard *Indigo 2* Silicon Graphics workstation.

- The third table shows similar information than the second one, but in a different way such that it is easy to visualize common features found by every $FSS(method,inducer)$ tested.
- The fourth table show accuracy of each inducer (mean and standard deviation) using 10-fold cross-validation (10-cv) and 10-fold stratified cross-validation (10-strat-cv) using all features as well as the features subset selected by every $FSS(method,inducer)$ considered. Each column represents the inducer used for accuracy estimation and each row represents the feature subset used. For instance, the first column indicates accuracies using $\mathcal{C}4.5$ as inducer; the first row of this column indicates accuracy of $\mathcal{C}4.5$ using all features in the dataset, the second row indicates accuracy using the feature subset selected by $FSS(wf,\mathcal{C}4.5)$ and so on.

6.2 Bupa

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	mcv	-	26	continuous
#1	alkphos	-	78	continuous
#2	sgpt	-	67	continuous
#3	sgot	-	47	continuous
#4	gammagt	-	94	continuous
#5	drinks	-	16	continuous

Table 3: Bupa Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf, $\mathcal{C}4.5$)	0 1 2 4 5	5	83.33%	28.70
FSS(wb, $\mathcal{C}4.5$)	0 1 2 4 5	5	83.33%	23.70
FSS(wf, $\mathcal{CN}2$)	0 2 3 4 5	5	83.33%	189.70
FSS(wb, $\mathcal{CN}2$)	0 2 3 4 5	5	83.33%	164.10
FSS(wf,IB)	0 2 3 4 5	5	83.33%	80.10
FSS(wb,IB)	0 2 3 4 5	5	83.33%	47.70
FSS(wf,NB)	0 1 2 4 5	5	83.33%	8.80
FSS(wb,NB)	0 1 2 4 5	5	83.33%	6.50
FSS(wf,TM)	2 5	2	33.33%	12.80
FSS(wb,TM)	2 5	2	33.33%	14.70

Table 4: Bupa Time for Selecting Features

Feature Number	FSS									
	(wf, $\mathcal{C}4.5$)	(wb, $\mathcal{C}4.5$)	(wf, $\mathcal{CN}2$)	(wb, $\mathcal{CN}2$)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0	•	•	•	•	•	•	•	•	•	•
#1	•	•	•	•	•	•	•	•	•	•
#2	•	•	•	•	•	•	•	•	•	•
#3	•	•	•	•	•	•	•	•	•	•
#4	•	•	•	•	•	•	•	•	•	•
#5	•	•	•	•	•	•	•	•	•	•
Total 6	5	5	5	5	5	5	5	5	2	2
100%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	33.33%	33.33%

Table 5: Bupa Wrapper Selected Features

bupa 10-cv	C4.5	CN2	IB	NB	TM
all features	67.30±2.79	64.65±2.01	61.08±2.40	53.99±2.69	59.68±1.48
FSS(wf,C4.5)	69.01±3.29	63.74±2.75	58.48±3.14	60.61±2.48	59.68±1.48
FSS(wb,C4.5)	69.01±3.29	63.74±2.75	58.48±3.14	60.61±2.48	59.68±1.48
FSS(wf,CN2)	67.85±3.29	67.83±2.96	66.68±1.98	53.70±2.66	59.68±1.48
FSS(wb,CN2)	67.85±3.29	67.83±2.96	66.68±1.98	53.70±2.66	59.68±1.48
FSS(wf,IB)	67.85±3.29	67.83±2.96	66.68±1.98	53.70±2.66	59.68±1.48
FSS(wb,IB)	67.85±3.29	67.83±2.96	66.68±1.98	53.70±2.66	59.68±1.48
FSS(wf,NB)	69.01±3.29	63.74±2.75	58.48±3.14	60.61±2.48	59.68±1.48
FSS(wb,NB)	69.01±3.29	63.74±2.75	58.48±3.14	60.61±2.48	59.68±1.48
FSS(wf,TM)	56.50±1.60	61.40±2.19	59.39±2.52	55.92±2.76	61.42±0.99
FSS(wb,TM)	56.50±1.60	61.40±2.19	59.39±2.52	55.92±2.76	61.42±0.99
bupa 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	68.71±1.73	67.82±2.11	61.66±2.82	55.44±2.95	59.71±0.44
FSS(wf,C4.5)	66.97±2.76	66.08±2.31	59.91±3.24	60.90±2.32	59.71±0.44
FSS(wb,C4.5)	66.97±2.76	66.08±2.31	59.91±3.24	60.90±2.32	59.71±0.44
FSS(wf,CN2)	66.37±1.86	65.81±1.83	65.82±1.68	55.45±2.72	59.71±0.44
FSS(wb,CN2)	66.37±1.86	65.81±1.83	65.82±1.68	55.45±2.72	59.71±0.44
FSS(wf,IB)	66.37±1.86	65.81±1.83	65.82±1.68	55.45±2.72	59.71±0.44
FSS(wb,IB)	66.37±1.86	65.81±1.83	65.82±1.68	55.45±2.72	59.71±0.44
FSS(wf,NB)	66.97±2.76	66.08±2.31	59.91±3.24	60.90±2.32	59.71±0.44
FSS(wb,NB)	66.97±2.76	66.08±2.31	59.91±3.24	60.90±2.32	59.71±0.44
FSS(wf,TM)	56.25±1.36	61.40±3.57	58.80±3.15	57.66±1.78	61.13±2.03
FSS(wb,TM)	56.25±1.36	61.40±3.57	58.80±3.15	57.66±1.78	61.13±2.03

Table 6: Bupa Accuracies

6.3 Pima

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	Number	-	17	continuous
#1	Plasma	-	136	continuous
#2	Diastolic	-	47	continuous
#3	Triceps	-	51	continuous
#4	Two	-	186	continuous
#5	Body	-	248	continuous
#6	Diabetes	-	517	continuous
#7	Age	-	52	continuous

Table 7: Pima Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	0 1 4 5 6	5	62.50%	81.90
FSS(wb,C4.5)	1 2 3 5 7	5	62.50%	89.20
FSS(wf,CN2)	0 1 2 4 5 6 7	7	87.50%	1292.10
FSS(wb,CN2)	0 1 2 4 5 6 7	7	87.50%	790.70
FSS(wf,IB)	0 1 7	3	37.50%	335.10
FSS(wb,IB)	1 3 4 5 7	5	62.50%	357.00
FSS(wf,NB)	1 3 5 6 7	5	62.50%	27.20
FSS(wb,NB)	1 3 5 6 7	5	62.50%	46.40
FSS(wf,TM)	1	1	12.50%	28.30
FSS(wb,TM)	0	1	12.50%	68.50

Table 8: Pima Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0	•		•	•	•					•
#1	•	•	•	•	•	•	•	•	•	•
#2		•	•	•						

continued on next page

continued from previous page

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#3		•				•	•	•		
#4	•		•	•		•				
#5	•	•	•	•		•	•	•		
#6	•		•	•		•	•	•		
#7		•	•	•	•	•	•	•		
Total 8	5	5	7	7	3	5	5	5	1	1
100%	62.50%	62.50%	87.50%	87.50%	37.50%	62.50%	62.50%	62.50%	12.50%	12.50%

Table 9: Pima Wrapper Selected Features

pima 10-cv	C4.5	CN2	IB	NB	TM
all features	74.13±1.28	74.88±1.97	68.79±2.19	75.29±1.09	64.89±1.25
FSS(wf,C4.5)	75.16±1.01	73.97±1.13	67.11±1.02	76.07±0.86	64.89±1.25
FSS(wb,C4.5)	76.99±1.07	70.97±1.37	69.82±1.23	76.59±1.23	64.89±1.25
FSS(wf,CN2)	74.12±0.99	76.31±1.22	70.09±1.89	75.68±1.13	64.89±1.25
FSS(wb,CN2)	74.12±0.99	76.31±1.22	70.09±1.89	75.68±1.13	64.89±1.25
FSS(wf,IB)	72.30±1.23	71.36±1.65	72.56±1.62	73.47±1.51	64.89±1.27
FSS(wb,IB)	73.47±0.80	73.58±1.08	71.00±1.42	75.55±0.86	64.89±1.25
FSS(wf,NB)	72.95±0.79	73.45±1.47	69.95±1.46	77.25±0.91	64.89±1.25
FSS(wb,NB)	72.95±0.79	73.45±1.47	69.95±1.46	77.25±0.91	64.89±1.25
FSS(wf,TM)	71.65±1.18	72.01±1.18	69.57±0.97	74.77±0.98	68.01±1.18
FSS(wb,TM)	67.75±1.68	68.11±1.56	67.75±1.56	66.97±2.02	67.49±1.58
pima 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	74.26±1.13	74.62±1.38	69.18±1.92	75.55±1.05	64.89±0.16
FSS(wf,C4.5)	74.77±1.04	73.06±1.32	66.98±1.16	76.20±0.96	64.89±0.16
FSS(wb,C4.5)	75.95±0.98	71.10±1.54	69.30±1.28	76.85±1.36	64.89±0.16
FSS(wf,CN2)	73.99±0.94	74.75±1.43	69.70±1.56	75.81±1.26	64.89±0.16
FSS(wb,CN2)	73.99±0.94	74.75±1.43	69.70±1.56	75.81±1.26	64.89±0.16
FSS(wf,IB)	72.30±1.04	71.75±1.83	72.56±1.44	73.60±1.43	64.89±0.25
FSS(wb,IB)	73.86±0.83	73.45±0.83	70.87±1.43	75.81±1.17	64.89±0.16
FSS(wf,NB)	72.69±0.98	73.19±1.03	68.91±1.38	77.25±0.99	64.89±0.16
FSS(wb,NB)	72.69±0.98	73.19±1.03	68.91±1.38	77.25±0.99	64.89±0.16
FSS(wf,TM)	72.42±1.11	71.62±1.40	69.18±1.08	75.03±1.13	67.75±1.31
FSS(wb,TM)	67.75±1.11	68.11±0.80	67.75±0.74	66.97±1.51	67.49±0.79

Table 10: Pima Accuracies

6.4 Breast Cancer

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	Sample code number	-	645	continuous
#1	Clump Thickness	-	10	continuous
#2	Uniformity of Cell Size	-	10	continuous
#3	Uniformity of Cell Shape	-	10	continuous
#4	Marginal Adhesion	-	10	continuous
#5	Single Epithelial Cell Size	-	10	continuous
#6	Bare Nuclei	-	10	continuous
#7	Bland Chromatin	-	10	continuous
#8	Normal Nucleoli	-	10	continuous
#9	Mitoses	-	9	continuous

Table 11: Breast Cancer Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	1 2 4 5 6 7 9	7	70.00%	135.80
FSS(wb,C4.5)	1 2 4 5 6 7 9	7	70.00%	116.70
FSS(wf,CN2)	1 2 6 8 9	5	50.00%	697.70
FSS(wb,CN2)	0 1 2 3 4 6 7 8 9	9	90.00%	564.30
FSS(wf,IB)	0 1 2 5 6 8	6	60.00%	537.10

continued on next page

continued from previous page

Inducer	Selected Features	#F	%F	Time (s)
FSS(wb,IB)	0 1 4 5 6 8	6	60.00%	566.70
FSS(wf,NB)	0 1 2 3 6	5	50.00%	33.20
FSS(wb,NB)	0 1 3 4 6 7 8	7	70.00%	37.70
FSS(wf,TM)	3	1	10.00%	36.70
FSS(wb,TM)	2 5	2	20.00%	61.90

Table 12: Breast Cancer Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,C \mathcal{N} 2)	(wb,C \mathcal{N} 2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0										
#1	•									
#2	•									
#3										
#4	•									
#5	•									
#6	•									
#7	•									
#8										
#9	•									
Total 10	7	7	5	9	6	6	5	7	1	2
100%	70.00%	70.00%	50.00%	90.00%	60.00%	60.00%	50.00%	70.00%	10.00%	20.00%

Table 13: Breast Cancer Wrapper Selected Features

breast cancer 10-cv	C4.5	C \mathcal{N} 2	IB	NB	TM
all features	94.14±0.84	95.27±1.11	95.28±1.02	96.14±0.77	65.52±1.72
FSS(wf,C4.5)	96.00±0.55	95.00±0.48	95.71±0.64	95.85±0.86	65.52±1.72
FSS(wb,C4.5)	96.00±0.55	95.00±0.48	95.71±0.64	95.85±0.86	65.52±1.72
FSS(wf,C \mathcal{N} 2)	95.43±0.67	96.43±0.86	95.85±0.78	95.85±0.89	67.95±1.65
FSS(wb,C \mathcal{N} 2)	93.85±0.82	96.30±0.68	95.00±0.83	96.28±0.68	65.52±1.72
FSS(wf,IB)	95.28±0.60	95.71±0.77	96.57±0.74	95.99±0.76	65.52±1.72
FSS(wb,IB)	94.57±1.02	94.84±0.83	96.71±0.60	95.42±0.73	65.52±1.72
FSS(wf,NB)	94.86±0.77	95.13±0.86	96.00±0.55	95.99±0.76	65.52±1.72
FSS(wb,NB)	95.85±0.54	95.56±0.75	95.42±0.63	96.42±0.61	65.52±1.72
FSS(wf,TM)	92.27±0.65	92.29±0.65	92.27±0.65	92.27±0.65	92.27±0.65
FSS(wb,TM)	92.99±0.58	93.57±0.64	93.42±0.57	93.57±0.65	92.13±0.65
breast cancer 10-strat-cv	C4.5	C \mathcal{N} 2	IB	NB	TM
all features	94.28±0.56	94.99±1.42	95.00±0.88	96.14±0.68	66.09±0.23
FSS(wf,C4.5)	95.00±0.83	95.00±0.71	95.71±0.56	95.85±0.62	66.09±0.23
FSS(wb,C4.5)	95.00±0.83	95.00±0.71	95.71±0.56	95.85±0.62	66.09±0.23
FSS(wf,C \mathcal{N} 2)	95.57±0.69	96.85±0.60	95.85±0.91	95.85±0.84	68.53±0.37
FSS(wb,C \mathcal{N} 2)	94.57±0.79	95.57±0.97	94.85±0.68	96.28±0.61	66.09±0.23
FSS(wf,IB)	95.28±0.77	95.56±1.03	96.57±0.80	95.99±0.67	66.09±0.23
FSS(wb,IB)	94.28±1.15	95.13±0.67	96.57±0.57	95.42±0.70	66.09±0.23
FSS(wf,NB)	95.00±0.83	95.00±0.77	96.00±0.70	95.85±0.84	66.09±0.23
FSS(wb,NB)	95.71±0.60	94.99±0.91	95.28±0.80	96.42±0.53	66.09±0.23
FSS(wf,TM)	92.27±0.69	92.27±0.68	92.27±0.69	92.27±0.69	92.27±0.69
FSS(wb,TM)	93.13±0.51	93.29±0.67	93.28±0.64	93.57±0.68	91.85±0.77

Table 14: Breast Cancer Accuracies

6.5 Hungaria

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	age	-	38	continuous
#1	sex	-	2	continuous
#2	cp	-	4	continuous
#3	trestbps	-	31	continuous
#4	chol	-	153	continuous
#5	fbs	-	2	continuous

continued on next page

<i>continued from previous page</i>				
Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#6	restecg	-	3	continuous
#7	thalach	-	71	continuous
#8	exang	-	2	continuous
#9	oldpeak	-	10	continuous
#10	slope	-	3	continuous
#11	ca	-	2	continuous
#12	thal	-	3	continuous

Table 15: Hungaria Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	0 9 10 11 12	5	38.46%	83.60
FSS(wb,C4.5)	0 4 5 6 9 10 11 12	8	61.54%	104.80
FSS(wf,CN2)	8 10 11 12	4	30.77%	314.20
FSS(wb,CN2)	1 2 3 7 10 11 12	7	53.85%	1242.90
FSS(wf,IB)	8 10	2	15.38%	112.20
FSS(wb,IB)	1 2 4 5 6 8 9 11 12	9	69.23%	185.00
FSS(wf,NB)	0 1 3 4 8 9 10 12	8	61.54%	23.10
FSS(wb,NB)	0 2 4 5 7 8 9 10 12	9	69.23%	20.70
FSS(wf,TM)	8 10 12	3	23.08%	22.20
FSS(wb,TM)	8 10 12	3	23.08%	40.40

Table 16: Hungaria Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0	•	•					•	•		
#1				•		•	•			
#2				•		•		•		
#3				•			•			
#4		•				•	•	•		
#5		•				•		•		
#6		•				•		•		
#7				•				•		
#8			•		•	•	•	•	•	•
#9	•	•				•	•	•		
#10	•	•	•	•	•		•	•	•	•
#11	•	•	•	•		•		•		
#12	•	•	•	•		•	•	•	•	•
Total 13	5	8	4	7	2	9	8	9	3	3
100%	38.46%	61.54%	30.77%	53.85%	15.38%	69.23%	61.54%	69.23%	23.08%	23.08%

Table 17: Hungaria Wrapper Selected Features

hungaria 10-cv	C4.5	CN2	IB	NB	TM
all features	79.92±2.69	78.56±2.19	78.57±2.03	84.01±2.04	63.92±3.49
FSS(wf,C4.5)	82.97±2.71	77.54±2.63	78.90±1.97	81.95±2.23	69.72±3.13
FSS(wb,C4.5)	82.97±2.71	75.15±2.86	73.20±2.96	83.00±2.26	63.57±3.45
FSS(wf,CN2)	81.28±1.88	83.99±2.00	82.95±2.01	81.97±2.05	83.63±1.99
FSS(wb,CN2)	78.60±2.26	84.03±2.59	78.93±1.79	80.00±2.22	65.63±2.83
FSS(wf,IB)	81.28±1.88	83.66±2.40	83.64±2.40	81.61±1.95	83.30±2.33
FSS(wb,IB)	80.61±2.45	79.28±1.43	83.37±1.90	82.00±2.46	63.57±3.14
FSS(wf,NB)	80.94±2.30	79.97±2.71	75.53±2.44	85.68±1.77	63.92±3.49
FSS(wb,NB)	79.95±2.38	76.27±3.24	73.83±2.00	85.05±1.69	63.92±3.49
FSS(wf,TM)	81.28±1.88	83.30±2.16	82.95±2.01	82.63±1.96	83.63±1.99
FSS(wb,TM)	81.28±1.88	83.30±2.16	82.95±2.01	82.63±1.96	83.63±1.99
hungaria 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	77.52±4.20	77.93±3.06	78.92±1.93	83.67±2.83	63.94±0.54
FSS(wf,C4.5)	82.97±3.27	78.27±2.92	78.55±2.69	82.29±2.78	69.75±1.42
FSS(wb,C4.5)	82.97±3.27	74.90±3.85	73.20±3.17	83.00±2.88	63.60±0.75
FSS(wf,CN2)	81.28±2.51	83.66±2.60	82.62±2.49	81.97±2.72	83.30±2.48
FSS(wb,CN2)	78.98±2.79	80.65±3.94	79.26±2.23	79.66±1.58	65.66±1.25

continued on next page

continued from previous page

	C4.5	CN2	IB	NB	TM
FSS(wf,IB)	81.28±2.51	83.66±3.02	83.64±3.02	81.61±2.60	83.30±2.92
FSS(wb,IB)	79.92±3.55	78.28±2.51	82.00±2.56	81.66±3.43	63.60±1.15
FSS(wf,NB)	79.57±3.47	73.15±2.93	75.89±2.25	85.68±2.27	63.94±0.54
FSS(wb,NB)	80.25±3.70	76.23±3.81	74.49±2.03	84.36±2.76	63.94±0.54
FSS(wf,TM)	81.28±2.51	82.98±2.66	82.62±2.49	82.63±2.56	83.30±2.48
FSS(wb,TM)	81.28±2.51	82.98±2.66	82.62±2.49	82.63±2.56	83.30±2.48

Table 18: Hungaria Accuracies

6.6 Crx

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	A1	2	3	nominal
#1	A2	-	349	continuous
#2	A3	-	215	continuous
#3	A4	4	4	nominal
#4	A5	3	4	nominal
#5	A6	14	15	nominal
#6	A7	9	10	nominal
#7	A8	-	132	continuous
#8	A9	2	2	nominal
#9	A10	2	2	nominal
#10	A11	-	23	continuous
#11	A12	2	2	nominal
#12	A13	3	3	nominal
#13	A14	-	170	continuous
#14	A15	-	240	continuous

Table 19: Crx Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	1 2 3 8 10 11 12	7	46.67%	416.50
FSS(wb,C4.5)	1 3 4 6 7 8 9 10 11 13 14	11	73.33%	324.80
FSS(wf,CN2)	3 4 8 9 12	5	33.33%	464.40
FSS(wb,CN2)	4 8 9 10 12 13 14	7	46.67%	3628.70
FSS(wf,IB)	3 4 8 9 12	5	33.33%	353.80
FSS(wb,IB)	0 1 2 3 4 5 6 7 8 9 10 12 13	13	86.67%	544.70
FSS(wf,NB)	3 4 8 9 12 13	6	40.00%	46.40
FSS(wb,NB)	0 4 6 8 9 12 13	7	46.67%	67.40
FSS(wf,TM)	3 4 8 9 12	5	33.33%	58.40
FSS(wb,TM)	3 8 9	3	20.00%	158.30

Table 20: Crx Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0						•		•		
#1	•	•				•				
#2	•					•				
#3	•		•		•	•	•		•	•
#4		•	•	•	•	•	•	•	•	
#5						•				
#6		•				•		•		
#7		•				•				
#8	•	•	•	•	•	•	•	•	•	•
#9		•	•	•	•	•	•	•	•	•
#10	•	•		•		•				
#11	•	•								
#12	•		•	•	•	•	•	•	•	
#13		•		•		•	•	•		
#14		•		•						

continued on next page

<i>continued from previous page</i>										
Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
Total 15	7	11	5	7	5	13	6	7	5	3
100%	46.67%	73.33%	33.33%	46.67%	33.33%	86.67%	40.00%	46.67%	33.33%	20.00%

Table 21: Crx Wrapper Selected Features

crx 10-cv	C4.5	CN2	IB	NB	TM
all features	85.36±0.88	81.90±1.55	81.74±1.13	77.68±1.56	55.51±1.97
FSS(wf,C4.5)	85.94±1.14	81.74±1.21	80.00±1.52	80.72±1.30	55.51±1.97
FSS(wb,C4.5)	87.83±1.04	82.33±0.86	80.72±1.38	75.80±1.56	55.51±1.97
FSS(wf,CN2)	85.51±1.43	86.97±1.12	86.96±1.22	86.38±1.44	86.52±1.04
FSS(wb,CN2)	85.22±0.77	87.41±1.18	83.48±1.42	76.09±1.44	58.55±1.98
FSS(wf,IB)	85.51±1.43	86.97±1.12	86.96±1.22	86.38±1.44	86.52±1.04
FSS(wb,IB)	83.91±1.11	80.74±1.50	83.62±1.18	81.01±1.09	55.51±1.97
FSS(wf,NB)	84.20±0.82	84.52±1.39	84.06±2.21	87.10±1.23	75.22±1.76
FSS(wb,NB)	84.93±1.04	83.79±1.37	82.46±1.13	87.25±1.26	70.29±1.09
FSS(wf,TM)	85.51±1.43	86.97±1.12	86.96±1.22	86.38±1.44	86.52±1.04
FSS(wb,TM)	85.65±1.53	86.69±1.36	86.96±1.10	86.67±1.36	86.96±1.10
crx 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	84.35±1.18	83.20±1.21	81.74±0.90	77.97±1.12	55.51±0.22
FSS(wf,C4.5)	85.65±1.17	82.04±0.84	80.43±1.61	80.87±0.91	55.51±0.22
FSS(wb,C4.5)	86.67±1.26	82.77±1.03	80.58±1.19	76.09±1.13	55.51±0.22
FSS(wf,CN2)	85.36±1.47	86.83±1.23	87.39±1.18	86.38±1.35	86.96±0.99
FSS(wb,CN2)	84.93±1.17	85.83±0.86	83.77±1.33	76.38±1.16	58.70±0.92
FSS(wf,IB)	85.36±1.47	86.83±1.23	87.39±1.18	86.38±1.35	86.96±0.99
FSS(wb,IB)	84.78±1.11	83.07±0.99	83.48±1.11	81.30±0.93	55.51±0.22
FSS(wf,NB)	84.78±1.21	84.95±1.47	84.49±2.12	87.10±1.25	75.36±1.67
FSS(wb,NB)	85.36±1.23	84.07±1.03	82.75±1.05	87.39±1.18	70.72±1.45
FSS(wf,TM)	85.36±1.47	86.83±1.23	87.39±1.18	86.38±1.35	86.96±0.99
FSS(wb,TM)	85.65±1.42	86.69±1.25	86.96±1.04	86.67±1.26	86.96±1.04

Table 22: Crx Accuracies

6.7 Letter

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	x-box	-	15	continuous
#1	y-box	-	16	continuous
#2	width	-	16	continuous
#3	high	-	16	continuous
#4	onpix	-	16	continuous
#5	x-bar	-	16	continuous
#6	y-bar	-	16	continuous
#7	x2bar	-	16	continuous
#8	y2bar	-	16	continuous
#9	xybar	-	16	continuous
#10	x2ybr	-	16	continuous
#11	xy2br	-	16	continuous
#12	x-ege	-	16	continuous
#13	xegvy	-	16	continuous
#14	y-ege	-	16	continuous
#15	yegvx	-	15	continuous

Table 23: Letter Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	1 5 7 8 9 10 11 12 13 14 15	11	68.75%	2003.5
FSS(wb,C4.5)	1 5 7 8 9 10 11 12 13 14 15	11	68.75%	1167.9
FSS(wf,CN2)	1 6 7 8 9 10 11 12 14	9	56.25%	33446.1

continued on next page

continued from previous page				
Inducer	Selected Features	#F	%F	Time (s)
FSS(wb,CN2)	1 6 7 8 9 10 11 12 14	9	56.25%	68115.1
FSS(wf,IB)	5 6 7 8 9 10 11 12 13 14 15	11	68.75%	53490.4
FSS(wb,IB)	5 6 7 8 9 10 11 12 13 14 15	11	68.75%	74728.6
FSS(wf,NB)	1 5 6 7 8 9 10 11 12 13 14 15	12	75.00%	784.2
FSS(wb,NB)	1 5 6 7 8 9 10 11 12 13 14 15	12	75.00%	865.3
FSS(wf,TM)	7 8 12 14	4	25.00%	2033.1
FSS(wb,TM)	7 8 12 14	4	25.00%	929.7

Table 24: Letter Wrapper Selected Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0										
#1	•	•	•	•			•	•		
#2										
#3										
#4										
#5	•	•			•	•	•	•		
#6			•	•	•	•	•	•		
#7	•	•	•	•	•	•	•	•	•	•
#8	•	•	•	•	•	•	•	•	•	•
#9	•	•	•	•	•	•	•	•		
#10	•	•	•	•	•	•	•	•		
#11	•	•	•	•	•	•	•	•		
#12	•	•	•	•	•	•	•	•	•	•
#13	•	•			•	•	•	•		
#14	•	•	•	•	•	•	•	•	•	•
#15	•	•			•	•	•	•		
Total 16	11	11	9	9	11	11	12	12	4	4
100%	68.75%	68.75%	56.25%	56.25%	68.75%	68.75%	75.00%	75.00%	25.00%	25.00%

Table 25: Letter Wrapper Selected Features

letter 10-cv	C4.5	CN2	IB	NB	TM
all features	86.66±0.41	70.46±0.40	95.13±0.13	64.51±0.51	12.10±0.31
FSS(wf,C4.5)	86.97±0.35	71.92±0.33	94.97±0.22	65.00±0.56	21.25±0.33
FSS(wb,C4.5)	86.97±0.35	71.92±0.33	94.97±0.22	65.00±0.56	21.25±0.33
FSS(wf,CN2)	85.77±0.31	75.74±0.51	92.56±0.22	62.85±0.54	24.49±0.39
FSS(wb,CN2)	85.77±0.31	75.74±0.51	92.56±0.22	62.85±0.54	24.49±0.39
FSS(wf,IB)	86.54±0.31	71.78±0.28	96.19±0.16	66.25±0.56	30.57±0.37
FSS(wb,IB)	86.54±0.31	71.78±0.28	96.19±0.16	66.25±0.56	30.57±0.37
FSS(wf,NB)	86.55±0.36	71.32±0.30	95.49±0.20	66.26±0.56	18.95±0.31
FSS(wb,NB)	86.55±0.36	71.32±0.30	95.49±0.20	66.26±0.56	18.95±0.31
FSS(wf,TM)	72.17±0.40	N/A	72.53±0.39	41.42±0.43	68.36±0.39
FSS(wb,TM)	72.17±0.40	N/A	72.53±0.39	41.42±0.43	68.36±0.39
letter 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	86.59±0.34	70.34±0.30	95.17±0.13	64.55±0.36	12.79±0.31
FSS(wf,C4.5)	86.93±0.31	71.35±0.19	95.00±0.19	65.01±0.38	21.84±0.33
FSS(wb,C4.5)	86.93±0.31	71.35±0.19	95.00±0.19	65.01±0.38	21.84±0.33
FSS(wf,CN2)	85.56±0.40	76.17±0.20	92.56±0.20	62.98±0.36	25.11±0.38
FSS(wb,CN2)	85.56±0.40	76.17±0.20	92.56±0.20	62.98±0.36	25.11±0.38
FSS(wf,IB)	86.83±0.38	72.17±0.25	96.19±0.13	66.27±0.44	31.15±0.34
FSS(wb,IB)	86.83±0.38	72.17±0.25	96.19±0.13	66.27±0.44	31.15±0.34
FSS(wf,NB)	86.71±0.38	71.54±0.33	95.51±0.15	66.31±0.43	19.59±0.36
FSS(wb,NB)	86.71±0.38	71.54±0.33	95.51±0.15	66.31±0.43	19.59±0.36
FSS(wf,TM)	72.13±0.48	N/A	72.61±0.45	41.49±0.30	68.61±0.46
FSS(wb,TM)	72.13±0.48	N/A	72.61±0.45	41.49±0.30	68.61±0.46

Table 26: Letter Accuracies

6.8 Hepatitis

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	age	-	49	continuous
#1	female	2	2	nominal
#2	steroid	2	3	nominal
#3	antivirals	2	2	nominal
#4	fatigue	2	3	nominal
#5	malaise	2	3	nominal
#6	anorexia	2	3	nominal
#7	liver-big	2	3	nominal
#8	liver-firm	2	3	nominal
#9	spleen-palpable	2	3	nominal
#10	spiders	2	3	nominal
#11	ascites	2	3	nominal
#12	varices	2	3	nominal
#13	bilirubin	-	34	continuous
#14	alk-phosphate	-	83	continuous
#15	sgot	-	84	continuous
#16	albumin	-	29	continuous
#17	protime	-	44	continuous
#18	histology	2	2	nominal

Table 27: Hepatitis Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	11 12 13 16 18	5	26.32%	77.20
FSS(wb,C4.5)	0 1 2 5 8 10 17	7	36.84%	149.60
FSS(wf,CN2)	1 3 4 6 9 11 16	7	36.84%	700.40
FSS(wb,CN2)	0 1 2 3 4 6 7 8 10 11 12 14 15 16 17 18	16	84.21%	583.00
FSS(wf,IB)	1 7 8 11 12	5	26.32%	112.60
FSS(wb,IB)	1 2 3 4 5 6 7 8 9 10 11 14 15 17 18	15	78.95%	136.90
FSS(wf,NB)	0 1 2 3 4 7 11 16 17	9	47.37%	17.40
FSS(wb,NB)	0 1 4 5 7 9 10 12 13 14 16	11	57.89%	26.00
FSS(wf,TM)	3 10 16	3	15.79%	16.00
FSS(wb,TM)	0 1 2 3 4 5 6 7 8 9	10	52.63%	25.00

Table 28: Hepatitis Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0		•		•				•		•
#1		•		•	•		•	•		•
#2		•		•			•			•
#3			•	•			•		•	•
#4			•	•			•			•
#5		•						•		•
#6			•	•				•		•
#7				•	•		•			•
#8		•		•	•			•		•
#9			•					•		•
#10		•		•				•		•
#11	•		•	•	•		•		•	
#12	•			•	•					
#13	•							•		
#14				•				•		
#15				•				•		
#16	•		•	•			•	•	•	
#17		•		•			•			
#18	•			•			•			
Total 19	5	7	7	16	5	15	9	11	3	10
100%	26.32%	36.84%	36.84%	84.21%	26.32%	78.95%	47.37%	57.89%	15.79%	52.63%

Table 29: Hepatitis Wrapper Selected Features

hepatitis 10-cv	$\mathcal{C}4.5$	$\mathcal{CN}2$	IB	NB	TM
all features	78.08±3.20	83.82±1.80	81.96±2.44	83.75±2.50	79.33±2.57
FSS(wf, $\mathcal{C}4.5$)	85.83±2.67	81.79±3.24	78.54±2.32	81.17±3.59	75.42±2.77
FSS(wb, $\mathcal{C}4.5$)	87.75±1.77	82.49±2.85	72.92±2.45	83.17±1.77	79.33±2.57
FSS(wf, $\mathcal{CN}2$)	81.21±3.01	91.59±2.18	81.29±4.70	85.79±2.80	78.71±3.34
FSS(wb, $\mathcal{CN}2$)	78.83±2.88	87.01±2.57	80.67±2.71	83.83±2.00	79.33±2.57
FSS(wf,IB)	84.50±3.07	82.64±3.42	87.13±2.98	80.62±3.01	87.13±2.98
FSS(wb,IB)	79.29±2.34	82.64±2.81	87.21±2.46	83.17±2.24	79.33±2.57
FSS(wf,NB)	84.50±3.07	80.53±2.26	78.54±3.80	87.08±2.52	79.33±2.57
FSS(wb,NB)	73.50±2.86	82.00±3.52	75.96±3.06	88.96±2.19	79.33±2.57
FSS(wf,TM)	80.67±3.42	83.26±3.05	82.71±3.52	85.79±3.05	81.38±3.32
FSS(wb,TM)	74.08±3.74	79.95±2.65	75.46±2.69	81.21±2.84	80.67±2.15
hepatitis 10-strat-cv	$\mathcal{C}4.5$	$\mathcal{CN}2$	IB	NB	TM
all features	79.38±2.27	81.75±3.83	81.29±2.23	83.08±3.72	79.33±0.89
FSS(wf, $\mathcal{C}4.5$)	84.50±2.00	79.25±3.67	77.88±3.38	81.17±4.22	75.42±1.96
FSS(wb, $\mathcal{C}4.5$)	88.38±1.62	84.49±2.45	72.88±3.06	83.79±2.32	79.33±0.89
FSS(wf, $\mathcal{CN}2$)	82.54±2.62	90.99±1.66	81.25±3.94	85.75±3.86	79.33±3.17
FSS(wb, $\mathcal{CN}2$)	80.08±2.75	87.13±2.81	80.00±2.46	83.83±2.80	79.33±0.89
FSS(wf,IB)	84.50±2.57	81.97±2.84	87.13±2.67	82.50±3.08	87.13±2.67
FSS(wb,IB)	81.88±2.57	81.25±2.03	86.54±2.17	83.79±3.44	79.33±0.89
FSS(wf,NB)	84.50±2.15	85.05±2.07	78.58±2.94	87.08±2.39	79.33±0.89
FSS(wb,NB)	76.00±2.52	78.72±3.71	75.29±3.37	88.29±1.93	79.33±0.89
FSS(wf,TM)	80.62±2.18	82.57±1.95	83.33±2.65	85.79±2.52	81.38±2.56
FSS(wb,TM)	76.00±2.84	78.04±2.18	75.46±2.69	83.21±2.93	80.67±0.89

Table 30: Hepatitis Accuracies

6.9 Anneal

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	family	10	3	nominal
#1	product-type	3	1	nominal
#2	steel	9	8	nominal
#3	carbon	-	10	continuous
#4	hardness	-	7	continuous
#5	temper-rolling	2	2	nominal
#6	condition	4	3	nominal
#7	formability	6	5	nominal
#8	strength	-	8	continuous
#9	non-ageing	2	2	nominal
#10	surface-finish	3	2	nominal
#11	surface-quality	5	5	nominal
#12	enamelability	6	3	nominal
#13	bc	2	2	nominal
#14	bf	2	2	nominal
#15	bt	2	2	nominal
#16	bw/me	3	3	nominal
#17	bl	2	2	nominal
#18	m	2	1	nominal
#19	chrom	2	2	nominal
#20	phos	2	2	nominal
#21	cbond	2	2	nominal
#22	marvi	2	1	nominal
#23	exptl	2	2	nominal
#24	ferro	2	2	nominal
#25	corr	2	1	nominal
#26	blue/bright/varn/clean	5	4	nominal
#27	lustre	2	2	nominal
#28	jurofm	2	1	nominal
#29	s	2	1	nominal
#30	p	2	1	nominal
#31	shape	2	2	nominal
#32	thick	-	50	continuous
#33	width	-	68	continuous

continued on next page

continued from previous page

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#34	len	-	24	continuous
#35	oil	3	3	nominal
#36	bore	4	3	nominal
#37	packing	4	3	nominal

Table 31: Anneal Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	0 2 3 4 6 7 8 12 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 32 33 35 36 37	28	73.68%	3721.00
FSS(wb,C4.5)	0 1 3 4 5 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 34 35 36 37	35	92.11%	3707.40
FSS(wf,CN2)	1 2 3 4 7 8 10 11 13 14 15 17 18 19 20 21 22 23 24 25 31 32 34	23	60.53%	87607.70
FSS(wb,CN2)	1 2 3 4 5 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 34 35 36 37	34	89.47%	71581.70
FSS(wf,IB)	0 2 4 6 7 8 10 11 12 19 24 32	12	31.58%	4138.50
FSS(wb,IB)	0 2 3 4 5 6 8 11 12 14 16 20 23 27 31 32 33 35	18	47.37%	6092.60
FSS(wf,NB)	0 2 10 12 15 19 20 23 24 31 33	11	28.95%	187.10
FSS(wb,NB)	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 27 28 29 30 32 35 36 37	34	89.47%	263.00
FSS(wf,TM)	0 1 3 4 5 8 10 11 12 13 16 17 18 19 20 22 23 24 25 28 29 30 35 36 37	25	65.79%	565.50
FSS(wb,TM)	0 1 2 3 4 5 6 8 9 10 11 12 13 14 18 19 20 21 22 23 24 25 28 29 30	25	65.79%	364.70

Table 32: Anneal Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0	•	•			•	•	•	•	•	•
#1		•		•				•	•	•
#2	•		•	•	•	•	•			•
#3	•	•	•	•		•			•	•
#4	•		•	•	•			•	•	•
#5		•		•		•		•	•	•
#6	•				•	•		•		•
#7	•	•	•	•	•			•		
#8	•	•	•	•	•	•			•	•
#9		•		•				•		•
#10		•	•	•	•		•	•	•	•
#11		•	•	•	•	•		•	•	•
#12	•	•			•	•	•	•	•	•
#13		•	•	•				•	•	•
#14		•	•	•		•		•		•
#15		•	•	•			•	•		
#16	•	•		•		•		•	•	
#17	•	•	•	•				•	•	
#18	•	•	•	•				•	•	
#19	•	•	•	•	•		•	•	•	•
#20	•	•	•	•		•	•	•	•	•
#21	•	•	•	•				•	•	•
#22	•	•	•	•				•	•	•
#23	•	•	•	•		•	•	•	•	•
#24	•	•	•	•	•		•	•	•	•
#25	•	•	•	•				•	•	•
#26	•	•		•						
#27	•	•		•		•		•		
#28	•	•		•				•	•	•
#29	•	•		•				•	•	•
#30	•	•		•				•	•	•
#31		•	•	•		•	•			
#32	•	•	•	•	•			•		
#33	•					•	•			
#34		•	•	•						
#35	•	•		•		•		•	•	
#36	•	•		•				•	•	
#37	•	•		•				•	•	
Total 38	28	35	23	34	12	18	11	34	25	25
100%	73.68%	92.11%	60.53%	89.47%	31.58%	47.37%	28.95%	89.47%	65.79%	65.79%

Table 33: Anneal Wrapper Selected Features

anneal 10-cv	$\mathcal{C}4.5$	$\mathcal{CN}2$	IB	NB	TM
all features	92.10±1.19	89.53±1.89	99.22±0.37	91.54±0.60	76.84±1.13
FSS(wf, $\mathcal{C}4.5$)	92.43±1.22	89.97±1.12	94.88±0.67	75.06±0.96	84.74±0.97
FSS(wb, $\mathcal{C}4.5$)	94.22±1.19	90.53±1.25	99.00±0.35	83.18±1.27	83.51±1.17
FSS(wf, $\mathcal{CN}2$)	90.54±1.65	97.57±0.49	98.78±0.20	87.75±0.82	86.63±1.13
FSS(wb, $\mathcal{CN}2$)	90.88±1.49	97.57±0.49	98.55±0.33	89.09±0.72	83.40±1.18
FSS(wf,IB)	91.32±1.26	84.42±1.40	99.78±0.15	87.75±0.91	97.55±0.37
FSS(wb,IB)	92.43±1.33	92.31±0.99	99.67±0.17	89.31±0.71	87.53±0.83
FSS(wf,NB)	78.62±1.24	79.40±1.33	90.42±0.88	88.76±1.00	87.31±0.77
FSS(wb,NB)	91.77±1.07	88.86±0.99	99.22±0.24	91.87±0.68	94.09±1.08
FSS(wf,TM)	81.07±0.98	79.17±1.13	99.67±0.17	73.61±1.38	99.67±0.17
FSS(wb,TM)	83.07±0.98	82.85±2.03	99.78±0.15	87.63±0.96	99.56±0.18
anneal 10-strat-cv	$\mathcal{C}4.5$	$\mathcal{CN}2$	IB	NB	TM
all features	92.54±1.18	90.42±1.08	99.33±0.25	91.65±0.71	76.84±0.22
FSS(wf, $\mathcal{C}4.5$)	92.55±1.24	89.53±0.76	94.88±0.67	75.06±0.92	85.19±0.59
FSS(wb, $\mathcal{C}4.5$)	94.21±0.86	90.75±0.44	98.67±0.40	83.29±1.32	83.52±0.82
FSS(wf, $\mathcal{CN}2$)	91.65±1.09	96.44±0.94	98.78±0.31	87.75±0.96	86.41±1.18
FSS(wb, $\mathcal{CN}2$)	91.77±1.07	96.44±0.94	98.44±0.44	89.09±0.87	83.40±0.86
FSS(wf,IB)	91.32±1.14	83.63±0.77	99.78±0.15	87.75±0.77	97.55±0.54
FSS(wb,IB)	91.10±1.21	90.18±1.29	99.67±0.17	89.31±0.65	87.53±0.72
FSS(wf,NB)	78.51±0.59	79.30±0.45	90.31±0.58	88.42±0.60	87.31±0.78
FSS(wb,NB)	91.55±1.29	87.44±0.97	99.11±0.22	91.87±0.66	94.32±0.57
FSS(wf,TM)	81.29±0.39	79.29±0.42	99.67±0.24	73.60±1.25	99.67±0.17
FSS(wb,TM)	83.07±0.43	81.97±1.26	99.78±0.15	87.41±0.84	99.56±0.25

Table 34: Anneal Accuracies

6.10 Sonar

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	a00	-	177	continuous
#1	a01	-	182	continuous
#2	a02	-	190	continuous
#3	a03	-	181	continuous
#4	a04	-	193	continuous
#5	a05	-	196	continuous
#6	a06	-	195	continuous
#7	a07	-	201	continuous
#8	a08	-	205	continuous
#9	a09	-	207	continuous
#10	a10	-	203	continuous
#11	a11	-	206	continuous
#12	a12	-	198	continuous
#13	a13	-	202	continuous
#14	a14	-	203	continuous
#15	a15	-	203	continuous
#16	a16	-	202	continuous
#17	a17	-	204	continuous
#18	a18	-	206	continuous
#19	a19	-	203	continuous
#20	a20	-	200	continuous
#21	a21	-	203	continuous
#22	a22	-	199	continuous
#23	a23	-	201	continuous
#24	a24	-	198	continuous
#25	a25	-	194	continuous
#26	a26	-	190	continuous
#27	a27	-	194	continuous
#28	a28	-	197	continuous
#29	a29	-	202	continuous
#30	a30	-	207	continuous
#31	a31	-	205	continuous
#32	a32	-	205	continuous
#33	a33	-	206	continuous

continued on next page

continued from previous page

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#34	a34	-	205	continuous
#35	a35	-	205	continuous
#36	a36	-	206	continuous
#37	a37	-	206	continuous
#38	a38	-	204	continuous
#39	a39	-	206	continuous
#40	a40	-	204	continuous
#41	a41	-	208	continuous
#42	a42	-	205	continuous
#43	a43	-	196	continuous
#44	a44	-	205	continuous
#45	a45	-	199	continuous
#46	a46	-	202	continuous
#47	a47	-	204	continuous
#48	a48	-	193	continuous
#49	a49	-	154	continuous
#50	a50	-	160	continuous
#51	a51	-	144	continuous
#52	a52	-	134	continuous
#53	a53	-	134	continuous
#54	a54	-	129	continuous
#55	a55	-	122	continuous
#56	a56	-	121	continuous
#57	a57	-	124	continuous
#58	a58	-	119	continuous
#59	a59	-	109	continuous

Table 35: Sonar Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	3 10 30 35 45 50 51	7	11.67%	569.20
FSS(wb,C4.5)	1 2 3 4 5 6 7 8 9 10 12 13 14 15 17 18 19 20 21 22 24 25 26 27 29 30 34 36 37 39 40 41 42 43 44 49 50 54 56 57	40	66.67%	3968.40
FSS(wf,CN2)	0 10 11 15 22 25 26 29 33 39 45 57	12	20.00%	5726.90
FSS(wb,CN2)	0 1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 18 19 21 22 24 25 26 27 28 30 31 33 35 36 37 38 39 40 41 44 45 46 47 49 50 51 52 53 54 55 56 57 58 59	50	83.33%	28153.00
FSS(wf,IB)	0 3 4 5 7 8 9 11 13 15 29 30 31 34 37 40 41 42 45 52 57	21	35.00%	2161.30
FSS(wb,IB)	0 1 2 3 4 7 8 9 10 11 12 14 15 16 17 19 20 21 22 23 24 25 26 27 29 31 32 33 34 35 36 37 38 39 40 44 45 46 47 48 49 50 51 52 53 54 55 57 58	49	81.67%	3751.70
FSS(wf,NB)	5 10 11 15 27 31 35 39 48 52	10	16.67%	126.70
FSS(wb,NB)	1 3 4 5 6 7 8 9 11 13 14 15 16 17 19 20 21 22 23 24 26 27 28 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 50 52 53 54 55 56 57 58	50	83.33%	768.70
FSS(wf,TM)	58	1	1.67%	83.00
FSS(wb,TM)	0	1	1.67%	140.30

Table 36: Sonar Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0			•	•	•	•				•
#1		•		•		•				
#2		•		•		•				
#3	•	•		•	•	•			•	
#4		•		•	•	•				
#5		•		•	•		•		•	
#6		•		•					•	
#7		•		•	•	•			•	
#8		•		•	•	•			•	
#9		•		•	•	•			•	
#10	•	•	•	•	•	•	•		•	
#11		•	•	•	•	•	•	•		
#12		•		•	•	•				
#13		•		•	•				•	
#14		•		•	•	•			•	
#15		•	•	•	•	•	•	•	•	

continued on next page

continued from previous page

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#16				•		•			•	
#17		•				•			•	
#18		•		•						
#19		•		•		•			•	
#20		•				•			•	
#21		•		•		•			•	
#22		•	•	•		•			•	
#23						•			•	
#24		•		•		•			•	
#25		•	•	•		•			•	
#26		•	•	•		•			•	
#27		•		•		•	•		•	
#28				•					•	
#29		•	•		•	•			•	
#30	•	•		•	•				•	
#31				•	•	•	•		•	
#32						•			•	
#33			•	•		•			•	
#34		•			•	•			•	
#35	•			•		•	•		•	
#36		•		•		•			•	
#37		•		•	•	•			•	
#38				•		•			•	
#39		•	•	•		•	•		•	
#40		•		•	•	•			•	
#41		•		•	•				•	
#42		•			•				•	
#43		•							•	
#44		•		•		•			•	
#45	•		•		•				•	
#46				•		•			•	
#47				•		•			•	
#48						•	•		•	
#49		•		•		•			•	
#50	•	•		•		•			•	
#51	•			•		•			•	
#52				•	•	•	•		•	
#53				•		•			•	
#54		•		•		•			•	
#55				•		•			•	
#56		•		•					•	
#57		•	•	•	•	•			•	
#58				•		•			•	•
#59				•					•	
Total 60	7	40	12	50	21	49	10	50	1	1
100%	11.67%	66.67%	20.00%	83.33%	35.00%	81.67%	16.67%	83.33%	1.67%	1.67%

Table 37: Sonar Wrapper Selected Features

sonar 10-cv	C4.5	CN2	IB	NB	TM
all features	67.83±2.79	73.16±3.44	84.64±2.42	68.79±3.60	53.36±4.45
FSS(wf,C4.5)	82.24±2.68	70.67±3.69	72.19±2.57	66.90±2.53	53.36±4.45
FSS(wb,C4.5)	87.07±2.55	74.05±1.88	83.71±3.33	67.81±2.63	53.36±4.45
FSS(wf,CN2)	73.17±3.87	85.13±2.86	84.60±2.35	72.17±4.46	53.36±4.45
FSS(wb,CN2)	67.36±3.65	82.79±3.00	83.24±2.93	65.88±2.94	53.36±4.45
FSS(wf,IB)	75.52±2.67	73.57±3.01	93.81±1.75	66.83±3.04	53.36±4.45
FSS(wb,IB)	69.26±3.08	74.17±4.11	90.38±2.26	68.33±2.80	53.36±4.45
FSS(wf,NB)	76.45±2.96	75.54±3.33	85.57±3.41	82.67±1.93	53.36±4.45
FSS(wb,NB)	76.50±3.56	74.60±2.51	84.17±1.86	76.00±2.99	53.36±4.45
FSS(wf,TM)	46.69±3.64	61.07±1.90	57.17±2.48	52.45±3.29	58.14±2.42
FSS(wb,TM)	49.95±4.21	49.18±4.38	52.55±4.16	57.29±5.03	56.36±4.49
sonar 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	69.74±1.97	71.19±3.30	85.60±2.00	69.26±4.53	53.38±0.54
FSS(wf,C4.5)	83.19±2.30	75.98±3.00	72.19±2.37	68.36±2.59	53.38±0.54
FSS(wb,C4.5)	83.74±3.75	72.62±3.32	83.71±2.75	68.24±2.98	53.38±0.54
FSS(wf,CN2)	72.14±3.05	81.32±3.46	84.12±2.03	71.69±2.74	53.38±0.54
FSS(wb,CN2)	68.31±2.53	75.52±2.75	84.19±2.83	66.36±3.60	53.38±0.54
FSS(wf,IB)	74.98±3.73	70.28±3.28	92.81±1.27	67.33±2.19	53.38±0.54
FSS(wb,IB)	71.64±1.47	72.10±3.01	89.43±1.99	67.81±2.82	53.38±0.54
FSS(wf,NB)	74.93±2.33	73.98±2.84	84.57±2.77	78.74±3.24	53.38±0.54
FSS(wb,NB)	73.60±2.14	76.02±3.14	85.12±1.64	74.07±3.61	53.38±0.54
FSS(wf,TM)	51.00±1.21	61.50±1.35	55.24±2.35	53.88±2.50	56.24±1.51
FSS(wb,TM)	53.90±2.70	46.75±3.43	52.52±3.51	57.74±3.19	56.36±2.45

Table 38: Sonar Accuracies

6.11 Genetics

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	A1	8	5	nominal
#1	A2	8	5	nominal
#2	A3	8	4	nominal
#3	A4	8	4	nominal
#4	A5	8	4	nominal
#5	A6	8	4	nominal
#6	A7	8	4	nominal
#7	A8	8	4	nominal
#8	A9	8	4	nominal
#9	A10	8	4	nominal
#10	A11	8	4	nominal
#11	A12	8	4	nominal
#12	A13	8	4	nominal
#13	A14	8	5	nominal
#14	A15	8	4	nominal
#15	A16	8	4	nominal
#16	A17	8	4	nominal
#17	A18	8	4	nominal
#18	A19	8	5	nominal
#19	A20	8	5	nominal
#20	A21	8	5	nominal
#21	A22	8	5	nominal
#22	A23	8	5	nominal
#23	A24	8	5	nominal
#24	A25	8	5	nominal
#25	A26	8	5	nominal
#26	A27	8	5	nominal
#27	A28	8	5	nominal
#28	A29	8	5	nominal
#29	A30	8	5	nominal
#30	A31	8	5	nominal
#31	A32	8	5	nominal
#32	A33	8	5	nominal
#33	A34	8	5	nominal
#34	A35	8	6	nominal
#35	A36	8	6	nominal
#36	A37	8	5	nominal
#37	A38	8	5	nominal
#38	A39	8	5	nominal
#39	A40	8	5	nominal
#40	A41	8	5	nominal
#41	A42	8	5	nominal
#42	A43	8	5	nominal
#43	A44	8	5	nominal
#44	A45	8	5	nominal
#45	A46	8	5	nominal
#46	A47	8	5	nominal
#47	A48	8	5	nominal
#48	A49	8	5	nominal
#49	A50	8	5	nominal
#50	A51	8	5	nominal
#51	A52	8	5	nominal
#52	A53	8	5	nominal
#53	A54	8	5	nominal
#54	A55	8	5	nominal
#55	A56	8	5	nominal
#56	A57	8	5	nominal
#57	A58	8	5	nominal
#58	A59	8	5	nominal
#59	A60	8	5	nominal

Table 39: Genetics Feature Description

Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	15 17 23 27 28 29 30 31 33 34 35 45 48 59	14	23,33%	8546.3
FSS(wb,C4.5)	0 1 2 3 4 6 7 8 9 10 11 12 13 14 15 18 21 23 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 48 49 50 51 52 54 56 57 58 59	49	81,67%	24841.0
FSS(wf,CN2)	6 28 29 30 31	5	8,33%	42479.4
FSS(wb,CN2)	0 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 28 29 30 31 32 33 34 35 36 38 39 42 43 45 46 47 49 50 51 52 54 55 56 59	48	80,00%	550329.7
FSS(wf,IB)	27 28 29 30 31	5	8,33%	36520.4
FSS(wb,IB)	0 1 2 3 5 6 7 9 10 12 13 16 17 18 20 21 22 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 42 43 46 48 50 52 58 59	42	70,00%	638657.0
FSS(wf,NB)	0 6 7 11 12 13 17 18 20 22 24 25 27 28 29 30 31 33 34 38 39 43 49 50 59	25	41,67%	2279.7
FSS(wb,NB)	0 2 3 4 5 6 7 9 10 11 12 13 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 40 42 44 45 46 47 49 50 51 52 53 54 55 57 58 59	52	86,67%	2026.6
FSS(wf,TM)	27 28 29 30 31 34	6	10,00%	1627.1
FSS(wb,TM)	27 28 29 30 31 34	6	10,00%	4668.4

Table 40: Genetics Time for Selecting Features

Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0		•		•		•	•	•		
#1		•				•				
#2		•				•				
#3		•		•		•		•		
#4		•		•				•		
#5				•		•		•		
#6		•	•	•		•	•	•		
#7		•		•		•	•	•		
#8		•		•				•		
#9		•		•		•		•		
#10		•		•		•		•		
#11		•		•			•	•		
#12		•				•	•	•		
#13		•		•		•	•	•		
#14		•		•				•		
#15	•	•		•				•		
#16				•		•		•		
#17	•			•		•	•	•		
#18		•		•		•	•	•		
#19				•				•		
#20				•		•	•	•		
#21		•		•		•		•		
#22				•		•	•	•		
#23	•	•		•		•	•	•		
#24				•		•	•	•		
#25				•		•	•	•		
#26		•		•		•		•		
#27	•	•			•	•	•	•	•	•
#28	•	•	•	•	•	•	•	•	•	•
#29	•	•	•	•	•	•	•	•	•	•
#30	•	•	•	•	•	•	•	•	•	•
#31	•	•	•	•	•	•	•	•	•	•
#32		•		•		•		•		•
#33	•	•		•		•	•	•		
#34	•	•		•		•	•	•	•	•
#35	•	•		•		•		•		•
#36		•		•		•		•		
#37		•				•		•		
#38		•		•		•	•	•		
#39		•		•		•	•	•		
#40		•						•		
#41		•						•		
#42		•		•		•		•		
#43		•		•		•	•	•		
#44		•						•		
#45	•	•		•				•		
#46		•		•		•		•		
#47		•		•				•		
#48	•	•				•		•		
#49		•		•			•	•		
#50		•		•		•	•	•		
#51		•		•				•		
#52		•		•		•		•		
#53								•		
#54		•		•				•		
#55		•		•				•		
#56		•		•				•		
#57		•						•		
#58		•				•		•		

continued on next page

<i>continued from previous page</i>										
Feature Number	FSS									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#59	•	•	•	•	•	•	•	•	•	•
Total 60	14	49	5	48	5	42	25	52	6	6
100%	23.33%	81.67%	8.33%	80.00%	8.33%	70.00%	41.67%	86.67%	10.00%	10.00%

Table 41: Genetics Wrapper Selected Features

genetics 10-cv	C4.5	CN2	IB	NB	TM
all features	94.08±0.52	76.39±1.65	78.78±0.53	95.52±0.45	60.82±0.98
FSS(wf,C4.5)	94.61±0.45	82.91±2.16	86.08±0.61	94.61±0.35	64.01±1.12
FSS(wb,C4.5)	93.92±0.45	78.35±1.83	76.90±0.81	94.67±0.53	60.94±0.96
FSS(wf,CN2)	89.66±0.60	89.39±0.66	88.59±0.72	89.03±0.72	89.31±0.65
FSS(wb,CN2)	93.17±0.53	85.74±1.08	79.47±0.70	95.55±0.45	60.94±1.02
FSS(wf,IB)	91.07±0.60	84.13±2.68	90.91±0.54	90.72±0.52	91.63±0.54
FSS(wb,IB)	94.20±0.47	78.36±1.77	83.35±0.70	95.71±0.47	61.35±1.04
FSS(wf,NB)	94.08±0.54	82.79±1.97	82.29±0.63	96.43±0.27	62.82±1.16
FSS(wb,NB)	94.29±0.43	76.97±1.89	79.22±0.74	96.21±0.35	60.91±1.00
FSS(wf,TM)	93.54±0.55	80.24±2.71	90.34±0.57	93.54±0.52	92.95±0.47
FSS(wb,TM)	94.08±0.42	80.24±2.71	90.34±0.57	93.54±0.52	92.95±0.47
genetics 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	94.17±0.39	79.53±1.45	78.75±0.71	95.45±0.32	60.94±0.28
FSS(wf,C4.5)	94.67±0.30	81.54±2.01	86.61±0.54	94.51±0.31	64.23±0.48
FSS(wb,C4.5)	94.17±0.39	76.58±1.77	77.18±0.59	94.67±0.40	61.07±0.25
FSS(wf,CN2)	89.69±0.50	89.47±0.53	88.71±0.63	89.09±0.58	89.44±0.55
FSS(wb,CN2)	93.17±0.53	83.26±1.73	79.78±0.64	95.55±0.35	61.13±0.32
FSS(wf,IB)	91.07±0.48	85.32±2.64	90.88±0.62	90.75±0.55	91.69±0.55
FSS(wb,IB)	94.23±0.31	75.39±1.65	83.13±0.68	95.71±0.34	61.47±0.34
FSS(wf,NB)	93.95±0.50	85.58±2.14	82.60±0.50	96.21±0.26	62.95±0.49
FSS(wb,NB)	94.29±0.33	77.93±1.61	79.56±0.43	96.11±0.23	61.03±0.26
FSS(wf,TM)	93.51±0.46	81.98±2.68	90.38±0.63	93.61±0.54	92.95±0.52
FSS(wb,TM)	94.08±0.44	81.98±2.68	90.38±0.63	93.61±0.54	92.95±0.52

Table 42: Genetics Accuracies

6.12 DNA

Feature Number	Feature Name	#Distinct Values		
		possible	actual	type
#0	a1	2	2	nominal
#1	a2	2	2	nominal
#2	a3	2	2	nominal
#3	a4	2	2	nominal
⋮	⋮	⋮	⋮	⋮
#179	a180	2	2	nominal

Table 43: DNA Feature Description

FSS Inducer	Selected Features	#F	%F	Time (s)
FSS(wf,C4.5)	62 64 71 81 82 83 84 89 92 93 94 95 96 99 104 108 117 119 127 143 146 150 153 165 167	25	13.89%	47896.5
FSS(wb,C4.5)	0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 20 21 22 26 27 28 31 33 35 37 39 40 41 43 44 45 47 50 51 52 55 56 58 61 64 65 68 69 71 72 73 74 75 78 79 81 82 83 84 85 86 87 88 89 91 92 93 94 95 96 99 100 101 103 104 105 106 107 109 110 111 113 115 116 117 118 119 120 121 122 123 127 128 129 130 132 133 134 135 140 145 150 151 155 156 158 159 166 172 174	107	59.44%	275187.0
FSS(wf,CN2)	41 54 63 81 83 84 85 89 92 93 94 95 96 99 104	15	8.33%	349135.5
FSS(wb,CN2)	N/A	N/A	N/A	N/A
FSS(wf,IB)	81 83 84 89 92 93 94 95 96 104	10	5.56%	225645.4
FSS(wb,IB)	N/A	N/A	N/A	N/A

continued on next page

<i>continued from previous page</i>												
FSS Inducer	Selected Features									#F	%F	Time (s)
FSS(wf,NB)	16 18 23 39 45 48 51 53 54 56 57 59 60 62 63 69 71 72 74 75 77 80 81 83 84 89 92 93 94 95 96 97 99 104 118 119 127 133 157 179									40	22,22%	7966.5
FSS(wb,NB)	1 3 4 5 6 8 9 10 11 12 13 14 16 17 18 20 21 22 23 25 26 27 28 29 30 31 32 33 34 35 37 38 39 40 41 44 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 104 105 106 107 108 109 110 111 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179											
FSS(wf,TM)	81 83 84 89 92 93 94 95 96 104									166	92,22%	18773.6
FSS(wb,TM)	81 83 84 89 92 93 94 95 96 104									10	5,56%	32123.2

Table 44: DNA Time for Selecting Features

Feature	FSS Inducer									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#0		•								
#1		•						•		
#2		•								
#3		•						•		
#4										
#5		•						•		
#6		•						•		
#7		•								
#8		•						•		
#9		•						•		
#10		•						•		
#11		•						•		
#12		•						•		
#13		•						•		
#14		•						•		
#15		•						•		
#16		•					•	•		
#17		•						•		
#18							•	•		
#19										
#20		•						•		
#21		•						•		
#22		•						•		
#23							•	•		
#24										
#25								•		
#26		•						•		
#27		•						•		
#28		•						•		
#29								•		
#30								•		
#31		•						•		
#32								•		
#33		•						•		
#34								•		
#35		•						•		
#36										
#37		•						•		
#38								•		
#39		•					•	•		
#40		•						•		
#41		•	•					•		
#42										
#43		•						•		
#44		•						•		
#45		•					•	•		
#46								•		
#47		•						•		
#48							•	•		
#49								•		
#50		•						•		
#51		•					•	•		
#52		•						•		
#53							•	•		
#54			•				•	•		
#55		•						•		
#56		•					•	•		
#57							•	•		
#58		•						•		
#59							•	•		
#60							•	•		
#61		•						•		

continued on next page

continued from previous page

Feature	FSS Inducer									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#62	•						•	•		
#63			•				•	•		
#64	•	•						•		
#65		•						•		
#66								•		
#67								•		
#68		•						•		
#69		•					•	•		
#70								•		
#71	•	•					•	•		
#72		•					•	•		
#73		•						•		
#74		•					•	•		
#75		•					•	•		
#76								•		
#77							•	•		
#78		•						•		
#79		•						•		
#80							•	•		
#81	•	•	•		•		•	•	•	•
#82	•	•						•		
#83	•	•	•		•		•	•	•	•
#84	•	•	•		•		•	•	•	•
#85		•	•					•		
#86		•						•		
#87		•						•		
#88		•						•		
#89	•	•	•		•		•	•	•	•
#90								•		
#91		•						•		
#92	•	•	•		•		•	•	•	•
#93	•	•	•		•		•	•	•	•
#94	•	•	•		•		•	•	•	•
#95	•	•	•		•		•	•	•	•
#96	•	•	•		•		•	•	•	•
#97							•	•		
#98								•		
#99	•	•	•				•	•		
#100		•						•		
#101		•						•		
#102								•		
#103		•						•		
#104	•	•	•		•		•	•	•	•
#105		•						•		
#106		•						•		
#107		•						•		
#108	•							•		
#109		•						•		
#110		•						•		
#111		•						•		
#112								•		
#113		•						•		
#114								•		
#115		•						•		
#116		•						•		
#117	•	•						•		
#118		•					•	•		
#119	•	•					•	•		
#120		•						•		
#121		•						•		
#122		•						•		
#123		•						•		
#124								•		
#125								•		
#126								•		
#127	•	•					•	•		
#128		•						•		
#129		•						•		
#130		•						•		
#131								•		
#132		•						•		
#133		•					•	•		
#134		•						•		
#135		•						•		
#136								•		
#137								•		
#138								•		
#139								•		
#140		•						•		
#141								•		
#142								•		
#143	•							•		
#144								•		
#145		•						•		
#146	•							•		
#147								•		

continued on next page

continued from previous page

Feature	FSS Inducer									
	(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
#148								•		
#149								•		
#150	•	•						•		
#151		•						•		
#152								•		
#153	•							•		
#154								•		
#155		•						•		
#156		•						•		
#157							•	•		
#158		•						•		
#159		•						•		
#160								•		
#161								•		
#162								•		
#163								•		
#164								•		
#165	•							•		
#166		•						•		
#167	•							•		
#168								•		
#169								•		
#170								•		
#171								•		
#172		•						•		
#173								•		
#174		•						•		
#175								•		
#176								•		
#177								•		
#178								•		
#179							•	•		
Total 180	25	107	15	N/A	10	N/A	40	166	10	10
100%	13.89%	59.44%	8.33%	N/A	5.56%	N/A	22.22%	92.22%	5.56%	5.56%

Table 45: DNA Wrapper Selected Features

dna 10-cv	C4.5	CN2	IB	NB	TM
all features	92.50±0.63	88.20±0.80	73.89±0.49	94.07±0.32	61.05±0.65
FSS(wf,C4.5)	95.70±0.41	92.18±0.61	88.26±0.59	94.32±0.48	66.20±0.85
FSS(wb,C4.5)	95.07±0.34	89.64±0.63	76.02±0.67	93.75±0.50	61.49±0.63
FSS(wf,CN2)	94.32±0.44	95.24±0.37	92.62±0.46	94.82±0.52	91.34±0.40
FSS(wb,CN2)	N/A	N/A	N/A	N/A	N/A
FSS(wf,IB)	94.07±0.41	93.85±0.44	93.82±0.37	94.16±0.50	94.32±0.48
FSS(wb,IB)	N/A	N/A	N/A	N/A	N/A
FSS(wf,NB)	93.82±0.52	91.88±0.56	84.93±0.53	96.92±0.23	64.15±0.73
FSS(wb,NB)	92.59±0.59	88.06±0.59	74.17±0.46	95.42±0.38	61.11±0.66
FSS(wf,TM)	94.88±0.37	94.71±0.36	94.07±0.32	94.22±0.41	94.73±0.39
FSS(wb,TM)	94.88±0.37	94.71±0.36	94.07±0.32	94.22±0.41	94.73±0.39
dna 10-strat-cv	C4.5	CN2	IB	NB	TM
all features	92.40±0.46	88.15±0.62	74.23±0.44	94.04±0.34	61.24±0.47
FSS(wf,C4.5)	95.45±0.38	91.94±0.61	88.23±0.58	94.35±0.44	66.32±0.55
FSS(wb,C4.5)	94.51±0.39	88.63±0.71	76.18±0.78	93.97±0.49	61.67±0.50
FSS(wf,CN2)	94.38±0.38	94.89±0.41	92.62±0.45	94.88±0.44	91.40±0.38
FSS(wb,CN2)	N/A	N/A	N/A	N/A	N/A
FSS(wf,IB)	93.97±0.34	93.80±0.47	93.72±0.34	94.13±0.51	94.29±0.46
FSS(wb,IB)	N/A	N/A	N/A	N/A	N/A
FSS(wf,NB)	93.82±0.35	91.91±0.53	85.31±0.43	96.83±0.23	64.34±0.69
FSS(wb,NB)	92.40±0.45	89.14±0.62	74.07±0.39	95.29±0.43	61.30±0.48
FSS(wf,TM)	94.85±0.35	94.76±0.35	94.07±0.31	94.22±0.37	94.73±0.35
FSS(wb,TM)	94.85±0.35	94.76±0.35	94.07±0.31	94.22±0.37	94.73±0.35

Table 46: DNA Accuracies

7 Results Comparison

The following sections show some tables, presenting a summary and comparing results. Some graphs, representing the numerical data in tables, are also shown.

In what follows it should be observed that results for dna using $\mathcal{CN}2$ and IB inducers with backward selection were not available, after running longer than 40 days in background. These results are reported as “N/A” in the tables and they are not considered on average calculations.

7.1 Number of Selected Features

Tables 47 and 48 show the number and the proportion of selected features for each dataset, respectively.

In all cases, not only on average, the number of selected features using backward selection is always greater or equal than using forward selection, *i.e.* $\#FSS(wb,inducer) \geq \#FSS(wf,inducer)$. On average, forward selection picks up 39.35% features against 62.36% on backward selection, an increasing factor of 58.48%. This result seems to confirm the idea that going backwards from the full set of features would favor to capture interactive features.

Dataset	#F	FSS									
		(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
bupa	6	5	5	5	5	5	5	5	5	2	2
pima	8	5	5	7	7	3	5	5	5	1	1
breast cancer	10	7	7	5	9	6	6	5	7	1	2
hungaria	13	5	8	4	7	2	9	8	9	3	3
crx	15	7	11	5	7	5	13	6	7	5	3
letter	16	11	11	9	9	11	11	12	12	4	4
hepatitis	19	5	7	7	16	5	15	9	11	3	10
anneal	38	28	35	23	34	12	18	11	34	25	25
sonar	60	7	40	12	50	21	49	10	50	1	1
genetics	60	14	49	5	48	5	42	25	52	6	6
dna	180	25	107	15	N/A	10	N/A	40	166	10	10

Table 47: Number of Selected Features

Dataset	#F	FSS									
		(wf,C4.5)	(wb,C4.5)	(wf,CN2)	(wb,CN2)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
bupa	6	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	33.33%	33.33%
pima	8	62.50%	62.50%	87.50%	87.50%	37.50%	62.50%	62.50%	62.50%	12.50%	12.50%
breast cancer	10	70.00%	70.00%	50.00%	90.00%	60.00%	60.00%	50.00%	70.00%	10.00%	20.00%
hungaria	13	38.46%	61.54%	30.77%	53.85%	15.38%	69.23%	61.54%	69.23%	23.08%	23.08%
crx	15	46.67%	73.33%	33.33%	46.67%	33.33%	86.67%	40.00%	46.67%	33.33%	20.00%
letter	16	68.75%	68.75%	56.25%	56.25%	68.75%	68.75%	75.00%	75.00%	25.00%	25.00%
hepatitis	19	26.32%	36.84%	36.84%	84.21%	26.32%	78.95%	47.37%	57.89%	15.79%	52.63%
anneal	38	73.68%	92.11%	60.53%	89.47%	31.58%	47.37%	28.95%	89.47%	65.79%	65.79%
sonar	60	11.67%	66.67%	20.00%	83.33%	35.00%	81.67%	16.67%	83.33%	1.67%	1.67%
genetics	60	23.33%	81.67%	8.33%	80.00%	8.33%	70.00%	41.67%	86.67%	10.00%	10.00%
dna	180	13.89%	59.44%	8.33%	N/A	5.56%	N/A	22.22%	92.22%	5.56%	5.56%
Average	38.64	47.15%	68.74%	43.20%	75.46%	36.83%	70.85%	48.11%	74.21%	21.46%	24.50%

Table 48: Proportion of Selected Features

It should be observed that in our experiments, for some datasets, a larger number of features than the ones needed to reach the halting criterion have been selected by the wrapper. This is due to the fact that we have used $MCC++$ FSS wrapper default parameter setting, which is zero for the *complexity penalty* parameter. This parameter allows penalizing feature subsets with many features such that, if the accuracy of two feature subsets is the same, the subset with small number of features is preferred. For instance, it should be observed that there are seven nominal features assuming just one value on the anneal dataset. All of them were selected by backward and six by forward selection for the C4.5 inducer even they cause no improvement in the accuracy. The other inducers have also selected more than one of these features. However, at most one of those features would be selected if a complexity penalty factor greater than zero was used.

7.2 Time for Selecting Features

Time taken, in seconds, for selecting features is shown in Table 49, related to a standard *Indigo 2* Silicon Graphics workstation. It is expected that forward should be less expensive than backward selection, since building classifiers when there are few features in the training data should be computationally faster. Also, Table 50 shows the time for running ten-fold cross-validation and ten-fold stratified-cross-validation using all features, *i.e.* the standard inducer.

Although not true for each individual case, the overall picture indicates that backward is about 5 times slower than forward selection. In any case, the wrapper approach is very slow. The overall time taken for running the experiments was almost 32 days of uninterrupted CPU processing time for all inducers and datasets (not considering FSS(wb, $\mathcal{CN}2$) and FSS(wb,IB) for dna). On average, the wrapper using NB was the fastest inducer followed by $\mathcal{C}4.5$, IB and $\mathcal{CN}2$, respectively.

Also it is possible to note that, for most of datasets, backward selection took more time than forward selection for IB and NB but not for $\mathcal{C}4.5$ and $\mathcal{CN}2$. Observe that only $\mathcal{C}4.5$ and $\mathcal{CN}2$ are able to select features by their own (embedded FSS), besides dealing with unknown values.

Dataset	#F	FSS									
		(wf, $\mathcal{C}4.5$)	(wb, $\mathcal{C}4.5$)	(wf, $\mathcal{CN}2$)	(wb, $\mathcal{CN}2$)	(wf,IB)	(wb,IB)	(wf,NB)	(wb,NB)	(wf,TM)	(wb,TM)
bupa	6	28.7	23.7	189.7	164.1	80.1	47.7	8.8	6.5	12.8	14.7
pima	8	81.9	89.2	1292.1	790.7	335.1	357.0	27.2	46.4	28.3	68.5
breast cancer	10	135.8	116.7	697.7	564.3	537.1	566.7	33.2	37.7	36.7	61.9
hungaria	13	83.6	104.8	314.2	1242.9	112.2	185.0	23.1	20.7	22.2	40.4
crx	15	416.5	324.8	464.4	3628.7	353.8	544.7	46.4	67.4	58.4	158.3
letter	16	2003.5	1167.9	33446.1	68115.1	53490.4	74728.6	784.2	865.3	2033.1	929.7
hepatitis	19	77.2	149.6	700.4	583.0	112.6	136.9	17.4	26.0	16.0	25.0
anneal	38	3721.0	3707.4	87607.7	71581.7	4138.5	6092.6	187.1	263.0	565.5	364.7
sonar	60	569.2	3968.4	5726.9	28153.0	2161.3	3751.7	126.7	768.7	83.0	140.3
genetics	60	8546.3	24841.0	42479.4	550329.7	36520.4	638657.0	2279.7	2026.6	1627.1	4668.4
dna	180	47896.5	275187.0	349135.5	N/A	225645.4	N/A	7966.5	18773.6	11012.2	32123.2
Average ⁴	38.64	1566.4	3449.4	17291.9	72515.3	9784.2	72506.8	353.4	412.8	448.3	647.2

Table 49: Time (in seconds) for Selecting Features

Dataset	C.45	$\mathcal{CN}2$	IB	NB	TM
10-cv					
bupa	1.6	8.1	2.0	0.3	0.3
pima	4.2	26.0	8.7	0.6	0.7
breast cancer	3.7	14.8	8.6	0.7	0.6
hungaria	2.0	12.2	2.3	0.4	0.4
crx	4.3	28.2	9.7	0.7	0.8
letter	219.5	8710.5	5223.4	32.5	15.7
hepatitis	1.2	5.0	1.1	0.3	0.3
anneal	10.1	168.7	31.7	1.5	1.8
sonar	9.5	38.2	4.8	1.1	0.7
genetics	37.9	831.0	502.7	7.8	10.0
dna	119.4	3086.6	1434.6	21.0	25.0
10-strat-cv					
bupa	1.7	8.1	2.0	0.4	0.7
pima	4.3	24.6	8.9	0.8	1.7
breast cancer	3.8	14.8	8.7	0.9	1.6
hungaria	2.1	12.7	2.4	0.4	0.8
crx	4.5	28.7	10.1	0.9	2.0
letter	230.2	8708.0	5209.9	38.6	43.2
hepatitis	1.2	5.1	1.1	0.3	0.5
anneal	10.2	173.7	32.2	1.9	3.8
sonar	9.7	39.2	4.9	1.1	1.2
genetics	38.9	820.3	513.3	9.7	19.2
dna	119.9	3107.6	1470.9	24.8	45.1

Table 50: Time for Running Ten-Fold Cross-Validation and Ten-Fold Stratified Cross-Validation Using all Features

⁴None of dna results were considered on average calculations.

7.3 Comparing No FSS, Forward and Backward Wrapper FSS

Looking to the numbers is not easy to understand if one algorithm is better than other one. So, to compare a pair of algorithms, the mean and standard deviation using ten fold stratified cross validation are used. To determine whether the difference between two algorithms — say A_1 and A_2 — is significant or not, we show one graph with three bars. Each bar corresponds to the mean accuracy divided by the standard deviation. When the length of the bars are higher than two, the results are significant at 95% confidence level.

Comparison are made such that A_2 is the algorithm proposed (such as $\mathcal{C4.5}$ using only the wrapper selected features) and A_1 is the standard algorithm (such as $\mathcal{C4.5}$ using all features). When the bar is above zero it means that A_2 outperforms A_1 — meaning in the above example that the wrapper did improve the accuracy of the standard algorithm.

For each dataset, the combined mean $m(A_2 - A_1)$ and standard deviation $sd(A_2 - A_1)$ are calculated according with Equations 1 and 2, respectively. The absolute difference in standard deviations is given by Equation 3.

$$m(A_2 - A_1) = m(A_2) - m(A_1) \quad (1)$$

$$sd(A_2 - A_1) = \sqrt{\frac{sd(A_2)^2 + sd(A_1)^2}{2}} \quad (2)$$

$$ad(A_2 - A_1) = \frac{m(A_2 - A_1)}{sd(A_2 - A_1)} \quad (3)$$

Table 51 shows the results obtained by Equation 3, for each accuracy inducer using no feature selection (*inducer*), forward (FSS(wf,*inducer*)) and backward (FSS(wb,*inducer*)) wrapper selected features for the same inducer (black box wrapper inducer equals accuracy estimator inducer).

Graphs from Table 51 are shown in Figures 5, 6, 7, 8 and 9 for each *inducer*. For each dataset, the first bar in the graph corresponds to the comparison of wrapper forward feature selection against no feature selection. The second one corresponds to the comparison of wrapper backward feature selection against no feature selection. The last bar compares backward against forward feature selection.

Dataset	$\mathcal{C4.5}$	FSS(wf, $\mathcal{C4.5}$)	FSS(wb, $\mathcal{C4.5}$)	FSS(wf, $\mathcal{C4.5}$) - $\mathcal{C4.5}$	FSS(wb, $\mathcal{C4.5}$) - $\mathcal{C4.5}$	FSS(wb, $\mathcal{C4.5}$) -FSS(wf, $\mathcal{C4.5}$)
bupa	68.71±1.73	66.97±2.76	66.97±2.76	-0.76	-0.76	0.00
pima	74.26±1.13	74.77±1.04	75.95±0.98	0.47	1.60	1.17
breast cancer	94.28±0.56	95.00±0.83	95.00±0.83	1.02	1.02	0.00
hungaria	77.52±4.20	82.97±3.27	82.97±3.27	1.45	1.45	0.00
crx	84.35±1.18	85.65±1.17	86.67±1.26	1.11	1.90	0.84
letter	86.59±0.34	86.93±0.31	86.93±0.31	1.05	1.05	0.00
hepatitis	79.38±2.27	84.50±2.00	88.38±1.62	2.39	4.56	2.13
anneal	92.54±1.18	92.55±1.24	94.21±0.86	0.01	1.62	1.56
sonar	69.74±1.97	83.19±2.30	83.74±3.75	6.28	4.67	0.18
genetics	94.17±0.39	94.67±0.30	94.17±0.39	1.44	0.00	-1.44
dna	92.40±0.46	95.45±0.38	94.51±0.39	7.23	4.95	-2.44
Average	83.09	85.70	86.32			

Dataset	$\mathcal{CN2}$	FSS(wf, $\mathcal{CN2}$)	FSS(wb, $\mathcal{CN2}$)	FSS(wf, $\mathcal{CN2}$) - $\mathcal{CN2}$	FSS(wb, $\mathcal{CN2}$) - $\mathcal{CN2}$	FSS(wb, $\mathcal{CN2}$) -FSS(wf, $\mathcal{CN2}$)
bupa	67.82±2.11	65.81±1.83	65.81±1.83	-1.02	-1.02	0.00
pima	74.62±1.38	74.75±1.43	74.75±1.43	0.09	0.09	0.00
breast cancer	94.99±1.42	96.85±0.60	95.57±0.97	1.71	0.48	-1.59
hungaria	77.93±3.06	83.66±2.60	80.65±3.94	2.02	0.77	-0.90
crx	83.20±1.21	86.83±1.23	85.83±0.86	2.98	2.51	-0.94
letter	70.34±0.30	76.17±0.20	76.17±0.20	22.87	22.87	0.00
hepatitis	81.75±3.83	90.99±1.66	87.13±2.81	3.13	1.60	-1.67
anneal	90.42±1.08	96.44±0.94	96.44±0.94	5.95	5.95	0.00
sonar	71.19±3.30	81.32±3.46	75.52±2.75	3.00	1.43	-1.86
genetics	79.53±1.45	89.47±0.53	83.26±1.73	9.11	2.34	-4.85
dna	88.15±0.62	94.89±0.41	N/A	12.82	N/A	N/A
Average	79.99	85.20	82.11			

continued on next page

continued from previous page						
Dataset	IB	FSS(wf,IB)	FSS(wb,IB)	FSS(wf,IB) -IB	FSS(wb,IB) -IB	FSS(wb,IB) -FSS(wf,IB)
bupa	61.66±2.82	65.82±1.68	65.82±1.68	1.79	1.79	0.00
pima	69.18±1.92	72.56±1.44	70.87±1.43	1.99	1.00	-1.18
breast cancer	95.00±0.88	96.57±0.80	96.57±0.57	1.87	2.12	0.00
hungaria	78.92±1.93	83.64±3.02	82.00±2.56	1.86	1.36	-0.59
crx	81.74±0.90	87.39±1.18	83.48±1.11	5.38	1.72	-3.41
letter	95.17±0.13	96.19±0.13	96.19±0.13	7.85	7.85	0.00
hepatitis	81.29±2.23	87.13±2.67	86.54±2.17	2.37	2.39	-0.24
anneal	99.33±0.25	99.78±0.15	99.67±0.17	2.18	1.59	-0.69
sonar	85.60±2.00	92.81±1.27	89.43±1.99	4.30	1.92	-2.02
genetics	78.75±0.71	90.88±0.62	83.13±0.68	18.20	6.30	-11.91
dna	74.23±0.44	93.72±0.34	N/A	49.57	N/A	N/A
Average	81.90	87.86	85.37			

Dataset	NB	FSS(wf,NB)	FSS(wb,NB)	FSS(wf,NB) -NB	FSS(wb,NB) -NB	FSS(wb,NB) -FSS(wf,NB)
bupa	55.44±2.95	60.90±2.32	60.90±2.32	2.06	2.06	0.00
pima	75.55±1.05	77.25±0.99	77.25±0.99	1.67	1.67	0.00
breast cancer	96.14±0.68	95.85±0.84	96.42±0.53	-0.38	0.46	0.81
hungaria	83.67±2.83	85.68±2.27	84.36±2.76	0.78	0.25	-0.52
crx	77.97±1.12	87.10±1.25	87.39±1.18	7.69	8.19	0.24
letter	64.55±0.36	66.31±0.43	66.31±0.43	4.44	4.44	0.00
hepatitis	83.08±3.72	87.08±2.39	88.29±1.93	1.28	1.76	0.56
anneal	91.65±0.71	88.42±0.60	91.87±0.66	-4.91	0.32	5.47
sonar	69.26±4.53	78.74±3.24	74.07±3.61	2.41	1.17	-1.36
genetics	95.45±0.32	96.21±0.26	96.11±0.23	2.61	2.37	-0.41
dna	94.04±0.34	96.83±0.23	95.29±0.43	9.61	3.22	-4.47
Average	80.62	83.67	83.48			

Dataset	TM	FSS(wf,TM)	FSS(wb,TM)	FSS(wf,TM) -TM	FSS(wb,TM) -TM	FSS(wb,TM) -FSS(wf,TM)
bupa	59.71±0.44	61.13±2.03	61.13±2.03	0.97	0.97	0.00
pima	64.89±0.16	67.75±1.31	67.49±0.79	3.06	4.56	-0.24
breast cancer	66.09±0.23	92.27±0.69	91.85±0.77	50.90	45.33	-0.57
hungaria	63.94±0.54	83.30±2.48	83.30±2.48	10.79	10.79	0.00
crx	55.51±0.22	86.96±0.99	86.96±1.04	43.86	41.84	0.00
letter	12.79±0.31	68.61±0.46	68.61±0.46	142.31	142.31	0.00
hepatitis	79.33±0.89	81.38±2.56	80.67±0.89	1.07	1.51	-0.37
anneal	76.84±0.22	99.67±0.17	99.56±0.25	116.13	96.48	-0.51
sonar	53.38±0.54	56.24±1.51	56.36±2.45	2.52	1.68	0.06
genetics	60.94±0.28	92.95±0.52	92.95±0.52	76.65	76.65	0.00
dna	61.24±0.47	94.73±0.35	94.73±0.35	80.82	80.82	0.00
Average	59.51	80.45	80.33			
Total Average	77.02	84.58	83.52			

Table 51: Absolute Difference in Standard Deviations of Accuracies

Although the quality of the results varies across datasets, in general, the wrapper approach outperforms the standard inducer. On total average, the standard inducer accuracy has improved from 77.02% to 84.58% for forward selection and to 83.52% for backward selection. This represents 32.87% and 28.29% relative reduction in error rates, respectively.

Regarding the $\mathcal{C}4.5$ as black box wrapper inducer, the wrapper has improved accuracy gracefully against the standard inducer, except for bupa dataset. In general there is no significant improvement when comparing backward with forward selection, except for hepatitis. For dna, $FSS(wb, \mathcal{C}4.5)$ degraded significantly over $FSS(wf, \mathcal{C}4.5)$.

Considering $\mathcal{CN}2$, the accuracy obtained by forward selection is equivalent or outperforms backward selection. For the genetics dataset, $FSS(wf, \mathcal{CN}2)$ significantly outperforms $FSS(wb, \mathcal{CN}2)$ using only 8.33% of features. The average accuracy increases from 83.26% (for backward selection) to 89.47% (for forward selection) which is a 37.10% relative reduction in error rate.

For IB, the wrapper outperforms the standard inducer for all datasets. In the majority of cases forward outperforms backward selection and for crx and genetics datasets it outperforms significantly. Improvement using forward selection for genetics was from 78.75% to 90.88% which is a 57.08% relative reduction in error rate, using only 8.33% of features.

An interesting result is that only the NB standard inducer significantly outperforms the forward selection for anneal dataset. For dna dataset both forward and backward selection significantly outperforms the standard inducer and $FSS(wf, NB)$ significantly outperforms $FSS(wb, NB)$. Similar results had been obtained with $\mathcal{C}4.5$. In fact, $FSS(wf, NB)$ obtained the best accuracy of 96.83% for dna dataset using only 22.22% of features. With a complexity penalty factor greater than zero an accuracy of 94.60% using 7.78% of features

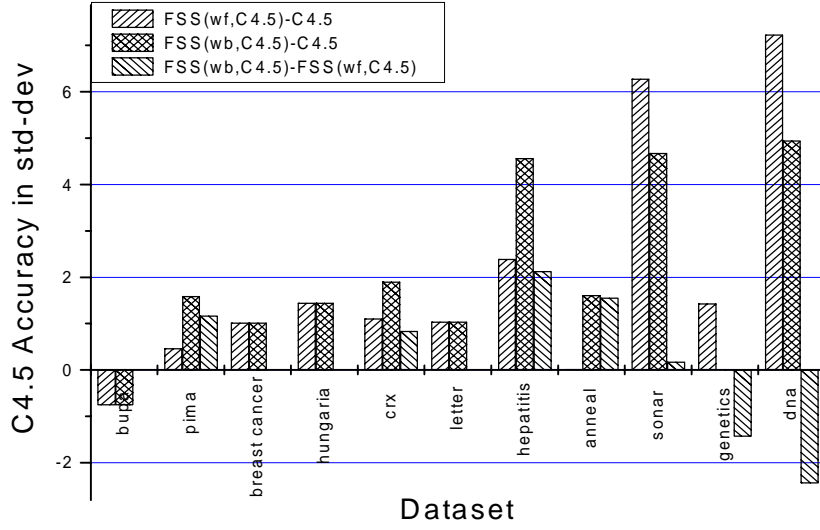


Figure 5: C4.5 Absolute Difference in Standard Deviations of Accuracies

is reported in (Kohavi, 1997). Indeed, no other known inducer outperforms NB for this dataset which includes several irrelevant features.

Table 52 shows improved accuracies at significance level (95% confidence) for forward and backward selection compared with the standard inducer. Improvements above two standard deviations are reported with Δ and those below with ∇ . Observe that the standard inducer outperforms significantly only once for anneal dataset using FSS(wf,NB).

Note that, in general C4.5 has improved (in the same datasets) less than the other inducers for both forward and backward selection. Indeed, $\mathcal{CN}2$, IB and NB always have improved their accuracies more in forward than in backward selection. As expected, it seems that datasets with many features tend to benefit more than those with few features. The overall scenario shows that the wrapper for FSS has significantly outperformed the standard inducer in 33 of 55 experiments for forward selection and in 24 of 53 experiments for backward selection for all studied domains and inducers.

Dataset	FSS					Count Δ	Count ∇					
	(wf,C4.5)	(wf,CN2)	(wf,IB)	(wf,NB)	(wf,TM)			(wb,C4.5)	(wb,CN2)	(wb,IB)	(wb,NB)	(wb,TM)
bupa				Δ								
pima					Δ	2	0					
breast cancer					Δ	3	0					
hungaria		Δ			Δ	3	0					
crx		Δ	Δ	Δ	Δ	7	0					
letter		Δ	Δ	Δ	Δ	8	0					
hepatitis	Δ	Δ	Δ			5	0					
anneal		Δ	Δ	∇	Δ	5	1					
sonar	Δ	Δ	Δ	Δ	Δ	6	0					
genetics		Δ	Δ	Δ	Δ	8	0					
dna	Δ	Δ	Δ	Δ	Δ	8	0					
Count Δ	3	8	7	6	9	3	4	4	5	8	57	
Count ∇	0	0	0	1	0	0	0	0	0	0		1

Table 52: Improved Accuracies at Significance Level

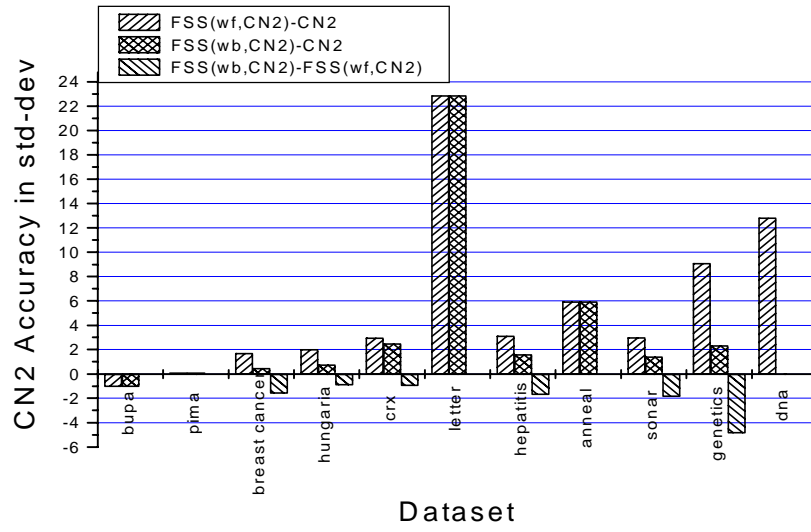


Figure 6: $\mathcal{CN}2$ Absolute Difference in Standard Deviations of Accuracies

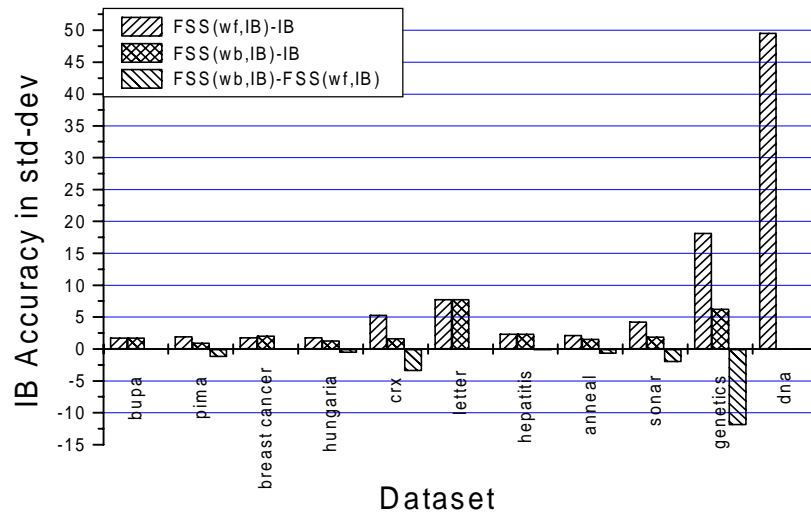


Figure 7: IB Absolute Difference in Standard Deviations of Accuracies

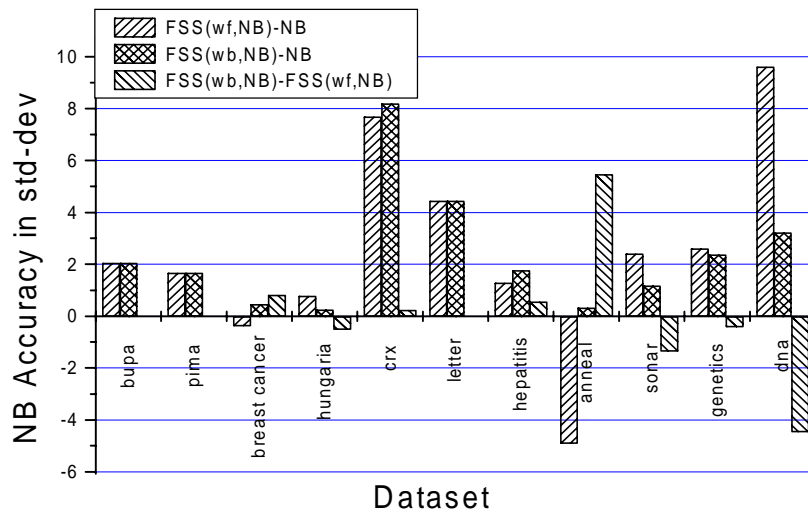


Figure 8: NB Absolute Difference in Standard Deviations of Accuracies

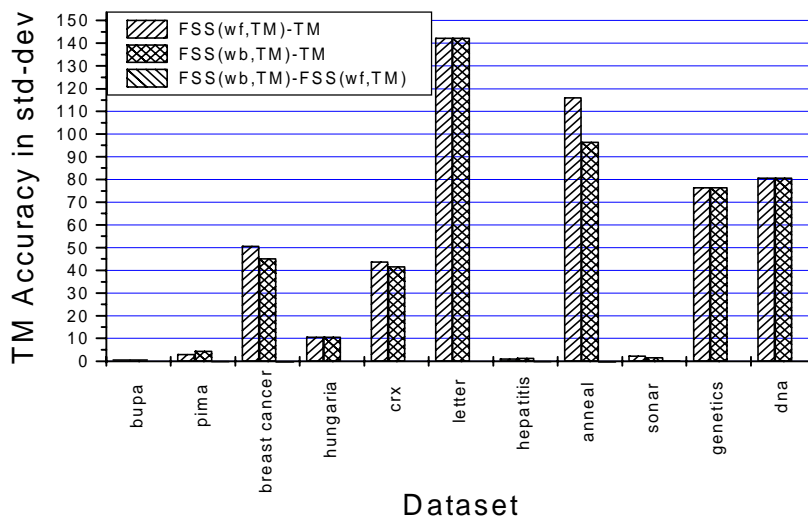


Figure 9: TM Absolute Difference in Standard Deviations of Accuracies

8 Conclusions

In this work we have used the wrapper approach for Feature Subset Selection. The FSS algorithm from *MCC++* library was used to run experiments with datasets ranging from few to many features and instances. Accuracies for the five inducers using all features and features found by FSS are presented. Results confirm the superiority of the FSS wrapper approach although in some cases the computational cost is excessive.

Feature Subset Selection can be defined as a problem of finding a subset of relevant features that describes the dataset as well as the original features do. The practical objective is to remove extraneous features, not necessarily to select the optimal subset, leaving the dataset reduced to a manageable dimension. It is also important to keep in mind that it is much safer to include more features than necessary, rather than fewer, since inducers can eventually cope with extra features but cannot make up for features that have been discarded. This view has been considered in this work.

For Data Mining, a special constraint for feature selection is that the size of a dataset is normally very large, both vertically and horizontally. For this sort of data, the running time would make the wrapper approach infeasible.

It could be argued that typical application of wrapper methods for feature selection occurs only once and for large datasets a less costly accuracy estimation method (*MCC++* default is ten-fold cross-validation) could be used. For instance, in our experiments for dna dataset using forward selection, more than 1,200 nodes were evaluated for each inducer. In fact, for large datasets holdout accuracy estimation can be used instead of cross-validation, improving execution time by a factor of 10. Also, as the *MCC++* wrapper is an anytime algorithm, the search can be stopped after a while and the current best feature subset selected so far can be used. Still, much work remains to be done to find approaches that are more effective for Data Mining tasks.

Most studies in supervised Machine Learning discuss accuracy on an unseen test set as the performance component and in our experiments the goodness of feature subsets were measured taking into account only the accuracy of the induced classifier. However, for Data Mining is also important to pay careful attention to the task of gaining the user/expert's acceptance and validation of the extracted knowledge. For this, not only *accuracy* but also *comprehensibility* of the knowledge produced should be considered.

We are currently working on this field of knowledge *interestingness*. We are checking if the rules generated by *C4.5* and *CN2* (symbolic inducers), when using all features and the ones selected by both FSS wrappers and filters approaches improve rule *interestingness* (Baranauskas et al., 1999a; Baranauskas et al., 1999b). This is being carried out considering not only the widely-used criteria of rule coverage, completeness and confidence factors but also considering new additional factors influencing rule *interestingness* (Freitas, 1998).

Acknowledgments: We are grateful to Jaqueline Brigladori Pugliesi and Chandler Caulkins for helpful comments on the draft of this paper.

References

- Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36:267–287.
- Aha, D. W. (1997). Lazy learning. *Artificial Intelligence Review*, 11:7–10.
- Almuallim, H. & Dietterich, T. G. (1991). Learning with many irrelevant features. In *Proceedings AAAI-91*, pages 547–552, Anaheim, CA. MIT Press, Cambridge, MA.
- Baranauskas, J. A., Monard, M. C., & Horst, P. S. (1999b). Evaluation of *CN2* induced rules using feature selection. *Argentine Symposium on Artificial Intelligence (ASAI)*. Submitted.
- Baranauskas, J. A., Monard, M. C., & Horst, P. S. (1999a). Evaluation of feature selection by wrapping around the *CN2* inducer. *Encontro Nacional de Inteligência Artificial (ENIA)*. Accepted.

- Blake, C., Keogh, E., & Merz, C. J. (1998). UCI repository of machine learning databases. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- Blum, A. L. & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, pages 245–271.
- Clark, P. & Boswell, R. (1991). Rule induction with CN2: Some recent improvements. In Kodratoff, Y., editor, *Proceedings of the 5th European Conference (EWSL 91)*, pages 151–163. Springer-Verlag.
- Clark, P. & Niblett, T. (1987). Induction in noise domains. In Bratko, I. & Lavrač, N., editors, *Proceedings of the 2nd European Working Session on Learning*, pages 11–30, Wilmslow, UK. Sigma.
- Clark, P. & Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, 3(4):261–283.
- Fayyad, U. M., Djorgovski, S. G., & Weir, N. (1996a). From digitized images to on-line catalogs: Data mining a sky survey. *AI Magazine*, 17(2):51–66.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R., editors (1996b). *Advances in Knowledge Discovery and Data Mining*. American Association for Artificial Intelligence, Menlo Park, CA.
- Felix, L. C. M., Rezende, S. O., Doi, C. Y., de Paula, M. F., & Romanato, M. J. (1998). *MCC++* biblioteca de aprendizado de máquina em C++. Technical Report 72, ICMC-USP.
- Freitas, A. A. (1998). A multi-criteria approach for the evaluation of rule interestingness. In *Proceedings of the International Conference on Data Mining*, pages 7–20, Rio de Janeiro, RJ.
- Gorman, R. P. & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1:75–89.
- KDD 95 (1995). *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, Menlo Park, CA. American Association for Artificial Intelligence.
- KDD 96 (1996). *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Menlo Park, CA. American Association for Artificial Intelligence.
- Kira, K. & Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the European Conference on Machine Learning*.
- Kira, K. & Rendell, L. A. (1998). The feature selection problem: Traditional methods and a new algorithm. In Press, M., editor, *Tenth National Conference on Artificial Intelligence*, pages 129–134.
- Kohavi, R. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324.
- Kohavi, R. & Sommerfield, D. (1995). Feature subset selection using the wrapper model: Overfitting and dynamic search space topology. In (KDD 95, 1995), pages 192–197.
- Kohavi, R., Sommerfield, D., & Dougherty, J. (1994). *MCC++: A Machine Learning Library in C++*. IEEE Computer Society Press.
- Kohavi, R., Sommerfield, D., & Dougherty, J. (1996). Data mining using *MCC++*: A machine learning library in C++. *Tools with Artificial Intelligence*, pages 234–245.
- Langley, P. (1996). *Elements of Machine Learning*. Morgan Kaufmann Publishers, Inc.
- Langley, P., Iba, W., & Thompson, K. (1992). An analysis of bayesian classifiers. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 223–228. AAAI Press and MIT Press.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1:81–106. Reprinted in Shavlik and Dieterich (eds.) *Readings in Machine Learning*.
- Quinlan, J. R. (1988). *C4.5 Programs for Machine Learning*. Morgan Kaufmann, CA.
- Taylor, C., Mitchie, D., & Spiegelhater, D. (1994). *Machine Learning, Neural and Statistical Classification*. Paramount Publishing International.