การเรียนรู้ด้วยต้นไม้การตัดสินใจโดยใช้เลนส์หลายลักษณะประจำ

นายเจริญชัย สิริสมบูรณ์รัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556

DECISION TREE-BASED LEARNING USING MULTI-ATTRIBUTED LENS

Mr. Charoenchai Sirisomboonrat

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in
Computer Science and Information Technology
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2013

Thesis Title	DECISION TREE-BASED LEARNING USING MULTI-ATTRIBUTED LENS
By	Mr. Charoenchai Sirisomboonrat
Field of Study	Computer Science and Information Technology
Thesis Advisor	Assistant Professor Krung Sinapiromsaran, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.....Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

.....Chairman

(Assistant Professor Jaruloj Chongstitvatana, Ph.D.)

.....Thesis Advisor

(Assistant Professor Krung Sinapiromsaran, Ph.D.)

.....Examiner

(Arthorn Luangsodsai, Ph.D.)

.....External Examiner

(Chumphol Bunkhumpornpat, Ph.D.)

เจริญชัย สิริสมบูรณ์รัตน์ : การเรียนรู้ด้วยต้นไม้การตัดสินใจโดยใช้เลนส์หลายลักษณะประจำ. (DECISION TREE-BASED LEARNING USING MULTI-ATTRIBUTED LENS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.กรุง สินอภิรมย์สราญ, 117 หน้า.

การสร้างต้นไม้การตัดสินใจเป็นหนึ่งในปัญหาการจำแนกข้อมูลที่เป็นที่รู้จักและถูกใช้กันอย่างกว้างขวาง ในกระบวนการสร้างต้นไม้การตัดสินใจชุดข้อมูลจะถูกแบ่งกั้น ด้วยลักษณะประจำที่ถูกเลือกว่าดีที่สุดจากตัววัดสารสนเทศ หลังจากนั้นจะมีการใช้ลักษณะประจำที่ดีที่สุดในผลแบ่งกั้นมาแบ่งข้อมูลซ้ำ ๆ จนกระทั่งเข้าเงื่อนไขการหยุด ในวิทยานิพนธ์เล่มนี้เสนอแนวคิดใหม่ในการสร้างตันไม้การตัดสินใจโดยใช้เลนส์หลายลักษณะประจำ แทนที่จะใช้ลักษณะประจำเพียงหนึ่งลักษณะในการแบ่งทุกครั้ง ลักษณะประจำทุกตัวจะถูกนำมาใช้ร่วมกันในการแบ่งชุดข้อมูล โดยจะหาคู่ระเบียนในชั้นเป้าหมายเดียวกันที่อยู่ห่างกันมากที่สุดมาสร้างเลนส์เพื่อนำมาแบ่งชุดข้อมูลเป็นสองส่วน โดยส่วนแรกคือส่วนของข้อมูลที่อยู่นอกเลนส์ซึ่งจะเป็นกลุ่มข้อมูลที่มีชั้นเป้าหมายตรงข้ามกับชั้นเป้าหมายของคู่ระเบียนที่ถูกนำมาใช้สร้างเลนส์นั้น และส่วนที่สองคือส่วนของข้อมูลที่อยู่ในเลนส์ซึ่งจำเป็นที่จะต้องถูกแบ่งต่อไป โดยการแบ่งในส่วนที่อยู่ในเลนส์จะฉายข้อมูลในส่วนดังกล่าวลงบนแกนของเวกเตอร์หลักที่สร้างมาจากคู่ระเบียนที่ไกลที่สุดและจะเลือกจุดแบ่งที่ดีที่สุดบนแกนดังกล่าวจากการวัดข้อมูลสารสนเทศของค่าที่ได้จากการฉาย ในการแบ่งส่วนนี้จะได้ข้อมูลย่อยสองส่วนเกิดขึ้นซึ่งแต่ละส่วนจะถูกดำเนินการตามขั้นตอนที่กล่าวมาซ้ำจนกระทั่งเข้าเงื่อนไขการหยุด ในการทดลอง t-tests ถูกใช้ทดสอบกับตัววัดประสิทธิภาพในชุดข้อมูลของ UCI และผลการทดลองแสดงให้เห็นว่าความแม่นยำของวิธีการจำแนกชุดข้อมูลด้วยขั้นตอนวิธีนี้ดีกว่ากว่าต้นไม้การตัดสินใจแบบ C4.5, k nearest neighbor, naïve Bayes, support vector machine และ logistic regression อย่างมีนัยสำคัญเชิงสถิติในชุดข้อมูล Blood Transfusion Service Center, Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis และ Horse Colic

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์

สาขาวิชา วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

ปีการศึกษา 2556

ลายมือชื่อนิสิต .....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

5472627123 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORDS: DECISION TREE / MULTI-ATTRIBUTED LENS / CORE VECTOR / CLASSIFICATION

CHAROENCHAI SIRISOMBOONRAT: DECISION TREE-BASED LEARNING USING MULTI-ATTRIBUTED LENS. ADVISOR: ASST. PROF. KRUNG SINAPIROMSARAN, Ph.D., 117 pp.

Decision tree induction is one of the popular and widely used method for classification. In a general decision tree algorithm, a dataset is split into partitions by the attribute which provides the best information measure value. Then, each partition recursively finds the best attribute until the terminal criteria are met. In this thesis, a new type of decision tree algorithm, multi-attributed lens, is presented. Instead of using one attribute at a time, all attributes are used to create a lens which is further used to split a dataset. A lens is generated from the farthest pair which splits a dataset into two regions. The first region is the outside lens region where its instances will be labeled as the opposite class of the farthest instances. The second region is the inside lens region where its instances need to be further partitioned. In splitting the instances of the inside lens, the best splitting point is computed by measuring the information measure of their projection values onto the core vector which is created from the farthest pair. Then, newly created partitions are recursively called until the termination conditions are met. Empirically, t-tests of the performance measures are gathered on the UCI datasets and the results show that accuracy of this algorithm is more statistically significant than C4.5 decision tree, k nearest neighbor, naïve Bayes, support vector machine and logistic regression in Blood Transfusion Service Center, Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis and Horse Colic datasets

Department: Mathematics and Computer Science
Field of Study: Computer Science and Information Technology
Academic Year: 2013

Student's Signature .....

Advisor's Signature .....

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest appreciation to Assistant Professor Dr. Krung Sinapiromsaran, my advisor who gave me all the supervision and encouragement throughout my research. Without his suggestion and persistent help, this thesis would not have been possible.

Moreover, I would like to thank my boss, Mark Bosscher, for understanding and allowing me to take some time off of work to focus on this thesis. I would also like to thank all of my friends and colleagues who have entertained me when I felt stressed.

Finally, I feel very grateful to my family for supporting me as well as cheering me up when things got rough. Last but most importantly, I feel happy to have my best friend, Sutira Limroscharoen, who helped me in every single way to finish this thesis, stayed beside me and provided me with unending encouragement and support.

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1 Motivation and literature surveys

Nowadays, classification in data mining is the essential tool for building up models based on the corresponding historical data. Many classification techniques have been proposed. All of those techniques mainly focus to predict as accurate as possible.

The main idea of classification is using a classifier to assign new unknown instances to one or more classes which are already defined in the dataset, for examples, classifying credit card transactions as legitimate or fraudulent [1], identifying loan applications as low, medium, or high credit risks [2], classifying tumor cells [3] as benign or malignant, categorizing news [4] as finance, weather, entertainment, sport, etc. The term of classifier sometimes refer to the mapper that maps an instance to a class. Most existing classifiers are built based on the values of the properties of instance which are known as attributes. The measurement scales of these properties can be mixed between nominal, ordinal, interval, as well as ratio.

The classification consists of two processes including a learning step, and a predictive step as shown in Figure 1. First, the dataset is separated into two sets. One is the training set and another is the test set. In the learning step, the training set is used to construct the classifier. There are many types of classifiers such as rule-based classifier, decision tree, nearest-neighbor classifier which represent the patterns in order to classify an unknown instance. In the predictive step, the outcome model from the first step will be used. To match the goal of classification that the classifier needs to classify unknown instance as accurate as possible. Examples of measurements are accuracy, precision, recall, F-measure. For a concrete example, in a decision tree, an instance will be tested with the condition from the root node and follow the appropriate branch based on the outcome of the test. The instance is repeatedly tested with internal nodes until it

reaches the leaf node. The class label of the instance is associated with that leaf node.

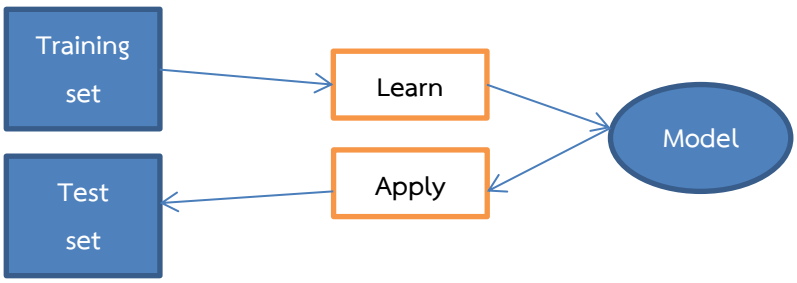

Figure 1 The steps of classification

Decision tree induction is one of the classic techniques and it has been widely used for many classification problems such as Impedance fault detection [5], cancer diagnosis [6], etc. In learning step, a decision tree generally uses a greedy strategy to select the best splitting attribute from all candidates by using an information measure and creates a node with the selected attribute. This attribute is also used to partition the dataset into subsets. Each subset is applied recursively until it meets the terminal criteria such as all instances in the same subset having the same class.

ID3 [7] was proposed by J. R. Quinlan, for discrete attributes. Afterward, other algorithms including C4.5 [8] were developed to work with both discrete and continuous attributes. Many researchers proposed the improvement of the decision tree such as optimizing the attribute selection. However, most researchers focus on selecting a single attribute to split dataset by measuring the information gain. Some datasets contain attributes that are correlated to one another. They can be used together for building a more accurate decision tree.

PCA-based [9] method is proposed by I. J. Jolliffe which can be used to reduce attributes of a dataset. It can be applied during the data preparation process

before the dataset is used by decision tree algorithms such as ID3, C4.5. Although, it increases relative performances and helps reduce number of attributes, the extra time is required to calculate the eigenvalues and eigenvectors. Moreover, W. Sun et al [10] has shown that even if PCA reduced 18 attributes to 6 eigenvectors but the total accuracy did not improve significantly.

The concept of simultaneous attributes has been proposed by K. Sinapiromsaran et al [11] called "the multi-attributed frame" and it has been applied in the network intrusion detection system. A core vector is built from the farthest instances in the dataset. Instances in the dataset are partitioned into three regions: left region, middle region, and right region. As the results, the performances were shown to detect more malicious accesses. However, this algorithm ignores the discrete attribute. This could make the classifier lose some important information.

Our experiments use C4.5 model and also the other classifiers including: k nearest neighbor, naïve Bayes, support vector machine, and logistic regression compare with our model.

- **K nearest neighbor**

In k nearest neighbor [12], all data instances are kept in training set correspond to points in the n-dimensional space where n is the number of attributes. Given a test instance, the algorithm finds distances to all instances in training set by a distance measure which is usually Euclidean distance. Then, it obtains the k instances closet to the test instance where k is the user-defined value. Once nearest instances are listed, the test instance is classified based on majority class of its nearest instances. In Figure 2, k is equal to 1 with the negative instance as a neighbor. Therefore, the instance is classified as the negative class. In Figure 3, k is equal to 3; its nearest list contains two positive instances and one negative instance. By the majority voting, this test instance is classified as the positive class. In case of the test instance cannot be classified by a majority vote, the algorithm randomly classifies the test instance.

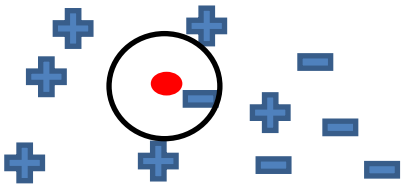

Figure 2 1-nearest neighbor from the given dataset

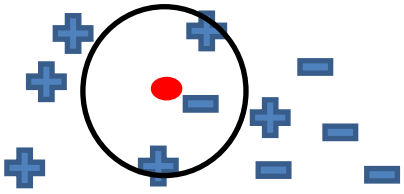

Figure 3 3-nearest neighbor from the same dataset

- **Naïve Bayes**

  A naïve Bayes classifier [13] uses probabilistic and statistic concepts based on Bayes theorem by assuming all attributes of the dataset are conditionally independent. Due to the use of probability, this classifier is robust to isolated noise instances. Besides, the classifier can handle missing values while building a model or classifying an instance

- **Support vector machine**

  A support vector machine or SVM [14] is a classifier which uses a hyperplane to linearly separate between two classes of the dataset with the largest margin. Since it needs to maximize margin of hyperplane or known as the maximum-margin hyperplane, it is considered as an optimization problem. When the classes cannot be separated linearly, there is a way known as nonlinear SVM to separate the classes by applying the kernel method to maximum-margin hyperplane. The

classifier will use a kernel function which is a function that transforms the original data into the higher dimension before applying the hyperplane.

- **Logistic regression**

A logistic regression [15] is a type of regression analysis used in statistics for predicting the result based on taking logistic function known as sigmoid function shown in Figure 4 on combination of our input attributes as equation (1) where X_i is a value of each attribute and w_i is a coefficient weight. The formula can also be written as a probability distribution as shown in equation (2) which its log likelihood function as shown in equation (3). Unfortunately, there is no analytical solution for determining all weights for this formula. However, if log likelihood function is convex, a global maximum exists. Thus, we may find the weights by using gradient descent which is taking the partial derivatives of likelihood function with respect to the weights.

$$w^T X^i = w_0 X_0^i + w_1 X_1^i + w_2 X_2^i + \ldots + w_n X_n^i \qquad X_0^i = 1 \tag{1}$$

$$P(Y \mid X, w) = \prod_{i=1}^{m} \frac{(e^{-w^T X^i})(1 - Y^i)}{1 + e^{-w^T X^i}} \tag{2}$$

$$l(w) = \sum_{i=1}^{m} (1 - Y^i)(-w^T X^i) - \log(1 + e^{-w^T X^i}) \tag{3}$$

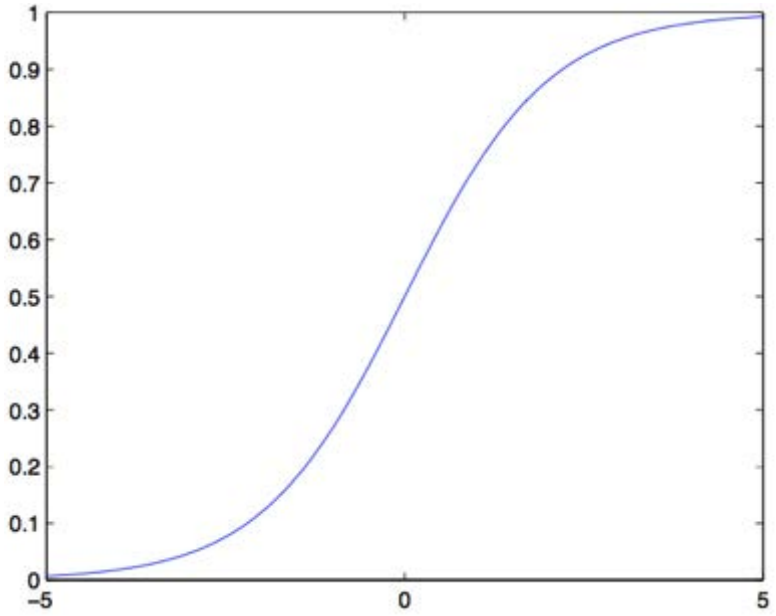

Figure 4 The sigmoid function

Due to the correlation between attributes in a dataset, Quinlan's decision tree may not split datasets appropriately. In our research study presented in this thesis, we proposed the new classifier that utilizes multiple attributes. The algorithm uses a lens, which is created from farthest instances, to split a dataset. In addition, a lens will separate the dataset exactly into two regions an inside lens region and an outside lens region. Then, the algorithm will continue to further split the data inside the lens by the new axis called core vector which is also created from the farthest pair. The algorithm will repeat these steps until the terminal criteria are met. Furthermore, the multi-attribute frame did not cover the discrete attribute while our algorithm can be applied to all attribute types whether they are discrete or continuous.

1.2 Research objective

The goal of this research is to propose a new decision tree algorithm which uses all attributes instead of a single attribute at a time in the partitioning steps. This algorithm uses the lens, created from the farthest instances, to partition the dataset. The performances of this algorithm are reported for accuracy, sensitivity, specificity, area under ROC curve (AUC), information score [16], F-measure,

precision, recall, brier score [17], and Matthews correlation coefficient (MCC) [18]. All these evaluators are used to compare with other classifiers including k nearest neighbor, naïve Bayes, support vector machine, and logistic regression on the UCI datasets. Besides, we perform paired t-test significantly indicate on our model with other models.

1.3 Thesis overview

The rest of the thesis is organized as follow. This thesis proposes the multi-attributed lens and applies this algorithm to UCI datasets as well as compares with other previous techniques. First, we will explain the fundamental of the decision tree and the classification evaluation in chapter 2. Afterwards, the concept of extreme poles, core vector, and lens as well as the multi-attributed lens algorithm will be described in chapter 3. The results on the UCI classification datasets will be shown in chapter 4. Finally, we will summarize the outcome of this thesis in chapter 5.

CHAPTER II

BACKGROUND KNOWLEDGE

2.1 Decision tree

In this section, we explain step-by-step to build a decision tree. C4.5 algorithm will be described for the comparison purpose.

A decision tree classifier or classification tree is commonly used in classification technique. The structure of this model can be visualized as a tree as shown in Figure 5 generated from the breast cancer dataset. The two types of tumors, benign and malignant, are the class labels of this dataset. The root node is the first internal node of the decision tree. All internal nodes contain the test conditions to partition a dataset into various subsets. The edges are represented by one of possible outcomes of the test condition from the internal node. The leaf nodes are labeled by the class. The example shows that the root node contains the "Cell Size" test condition. If an instance has "Cell Size" less than or equal to 0.390, then it will go to the left node which will be tested against the "Bare Nuclei" condition. Otherwise, the test instance will go to the right node which is the leaf node that is classified as Malignant.

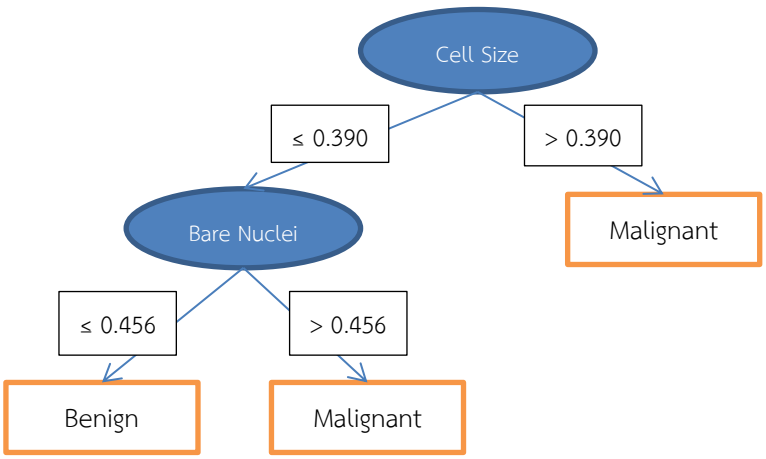

Figure 5 An example of the decision tree

```
1.  CreateDecisionTree(D)
2.      if StoppingCond(D) is true then
3.          node = LeafNode()
4.          node.label = Majority(D)
5.      else
6.          node = InnerNode()
7.          node.condition = FindBestSplit(D)
8.          foreach D_v in D do
9.              node.children.Add(CreateDecisionTree(D_v))
10.         end for
11.     end if
12.     return node
```

Figure 6 A general decision tree algorithm

Figure 6 shows a decision tree induction algorithm. It starts with the dataset *D* that is partitioned recursively using the best split (Line 6-7) where D_v is the subpartition that the dataset *D* is split by the condition *v*. Then, the tree expands to other nodes (Line 8-10) until the stopping condition (Line 2-4) is reached. The *StoppingCond()* function will check all instances whether they are the same class or meet any specific conditions. The *Majority()* function is used to find the class of the leaf node by majority votes. The *FindBestSplit()* function is used to identify a splitting value by computing the information measure. For example, the information gain is used by ID3, the gain ratio is used by C4.5 and GINI index is used by CART [19].

- **C4.5**

ID3 is an algorithm proposed by Quinan in 1986. Unfortunately, it only works with discrete datasets. However, he had improved ID3 algorithm called C4.5 algorithm in order to work with continuous datasets in 1993 and it has become a famous classifier.

In order to find the best split value, the attribute that has the highest information gain will be selected from a pool. The information gain can be calculated by equation (4) where D is a dataset and V is a selected attribute and the entropy, which is the impurity measure, can be calculated by equation (5) where $|C|$ is a number of classes, and p_i is the probability of ith class in D.

$$InfoGain_V(D) = Entropy(D) - \sum_{v \in V}\left(\frac{|D_v|}{|D|} \times Entropy(D_v)\right) \quad (4)$$

$$Entropy(D) = -\sum_{i}^{|C|}\left(p_i \times \log p_i\right) \quad (5)$$

In addition, C4.5 algorithm will use the gain ratio as the information measure which is the proportion between the information gain and the split information. The split information can be computed by equation (6) where p_v is a proportion of instances in D with the outcome v.

$$SplitInfo_V(D) = -\sum_{v \in V}\left(p_v \times \log p_v\right) \quad (6)$$

The corresponding gain ratio is calculated by equation (7).

$$GainRatio_V(D) = \frac{InfoGain_V(D)}{SplitInfo_V(D)} \quad (7)$$

The illustration of the information gain and the gain ratio of C4.5 algorithm displays in Table 2 where the data comes from Table 1. For the continuous attribute, the values of the attribute “A” needs to be sorted and the middle value between two adjacent values will be computed. Then, it is used to partition the dataset into two sets: a less than or equal to the middle value set and a greater than the middle value set. In Table 2, the attribute “A” with the splitting value

equal to 5.5 will be selected as both of the information gain and the gain ratio have the highest score.

Table 1 Example of a training dataset

A	B	Class
9	u	+
7	u	+
7	y	+
4	y	-
3	y	-

Table 2 Example of information measures

	A			B
	A=8	A=5.5	A=3.5	
Information gain	0.1185	0.6730	0.2231	0.2911
Gain ratio	0.2368	1.0000	0.4459	0.4325

The *FindBestSplit()* function of C4.5 algorithm shown in Figure 7 which will compare the best split between the best discrete attribute in Figure 8 and the best continuous attribute in Figure 9.

```
1. FindBestSplit(D)
2.      bestDis = FindBestDisc(D)
3.      bestCon = FindBestCont(D)
4.      return Max(bestDis.score, bestCon.score)
```

Figure 7 Find the best split function in C4.5 algorithm

```
1. FindBestDisc(D)
2.      foreach V in the possible discrete attributes in D do
3.        for v in V do
                    D_v = { x|x is a instance where V value equal to v }
5.        end for
6.        score = the information score of D
7.        if(score > maxScore) then
8.                  maxScore = score
9.                  bestAttr = V
10.       end if
11.     end for
12.     return { attribute:bestAttr, score:maxScore }
```

Figure 8 Find the best discrete attribute function in C4.5 algorithm

```
1. FindBestCont(D)
2.      foreach V in the continuous attributes in D do
3.          T = { t|t = (v_i + v_{i+1})/2 where v_1, v_2, ..., v_n are the ordered
distinct values of attribute V in D }
4.          foreach t in T do
5.                  D_l = { x|x is an instance in the dataset D that has
the value of the attribute V value less than or equal to t }
6.                  D_r = { x|x is an instance in the dataset D that has
the value of the attribute V greater than t }
7.                  score = the information score of D
8.                  if(score > maxScore) then
9.                          maxScore = score
10.                         bestAttr = V
11.                         bestThres = t
12.                 end if
13.         end for
14.     end for
15.     return { attr:bestAttr, thres:bestThres, score:maxScore }
```

Figure 9 Find the best continuous attribute function in C4.5 algorithm

2.2 Performance measures

Table 3 A confusion matrix

		Actual Class	
		+	**-**
Predicted Class	**+**	TP	FP
	-	FN	TN

In order to measure the classification performance whether it can effectively and efficiently predict the result, the evaluators will be used. Such evaluators are calculated based on the values from the confusion matrix.

Table 3 exhibits the confusion matrix which we will be using. The confusion matrix contains 4 cells of numbers, which each represents the relationship between the actual and the predicted classes of the classification. The upper-left cell, TP or True Positive, is the number of positive instances that are correctly classified. The upper-right cell, FP or False Positive, is the number of negative instances that are incorrectly classified. The lower-left cell, FN or False Negative, is the number of positive instances that are incorrectly classified. The lower-right cell, TN or True Negative, is the number of negative instances that are correctly classified.

The following performance measures are derived from the confusion matrix.

- **Accuracy**

Accuracy determines the level of true results from the total population. The calculation is shown in equation (8):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

- **Sensitivity or recall or true positive rate (TPR)**

Sensitivity, sometimes it is called recall, is the level of correct prediction given the positive actual class. The calculation is shown in equation (9):

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

- **Specificity or true negative rate (TNR)**

Specificity is the level of correct prediction given the negative actual class. The calculation is shown in equation (10):

$$Specificit\,y = \frac{TN}{TN + FP} \tag{10}$$

Moreover, we can find false positive rate (FPR) which is one minus the value of specificity.

- **Precision**

Precision is the level of correct result among the positive predictions. The calculation is shown in equation (11):

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

- **F-measure**

F-measure or F_1 is the score that balances the level of precision and recall. The calculation is shown in equation (12):

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

- **Matthews correlation coefficient**

While F-measure left out True Negative, Matthews correlation coefficient considers all four values in the confusion matrix which allows it to be able to apply to various sizes of the population. The calculation is shown in equation (13):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{13}$$

- **Area under ROC curve**

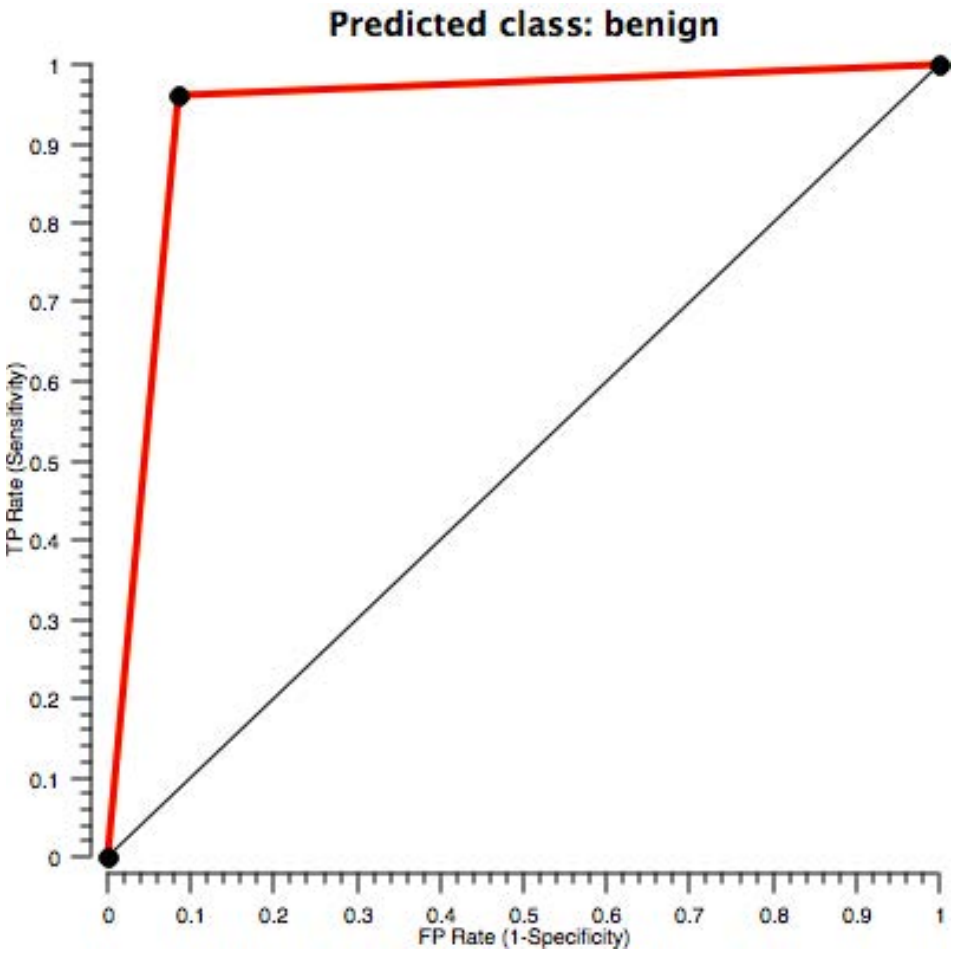

Figure 10 Area under ROC curve

The ROC curve represents the tradeoff between the true positive rate (TPR) which is plotted on the y axis and the false positive rate (FPR) which is plotted on the x axis as shown in Figure 10. The common critical points are the followings:

The model predicts every instance as negative (TPR=0, FPR=0)

The model predicts every instance as positive (TPR=1, FPR=1)

The model predicts every instance correctly (TPR=1, FPR=0)

Hence, a good model should be located on the upper left corner of the graph, while a random model should be placed along the diagonal line of the graph.

- **Information score**

Information score or IS was introduced by Kononenko and Bratko 1991 which measures the performance of the classifier using the entropy and the probability. Let $P(V+)$ be the prior probability of class $V+$ and $P'(V+)$ be the posterior probability of class $V+$ from the classifier. There are three possible cases as follows:

1. if $P'(V+) > P(V+)$ then

$$IS = -log^2 P(V+) + log^2 P'(V+)$$

2. if $P'(V+) < P(V+)$ then

$$IS = log^2(1 - P(V+)) - log^2(1 - P'(V+))$$

3. if $P'(V+) = P(V+)$ then

$IS = 0$, since, it has no information.

- **Brier score**

Brier score is the mean squared error of the probability forecast which is commonly used to measure the classification accuracy. It can be computed by equation (11):

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2 \qquad (11)$$

f_t is the probability that was forecasted, o_t is the event outcome which is 1 if the event occurs and 0 if it does not occur. In addition, Brier score is 0 which is called the perfect score; on the other hand, 1 is called the worst score.

2.3 T-test

T-test is a simple method used to determine whether one technique is competitive to another based on the same circumstances. In this case, the N training and testing sets are iterated by given the proportion between training and test sets of the dataset. The i^{th} training set is used to create two models while the i^{th} test set is evaluated in order to identify the performance measures. Now, let M_{1i} and M_{2i} be the performance measures of the two classifiers on the i^{th} set. The significance of the t-statistic is computed as in equation (15) where the required variables are calculated using the formulas including the mean difference between two performance measures $\overline{d}$ in equation (16), the standard deviation of $\overline{d}$ in equation (17), the variance of the classifier's measure in equation (18), and the average of the classifier's measure in equation (19).

$$t = \frac{\bar{d}}{\sigma_{\bar{d}}} \qquad (15)$$

$$\bar{d} = \overline{M}_1 - \overline{M}_2 \qquad (16)$$

$$\sigma_{\bar{d}} = \sqrt{\frac{\sigma_{M_1}^2 + \sigma_{M_2}^2}{N}} \qquad (17)$$

$$\sigma_{M_k}^2 = \frac{\sum_{i=1}^{N}\left(M_{ki} - \overline{M}_k\right)^2}{N-1} \tag{18}$$

$$\overline{M}_k = \frac{\sum_{i=1}^{N} M_{ki}}{N} \tag{19}$$

CHAPTER III

MULTI-ATTRIBUTED LENS

A multi-attributed lens ("MAL") decision tree is derived from the concept of extreme poles by finding the farthest pair or poles of the same class instances. Consequently, a lens and a core vector are created from the farthest pair. A lens will be used to partition a dataset into two regions: the outside region and the inside region. The outside region is labeled as the opposite class of the poles. Meanwhile, the inside region will be partitioned with the best split point along the core vector. The best split point is calculated from the projection values with the information measure and each partition is repeated the process in recursive manner.

3.1 Concept of extreme poles, lens and core vector

The extreme poles concept is basically the two instances in the space having the maximum distance. For this reason, there is no other pair of the same target class can place further away than the distance of the extreme poles. We have theoretically applied this concept to the space of a binary dataset. Let a set D contains $p_1,p_2,..p_s$ and $n_1, n_2, .. n_t$ where p_i is the i^{th} positive instance and n_j is the j^{th} negative instance. In this thesis, we will use the Euclidean distance as a distance measure defined in equation (20) where $x = (x_1,x_2,...,x_n)$ and $y = (y_1,y_2,...,y_n)$ are the instances in the dataset.

$$Distance(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (20)$$

Properties:

1. Given the farthest positive pair p_l and p_r. For any instance o, if the distance from o to p_l or the distance from o to p_r is greater than the maximum distance between the two positive instances p_l and p_r, then the instance o must be the negative instance.

2. Given the farthest negative pair n_l and n_r. For any instance o, if the distance from o to n_l or the distance from o to n_r is greater than the maximum distance between the two negative instances n_l and n_r, then the instance o must be the positive instance.

Proof property 1:

Suppose the farthest positive pair p_l and p_r. Let o be an instance that $Distance(o,p_l) > Distance(p_l,p_r)$ or $Distance(o,p_r) > Distance(p_l,p_r)$. Assume o is a positive instance. Then o, p_l is the farthest positive pair which contradicts the farthest positive pair of p_l and p_r.

Property 2 can be proved in the same way as property 1.

The lens is created based on the properties as mentioned. There are two type of the lens including a positive lens and a negative lens:

- A positive lens which is the lens created from the farthest positive instances p_l and p_r that all positive instances must stay inside.
- A negative lens which is the lens created from the farthest negative instances n_l and n_r that all negative instances must stay inside.

The positive lens is created from the farthest positive instances and the n-dimensional balls. If the first n-dimensional ball is created around the p_l with the radius equal to the distance from p_l to p_r, then we can ensure that the outside ball's instances are not positive instances by property 1; however, the negative instances may remain inside and outside of this ball as shown in Figure 12. Next, another n-dimensional ball is created and centered at the p_r with the same radius. The positive instances must only remain inside this ball, while the negative instances can stay in both regions. As a result, the intersection area between these two positive balls is geometrically called a positive lens as shown

in Figure 13. The negative lens can be created by the n_l and n_r instances in the same way as the positive lens.

There is no positive instance that stays outside the positive lens, and this property will later be used to partition the dataset into two regions. The outside lens region is labeled as negative; however, the inside lens region needs to be further partitioned. Instead of determining the best vector projection for splitting, our method utilizes the vector that generate from p_l and p_r which is named core vector. In case of the negative lens, the outside lens region is labeled as positive while the inside lens region will be processed similarly to the positive lens.

The core vector is the vector which is defined from p_l to p_r for the positive lens. Each instance within the lens needs to be projected onto the core vector. The projection values can be computed by equation (21) where $\boldsymbol{u}$ is the core vector and $\boldsymbol{v}$ is the vector from p_l to an instance. Those values will be ordered, and then will be used to find the best split point in the same way as C4.5. The outside lens partition, the less than or equal to the best split value partition, and the greater than the best split value partition are included while calculating the information measure. All of the above three partitions are shown in Figure 14. For the negative lens, it follows the similar steps as the positive lens despite the change of pole type.

Note that other vectors beside the core vector may not be applicable for all instances inside the lens. When this classifier is used, the instance inside the lens is projected on the vector. Figure 15 shows an example of the projection from an instance to the core vector (green line) and the other vector (purple line). In this example, the instance can be projected on both core vector and other vector indifferently. However, there might be the case that the instance is projected on the other vector but such vector is not able to cover every instance as shown in Figure 16. Hence, there is an uncertain area where the split value cannot be clearly identified.

Moreover, the process of seeking the vector to be used instead of the core vector definitely requires more time. Suppose there are n instances, we need to

pair up all of the instances in order to obtain all possible vectors. The best split value is calculated for each vector with the time complexity of n^3.

$$Magnitude\ (v) = \frac{|u \bullet v|}{|u|} \tag{21}$$

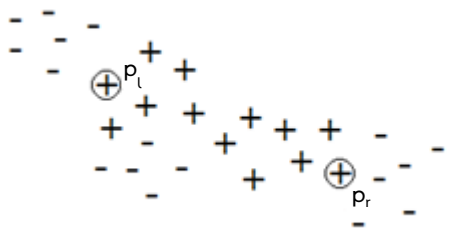

Figure 11 The farthest positive instances

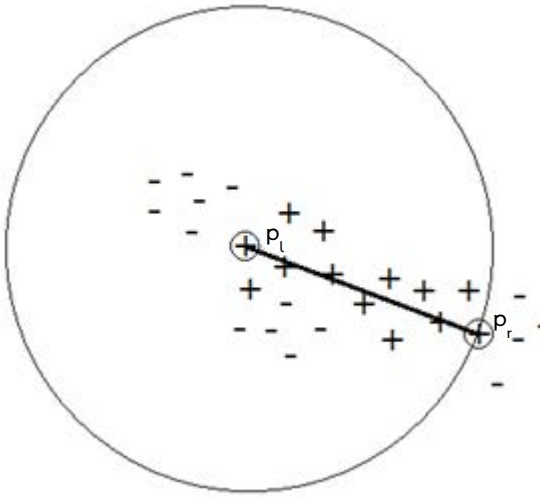

Figure 12 A ball is around p_l

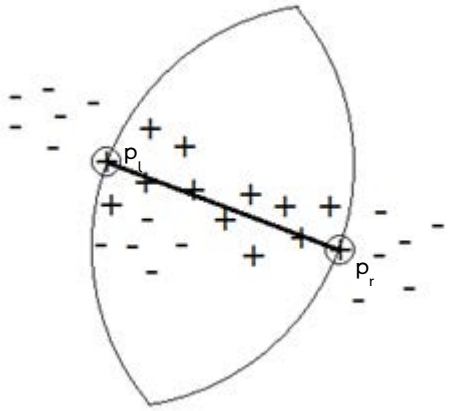

Figure 13 A positive lens

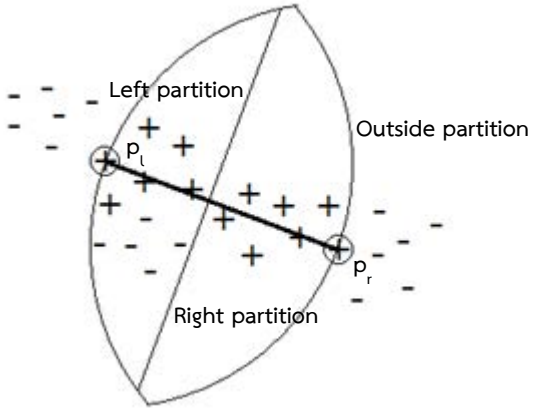

Figure 14 Data partitions

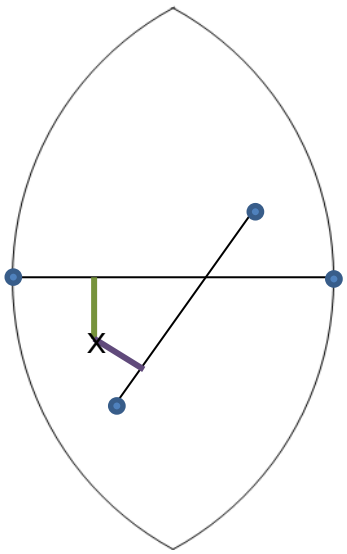

Figure 15 Projection without uncertain area

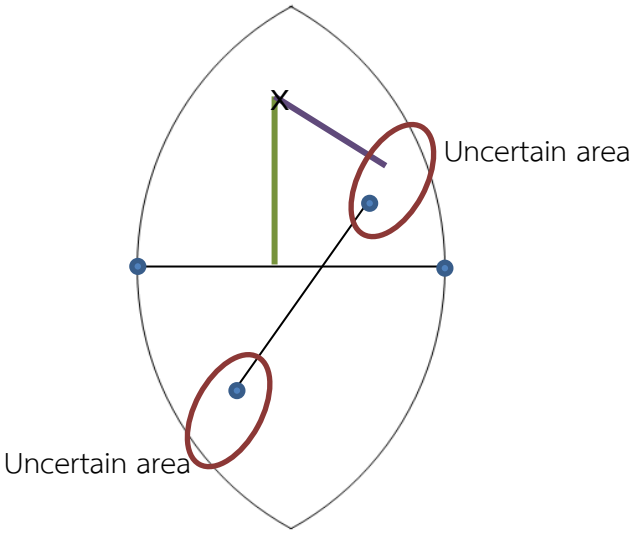

Figure 16 Projection with uncertain area

3.2 Algorithm

To build a tree, MAL algorithm applies a simple decision tree algorithm. In Figure 17, the modification of the *FindBestSplit()* function where *D* is the input dataset and *poleType* is the parameter that identifies the lens type which can either be positive or negative. First, it finds the best split of discrete attributes (Line 2). Next, the best positive lens or the best negative lens (Line 3) is determined based on the parameter *poleType*. Then, the algorithm will select the best split from the candidates according to their information scores (Line 4).

In Figure 18, the *FindBestSplitLens()* function shows the way to find the best split lens. First, the algorithm determines the farthest pair depending on the parameter *poleType* (Line 2). In addition, the positive lens algorithm or posMAL will use the *poleType* as the positive class; on the other hand, the negative lens algorithm or negMAL will use the *poleType* as the negative class. The lens will be used to partition the data into two sets: the inside lens and the outside lens (Line 3-4). Next, the algorithm will create the core vector (Line 5), the projection values of the inside lens instances onto the core vector (Line 6) and the sorted middle values of the projection values (Line 7). Each middle value is used to split the data into left and right partitions, and then the algorithm will calculate the information score (Line 8-11). Afterward, the algorithm will choose the middle value that has the highest information score and will return this value with the core vector (Line 12-17).

```
1. FindBestSplit(D, poleType)
2.       bestDis = FindBestSplitDisc(D)
3.       bestLens = FindBestSplitLens(D, poleType)
5.       return Max(bestDis.score, bestLens.score)
```

Figure 17 Find the best split function in MAL algorithm

```
1. FindBestSplitLens(D, poleType)
2.      pair = the farthest pair of the given pole type poleType
3.      D_in = { x|x is an instance that stay inside lens }
4.      D_out = { x|x is an instance that stay outside lens }
5.      cv = the vector from pair.p1 to pair.p2
6.      inProj = { v|v is a projection values of D_in that stay in the
lens onto the core vector cv }
7.      T = { t|t = (v_i + v_i+1)/2 where v_1, v_2, ..., v_n are the ordered
distinct values of projection values inProj }
8.      foreach t in T do
9.          D_left = { x|x is an inside lens instance that has the
magnitude less than or equal t }
10.         D_right = { x|x is an inside lens instance that has the
magnitude greater than t }
11.         score = the information score of D
12.         if(score > maxScore) then
13.             maxScore = score
14.             bestThres = t
15.         end if
16.     end for
17.     return { coreVector:cv, thres:bestThres, score:maxScore}
```

Figure 18 Find the best split lens function in MAL algorithm

CHAPTER IV

EXPERIMENTS AND RESULTS

In this chapter, we will show the results of the aforementioned performance measures in chapter 2 with the classification techniques: k nearest neighbor, naïve Bayes, SVM, logistic regression, as well as C4.5 algorithms.

4.1 Experimental environment

The MAL algorithm is written using Python language with Xcode IDE. Orange library [20] is a data mining library that is used to analyze, visualize, preprocess, and sample the datasets. It also provides many objects and utilities to easily manipulate and extend decision tree object as well as provides other classifiers that are used in this experiment. The experiment runs under Mac OSX system with 1.6 GHz Intel CPU and 4 GB of RAM.

4.2 Dataset descriptions

In this section, we will explain the datasets for our experiments. The datasets for classification are obtained from UCI Machine Learning Repository [21] that have the specified characteristics, including having multivariate, numeric and mixed attribute types, having less than 40 attributes, and having a number of instances between 100 and 1000 as shown in Table 4. There are a total of 24 datasets that satisfy with the conditions mentioned above, and are used in the experiments.

Table 4 UCI datasets

Name	Attribute Type	#instances	#attributes	#classes
Blood Transfusion Service Center	Real	748	5	2
Breast Cancer Wisconsin (Diagnostic)	Real	569	32	2
Breast Cancer Wisconsin (Original)	Integer	699	10	2
Breast Cancer Wisconsin (Prognostic)	Real	198	34	2
Breast Tissue	Real	106	10	6
Ecoli	Integer	336	8	8
Glass Identification	Real	214	10	7
Haberman's Survival	Integer	306	3	2
ILPD (Indian Liver Patient Dataset)	Integer, Real	583	10	2
Ionosphere	Integer, Real	351	34	2
Iris	Real	150	4	3
Mammographic Mass	Integer	961	6	2
Pima Indians Diabetes	Integer, Real	768	8	2
Seeds	Real	210	7	3
Vertebral Column	Real	310	6	2
Wine	Integer, Real	178	13	3
Credit Approval	Categorical, Integer, Real	690	15	2
Cylinder Bands	Categorical, Integer, Real	512	39	2
Dermatology	Categorical, Integer	366	33	6
Hepatitis	Categorical, Integer, Real	155	19	2
Horse Colic	Categorical, Integer, Real	368	27	2
Statlog (Australian Credit Approval)	Categorical, Integer, Real	690	14	2
Statlog (Heart)	Categorical, Real	270	13	2
Teaching Assistant Evaluation	Categorical, Integer	151	5	3

- **Blood transfusion service center**

Blood transfusion service center dataset [22] is data of blood donation collected from the blood donors in Hsin-Chu City, Taiwan in March 2007. It consists of 748 instances and 5 attributes with no missing value.

- **Breast Cancer Wisconsin (Diagnostic)**

Breast cancer Wisconsin (diagnostic) dataset [23], [24], [25], [26], [27], [28] collected the attributes of cell nuclei of a breast mass extracted from a digitized image using a fine needle aspirate (FNA). It consists of 569 instances and 32 attributes with no missing value. The ID attribute will not be accounted in our computation; therefore, there will be only 31 attributes remaining.

- **Breast Cancer Wisconsin (Original)**

Breast cancer Wisconsin (original) dataset [29] was originally collected chronologically from Dr. Wolberg's clinical cases from January 1989 to July 1992. All of the attributes were normalized and scaled with a value from 1 to 10. It consists of 699 instances and 11 attributes with 16 missing values. The sample code number attribute will not be accounted in our computation; therefore, there will be only 10 attributes remaining

- **Breast Cancer Wisconsin (Prognostic)**

Breast cancer Wisconsin (prognostic) dataset [24], [26], [27], [30], [31] was a follow-up data gathered from the breast cancer cases of Dr.Wolberg's patients. Only the cases with invasive breast cancer and no evidence of distant metastases at the time of diagnosis will be added to this dataset, with the data dated back in 1984. It consists of 198 instances and 35 attributes with 4 missing values. The sample code number attribute will not be accounted in our computation; therefore, there will be only 34 attributes remaining.

- **Breast Tissue**

Breast tissue dataset [32], [33] represented the impedance measurements of the breast tissue that were newly excised. It consists of 106 instances and 10 attributes with no missing value.

- **Ecoli**

Ecoli dataset [34], created by Kenta Nakai, comprises of the attributes that are used to predict the localization site of protein. It consists of 336 instances and 8 attributes with no missing value. The name attribute will not be accounted in our computation; therefore, there will be only 7 attributes remaining.

- **Glass Identification**

Glass identification [35] was first used by Vina in rule-based system in order to determine whether the glass was a type of “float” glass or not. It consists of 11 attributes with no missing value. The id attribute will not be accounted in our computation; there will be only 10 attributes remaining.

- **Haberman's Survival**

The dataset [36], [37], [38] keeps the records of patients who had survived on surgery for breast cancer at the University of Chicago’s Billings Hospital between 1958 and 1970. It consists of 4 attributes with no missing value.

- **ILPD (Indian Liver Patient Dataset)**

ILPD dataset [39], [40] was gathered from 416 liver patient records and 167 non liver patient records in the north east of Andhra Pradesh, India. In this dataset, 441 of the patient records are male; the remaining 167 are female patient records. It consists of 11 attributes with no missing value.

- **Ionosphere**

Ionosphere dataset [41] contains the radar data based in Goose Bay, Labrador. The data was assembled by the system with a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts, targeting at the free electrons in the ionosphere. Those with evidence of some type of structure in the ionosphere were identified as “Good” radar returns, while those that do not were “Bad” returns. It consists of 35 attributes with no missing value.

- **Iris**

Iris dataset [42], [43] [44] [45] contains 3 classes of types of Iris plant, where each class has a total of 50 instances. It is the dataset best known in the pattern recognition literature. It consists of 5 attributes with no missing value.

- **Mammographic Mass**

The dataset of Mammographic Mass [46] was collected from 516 benign and 445 malignant masses at the Institute of Radiology of the University Erlangen-Nuremberg during the year 2003 to 2006. Its feature is the use to discriminate benign and malignant mammographic masses based on BI-RADS attributes together with the age of the patient. It consists of 6 attributes with 131 missing values out of 961 instances.

- **Pima Indians Diabetes**

Pima Indian Diabetes dataset [47] was conducted based on the data from Pima Indian female that were 21 years old or older, with a total of 768 instances. It consists of 9 attributes with no missing value.

- **Seeds**

Seeds dataset [48] focused on three verities of wheat, including Kama, Rosa and Canadian, which a soft X-ray technique was used to visualize the internal kernel structure. The data was gathered at the Institute of Agrophysics of the

Polish Academy of Sciences in Lublin, and can be used for the purpose of classification and cluster analysis. It consists of 7 attributes with no missing value.

- **Vertebral Column**

Dr. Henrique da mota record the dataset during a medical residence period in the Group of Applied Research in Orthopaedics (GARO), Lyon, France [49], [50], [51]. The data were kept into two classes: Normal (100 patients) and Disk Hernia and Spondylolisthesis were labeled as Abnormal (210 patients). It consists of 7 attributes with no missing value.

- **Wine**

Three types of wines were defined based on a chemical analysis of wines in the same region in Italy in there different cultivars [52], [53]. There are initially 30 attributes, but in this dataset only 13 attributes were used with no missing value.

- **Credit Approval**

The file was submitted by Quilan [54], [55]. All attribute names and values had been changed to meaningless due to the confidential issue. There are 16 attributes with 37 missing values out of 690 instances.

- **Cylinder Bands**

The mitigating process delays known as “cylinder bands” in rotogravure printing was published with the decision tree classifier by Evans, B and al et [56]. It consists of 19 attributes with no missing value.

- **Dermatology**

The real problem in dermatology is the little differences in the diagnosis of the erythemato-squamous diseases. The diseases in this group are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris. The patients were fist evaluated clinically with 12 attributes and then 22 attributes were sampled histopathologically by an analysis of the

samples under a microscope [57]. The dataset consists of 35 attributes with 8 missing values out of 366 instances.

- **Hepatitis**

Hepatitis dataset was published by G. Gong [58], [59]. It consists of 20 attributes with 75 missing values out of 155 instances.

- **Horse Colic**

Horse Colic dataset is a study on colic in horses based on 368 instances which can be divided into 300 training instances and 68 test instances. It consists of 23 attributes with 361 missing values.

- **Statlog (Australian Credit Approval)**

Statlog (Australian Credit Approval) [54], [55] represents 690 instances regarding the credit card applications. It consists of 15 attributes with no missing value.

- **Statlog (Heart)**

Statlog (Heart) dataset is a dataset for a heart disease, with 270 instances collected. It consists of 13 attributes with no missing value.

- **Teaching Assistant Evaluation**

Teaching Assistant Evaluation dataset [60], [61] obtained the teaching performance evaluations from 151 teaching assistants at the Department of Statistics of the University of Wisconsin-Madison. The scores were rated over three regular semesters and two summer semesters and were categorized as “low”, “medium” and “high”. It consists of 6 attributes with no missing value.

4.3 Experimental design

Our thesis uses the holdout sampling technique to partition a dataset into two disjoint sets, called the training set and the test set. There are four

proportions between training set and the test set which are 60:40, 70:30, 80:20 and 90:10. To validate the datasets, each dataset is randomly sampled thirty times. For each time, the test set is applied to all classifiers including: posMAL, negMAL, C4.5, k nearest neighbor, naïve Bayes, SVM, as well as logistic regression and then the corresponding test set is used to evaluate each classifier. The performance measures in this experiment are accuracy, sensitivity, specificity, area under ROC curve (AUC), information score, F-measure, precision, recall, brier score, and Matthews correlation coefficient (MCC) (see also Appendix A). Afterward, we perform a paired t-test for each measure on posMAL against other classifiers (see also Appendix B) and negMAL against other classifiers (see also Appendix C). For each paired t-test, the null and alternative hypotheses are the followings.

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

where μ_1 is posMAL or negMAL mean and μ_2 is naïve Bayes, C4.5, KNN, or logistic regression mean. In this thesis, the significance level is equal to 0.025. Hence, if two-tailed probability or $p < 0.025$, H_0 is rejected which means that there is a difference in means across the paired observation.

Furthermore, we preprocess the missing value by the k nearest neighbor model as well as convert the datasets that have more than 2 classes by using one against all method since our algorithm only supports binary class datasets.

4.4 Experimental results

In this section, we compare the accuracy based on a paired t-test of our models: posMAL and negMAL with other classifiers.

- **The results of the paired t-test of accuracy**

Table 5 and Table 6 summarize the average value of accuracy for posMAL and negMAL for each dataset and each classifier. μ_1 is the average value of accuracy for posMAL or negMAL, μ_2 is the average value of accuracy for the other algorithms. The difference between the average value of accuracy of posMAL or negMAL and the other classifiers is shown as Δ, which the positive value of Δ

means that our algorithm is superior to the others and the negative value of Δ means that the other algorithms are superior to ours.

In the case of posMAL in Table 5, from 120 cases, we found that posMAL has better accuracy value of 30 cases, while the other 88 cases are worse than the other algorithms and 2 cases are equal. In the case of negMAL in Table 6, 28 cases have better accuracy value, 90 cases have worse accuracy value and the remaining 2 cases are equal. However, the Δ merely represents the performance of our classifier compared to the others; the paired t-tests are then applied to the accuracy as well as the other performance measures to further explain whether the average accuracy result of Δ and the other performance measure is significant or not. (See also Appendix B and C).

Table 7 and Table 8 summarize the average p value of accuracy for posMAL and negMaL for each dataset and each classifier. As mentioned earlier, the positive number of Δμ or the mean difference shows that our classifier is better than the others; moreover, the average p value of less than or equal to 0.025 is statistically considered as significantly better. On the other hand, those with negative Δμ shows that other classifiers are better than ours; and the p value of less than or equal to 0.025 shows that the other classifiers are considered to be significantly better.

From Table 7, it shows that posMAL is considered to be significantly better than the other classifiers in 8 datasets, including Blood Transfusion Services Center, Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis and Horse Colic. PosMAL is considered to be worse than the other classifiers in 13 datasets, including Breast Cancer Wisconsin (Diagnostic), Breast Tissue, Ionosphere, Iris, Mammographic Mass, Pima Indian Diabetes, Seeds, Vertebral Column, Wine, Cylinder Bands, Statlog (Australian Credit Approval), Statlog (Heart) and Teaching Assistant Evaluation. The result of the remaining 3 datasets; Ecoli, Glass Identification and Dermatology; is indifferent.

From Table 8, it shows that negMAL is considered to be significantly better than the other classifiers in 7 datasets, including Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis and Horse Colic. NegMAL is considered to be worse than the other classifiers in 15 datasets, including Blood Transfusion Services Center, Breast Cancer Wisconsin (Diagnostic), Breast Tissue, Glass Identification, Ionosphere, Iris, Mammographic Mass, Pima Indian Diabetes, Seeds, Vertebral Column, Wine, Cylinder Bands, Statlog (Australian Credit Approval), Statlog (Heart) and Teaching Assistant Evaluation. The result of the remaining 2 datasets; Ecoli and Dermatology; does not yield a significant different result.

Table 5 Average values of accuracy - posMAL

DB		*μ1*	*μ2*	*Δ*
01.Blood Transfusion Service Center				
	Bayes	0.761040	0.756145	+
	C4.5	0.761040	0.716333	+
	KNN	0.761040	0.771098	-
	LogR	0.761040	0.771988	-
	SVM	0.761040	0.760635	+
02.Breast Cancer Wisconsin (Diagnostic)				
	Bayes	0.900459	0.944237	-
	C4.5	0.900459	0.917288	-
	KNN	0.900459	0.970882	-
	LogR	0.900459	0.957627	-
	SVM	0.900459	0.949281	-
03.Breast Cancer Wisconsin (Original)				
	Bayes	0.961660	0.971151	-
	C4.5	0.961660	0.940744	+
	KNN	0.961660	0.966746	-
	LogR	0.961660	0.961319	+
	SVM	0.961660	0.963938	-
04.Breast Cancer Wisconsin (Prognostic)				
	Bayes	0.766231	0.674793	+
	C4.5	0.766231	0.661255	+
	KNN	0.766231	0.770811	-
	LogR	0.766231	0.789041	-
	SVM	0.766231	0.769040	-
05.Breast Tissue				
	Bayes	0.835794	0.831934	+
	C4.5	0.835794	0.869280	-
	KNN	0.835794	0.897246	-
	LogR	0.835794	0.880589	-
	SVM	0.835794	0.817381	+
06.Ecoli				
	Bayes	0.910602	0.564055	+
	C4.5	0.910602	0.937836	-
	KNN	0.910602	0.963445	-
	LogR	0.910602	0.954580	-
	SVM	0.910602	0.917376	-
07.Glass Identification				
	Bayes	0.837290	0.864103	-
	C4.5	0.837290	0.880781	-
	KNN	0.837290	0.894196	-
	LogR	0.837290	0.862751	-
	SVM	0.837290	0.851039	-
08.Haberman's Survival				
	Bayes	0.739056	0.753571	-
	C4.5	0.739056	0.674966	+
	KNN	0.739056	0.713445	+
	LogR	0.739056	0.748116	-
	SVM	0.739056	0.739119	-
09.ILPD (Indian Liver Patient Dataset)				
	Bayes	0.711085	0.663890	+
	C4.5	0.711085	0.675194	+
	KNN	0.711085	0.666738	+
	LogR	0.711085	0.717836	-
	SVM	0.711085	0.704375	+
10.Ionosphere				
	Bayes	0.666010	0.876457	-
	C4.5	0.666010	0.922461	-
	KNN	0.666010	0.852640	-
	LogR	0.666010	0.874709	-
	SVM	0.666010	0.921118	-
11.Iris				
	Bayes	0.883426	0.933256	-
	C4.5	0.883426	0.958611	-
	KNN	0.883426	0.966034	-
	LogR	0.883426	0.895586	-
	SVM	0.883426	0.958966	-
12.Mammographic Mass				
	Bayes	0.547873	0.816654	-
	C4.5	0.547873	0.763267	-
	KNN	0.547873	0.802567	-
	LogR	0.547873	0.826154	-
	SVM	0.547873	0.795975	-
13.Pima Indians Diabetes				
	Bayes	0.648902	0.748026	-
	C4.5	0.648902	0.689179	-
	KNN	0.648902	0.738032	-
	LogR	0.648902	0.771332	-
	SVM	0.648902	0.765576	-
14.Seeds				
	Bayes	0.839242	0.929056	-
	C4.5	0.839242	0.944940	-
	KNN	0.839242	0.954343	-
	LogR	0.839242	0.974217	-
	SVM	0.839242	0.944378	-
15.Vertebral Column				
	Bayes	0.729017	0.768168	-
	C4.5	0.729017	0.791891	-
	KNN	0.729017	0.797939	-
	LogR	0.729017	0.834409	-
	SVM	0.729017	0.840367	-
16.Wine				
	Bayes	0.773624	0.981396	-
	C4.5	0.773624	0.946733	-
	KNN	0.773624	0.970629	-
	LogR	0.773624	0.981089	-
	SVM	0.773624	0.937998	-
17.Credit Approval				
	Bayes	0.856401	0.855616	+
	C4.5	0.856401	0.796548	+
	KNN	0.856401	0.864543	-
	LogR	0.856401	0.850513	+
	SVM	0.856401	0.856401	0
18.Cylinder Bands				
	Bayes	0.599580	0.688272	-
	C4.5	0.599580	0.661908	-
	KNN	0.599580	0.670010	-
	LogR	0.599580	0.628279	-
	SVM	0.599580	0.677803	-
19.Dermatology				
	Bayes	0.938471	0.976004	-
	C4.5	0.938471	0.962750	-
	KNN	0.938471	0.987546	-
	LogR	0.938471	0.966011	-
	SVM	0.938471	0.879687	+
20.Hepatitis				
	Bayes	0.834253	0.849426	-
	C4.5	0.834253	0.785950	+
	KNN	0.834253	0.822060	+
	LogR	0.834253	0.821010	+
	SVM	0.834253	0.775438	+
21.Horse Colic				
	Bayes	0.819970	0.778928	+
	C4.5	0.819970	0.765090	+
	KNN	0.819970	0.792247	+
	LogR	0.819970	0.779167	+
	SVM	0.819970	0.839717	-
22.Statlog (Australian Credit Approval)				
	Bayes	0.853915	0.860205	-
	C4.5	0.853915	0.801198	+
	KNN	0.853915	0.865328	-
	LogR	0.853915	0.852838	+
	SVM	0.853915	0.853915	0
23.Statlog (Heart)				
	Bayes	0.551192	0.808745	-
	C4.5	0.551192	0.717078	-
	KNN	0.551192	0.797814	-
	LogR	0.551192	0.828369	-
	SVM	0.551192	0.833539	-
24.Teaching Assistant Evaluation				
	Bayes	0.665493	0.686416	-
	C4.5	0.665493	0.744362	-
	KNN	0.665493	0.711739	-
	LogR	0.665493	0.666311	-
	SVM	0.665493	0.683294	-

Table 6 Average values of accuracy - negMAL

DB	$\mu 1$	$\mu 1$	Δ
01.Blood Transfusion Service Center			
Bayes	0.760126	0.756145	+
C4.5	0.760126	0.716333	+
KNN	0.760126	0.771098	-
LogR	0.760126	0.771988	-
SVM	0.760126	0.760635	-
02.Breast Cancer Wisconsin (Diagnostic)			
Bayes	0.898526	0.944237	-
C4.5	0.898526	0.917288	-
KNN	0.898526	0.970882	-
LogR	0.898526	0.957627	-
SVM	0.898526	0.949281	-
03.Breast Cancer Wisconsin (Original)			
Bayes	0.956300	0.971151	-
C4.5	0.956300	0.940744	+
KNN	0.956300	0.966746	-
LogR	0.956300	0.961319	-
SVM	0.956300	0.963938	-
04.Breast Cancer Wisconsin (Prognostic)			
Bayes	0.767291	0.674793	+
C4.5	0.767291	0.661255	+
KNN	0.767291	0.770811	-
LogR	0.767291	0.789041	-
SVM	0.767291	0.769040	-
05.Breast Tissue			
Bayes	0.837524	0.831934	+
C4.5	0.837524	0.869280	-
KNN	0.837524	0.897246	-
LogR	0.837524	0.880589	-
SVM	0.837524	0.817381	+
06.Ecoli			
Bayes	0.887421	0.564055	+
C4.5	0.887421	0.937836	-
KNN	0.887421	0.963445	-
LogR	0.887421	0.954580	-
SVM	0.887421	0.917376	-
07.Glass Identification			
Bayes	0.833606	0.864103	-
C4.5	0.833606	0.880781	-
KNN	0.833606	0.894196	-
LogR	0.833606	0.862751	-
SVM	0.833606	0.851039	-
08.Haberman's Survival			
Bayes	0.736373	0.753571	-
C4.5	0.736373	0.674966	+
KNN	0.736373	0.713445	+
LogR	0.736373	0.748116	-
SVM	0.736373	0.739119	-
09.ILPD (Indian Liver Patient Dataset)			
Bayes	0.712095	0.663890	+
C4.5	0.712095	0.675194	+
KNN	0.712095	0.666738	+
LogR	0.712095	0.717836	-
SVM	0.712095	0.704375	+
10.Ionosphere			
Bayes	0.658998	0.876457	-
C4.5	0.658998	0.922461	-
KNN	0.658998	0.852640	-
LogR	0.658998	0.874709	-
SVM	0.658998	0.921118	-
11.Iris			
Bayes	0.838488	0.933256	-
C4.5	0.838488	0.958611	-
KNN	0.838488	0.966034	-
LogR	0.838488	0.895586	-
SVM	0.838488	0.958966	-
12.Mammographic Mass			
Bayes	0.547386	0.816654	-
C4.5	0.547386	0.763267	-
KNN	0.547386	0.802567	-
LogR	0.547386	0.826154	-
SVM	0.547386	0.795975	-
13.Pima Indians Diabetes			
Bayes	0.649862	0.748026	-
C4.5	0.649862	0.689179	-
KNN	0.649862	0.738032	-
LogR	0.649862	0.771332	-
SVM	0.649862	0.765576	-
14.Seeds			
Bayes	0.799625	0.929056	-
C4.5	0.799625	0.944940	-
KNN	0.799625	0.954343	-
LogR	0.799625	0.974217	-
SVM	0.799625	0.944378	-
15.Vertebral Column			
Bayes	0.683333	0.768168	-
C4.5	0.683333	0.791891	-
KNN	0.683333	0.797939	-
LogR	0.683333	0.834409	-
SVM	0.683333	0.840367	-
16.Wine			
Bayes	0.777525	0.981396	-
C4.5	0.777525	0.946733	-
KNN	0.777525	0.970629	-
LogR	0.777525	0.981089	-
SVM	0.777525	0.937998	-
17.Credit Approval			
Bayes	0.856401	0.855616	+
C4.5	0.856401	0.796548	+
KNN	0.856401	0.864543	-
LogR	0.856401	0.850513	+
SVM	0.856401	0.856401	0
18.Cylinder Bands			
Bayes	0.593904	0.688272	-
C4.5	0.593904	0.661908	-
KNN	0.593904	0.670010	-
LogR	0.593904	0.628279	-
SVM	0.593904	0.677803	-
19.Dermatology			
Bayes	0.938471	0.976004	-
C4.5	0.938471	0.962750	-
KNN	0.938471	0.987546	-
LogR	0.938471	0.966011	-
SVM	0.938471	0.879687	+
20.Hepatitis			
Bayes	0.834253	0.849426	-
C4.5	0.834253	0.785950	+
KNN	0.834253	0.822060	+
LogR	0.834253	0.821010	+
SVM	0.834253	0.775438	+
21.Horse Colic			
Bayes	0.819970	0.778928	+
C4.5	0.819970	0.765090	+
KNN	0.819970	0.792247	+
LogR	0.819970	0.779167	+
SVM	0.819970	0.839717	-
22.Statlog (Australian Credit Approval)			
Bayes	0.853915	0.860205	-
C4.5	0.853915	0.801198	+
KNN	0.853915	0.865328	-
LogR	0.853915	0.852838	+
SVM	0.853915	0.853915	0
23.Statlog (Heart)			
Bayes	0.551620	0.808745	-
C4.5	0.551620	0.717078	-
KNN	0.551620	0.797814	-
LogR	0.551620	0.828369	-
SVM	0.551620	0.833539	-
24.Teaching Assistant Evaluation			
Bayes	0.664258	0.686416	-
C4.5	0.664258	0.744362	-
KNN	0.664258	0.711739	-
LogR	0.664258	0.666311	-
SVM	0.664258	0.683294	-

Table 7 paired t-test on accuracy - posMAL

DB	Δμ	p	sig
01.Blood Transfusion Service Center			
Bayes	0.006722	0.318380	
C4.5	0.046534	0.000000	+
KNN	-0.008230	0.068098	
LogR	-0.009121	0.037876	
SVM	0.002232	0.199553	
02.Breast Cancer Wisconsin (Diagnostic)			
Bayes	-0.039912	0.000006	-
C4.5	-0.012963	0.096133	
KNN	-0.066557	0.000000	-
LogR	-0.053302	0.000000	-
SVM	-0.044956	0.000000	-
03.Breast Cancer Wisconsin (Original)			
Bayes	-0.006270	0.063978	
C4.5	0.024137	0.000059	+
KNN	-0.001865	0.515264	
LogR	0.003562	0.244775	
SVM	0.000942	0.527138	
04.Breast Cancer Wisconsin (Prognostic)			
Bayes	0.090579	0.000004	+
C4.5	0.104117	0.000000	+
KNN	-0.005440	0.498136	
LogR	-0.023669	0.103934	
SVM	-0.003668	0.474360	
05.Breast Tissue			
Bayes	0.001806	0.074038	
C4.5	-0.035540	0.070139	
KNN	-0.063506	0.021363	-
LogR	-0.046849	0.125020	
SVM	0.016358	0.172220	
06.Ecoli			
Bayes	0.358827	0.057218	
C4.5	-0.014954	0.268671	
KNN	-0.040564	0.245772	
LogR	-0.031698	0.144367	
SVM	0.005505	0.179059	
07.Glass Identification			
Bayes	-0.025285	0.093067	
C4.5	-0.041964	0.070584	
KNN	-0.055378	0.064484	
LogR	-0.023933	0.054015	
SVM	-0.012222	0.188994	
08.Haberman's Survival			
Bayes	-0.010434	0.276179	
C4.5	0.068170	0.000000	+
KNN	0.029691	0.004564	+
LogR	-0.004980	0.301803	
SVM	0.004017	0.554174	
09.ILPD (Indian Liver Patient Dataset)			
Bayes	0.046589	0.000096	+
C4.5	0.035286	0.040344	
KNN	0.043741	0.000019	+
LogR	-0.007357	0.376423	
SVM	0.006104	0.185920	
10.Ionosphere			
Bayes	-0.207198	0.000000	-
C4.5	-0.253203	0.000000	-
KNN	-0.183381	0.000000	-
LogR	-0.205450	0.000000	-
SVM	-0.251859	0.000000	-
11.Iris			
Bayes	-0.036790	0.036185	
C4.5	-0.062145	0.005077	-
KNN	-0.069568	0.004191	-
LogR	0.000880	0.003955	+
SVM	-0.062500	0.007957	-
12.Mammographic Mass			
Bayes	-0.268609	0.000000	-
C4.5	-0.215222	0.000000	-
KNN	-0.254522	0.000000	-
LogR	-0.278109	0.000000	-
SVM	-0.247929	0.000000	-
13.Pima Indians Diabetes			
Bayes	-0.100521	0.000000	-
C4.5	-0.041674	0.000004	-
KNN	-0.090527	0.000000	-
LogR	-0.123826	0.000000	-
SVM	-0.118070	0.000000	-
14.Seeds			
Bayes	-0.092879	0.004880	-
C4.5	-0.108763	0.015116	-
KNN	-0.118166	0.000205	-
LogR	-0.138040	0.000003	-
SVM	-0.108201	0.008745	-
15.Vertebral Column			
Bayes	-0.016308	0.273626	
C4.5	-0.040031	0.044619	
KNN	-0.046080	0.009423	-
LogR	-0.082549	0.000008	-
SVM	-0.088508	0.000008	-
16.Wine			
Bayes	-0.211165	0.000000	-
C4.5	-0.176503	0.006452	-
KNN	-0.200398	0.000000	-
LogR	-0.210858	0.000000	-
SVM	-0.167768	0.002646	-
17.Credit Approval			
Bayes	0.000785	0.511023	
C4.5	0.059853	0.000000	+
KNN	-0.008142	0.177818	
LogR	0.005888	0.360724	
SVM	0.000000	0.662790	
18.Cylinder Bands			
Bayes	-0.085854	0.000000	-
C4.5	-0.059491	0.000001	-
KNN	-0.067593	0.000000	-
LogR	-0.025862	0.004211	-
SVM	-0.075386	0.000000	-
19.Dermatology			
Bayes	-0.037533	0.091478	
C4.5	-0.024279	0.155556	
KNN	-0.049075	0.132351	
LogR	-0.027540	0.128958	
SVM	0.058783	0.055203	
20.Hepatitis			
Bayes	-0.015172	0.258235	
C4.5	0.048303	0.007376	+
KNN	0.012193	0.334489	
LogR	0.013244	0.434106	
SVM	0.058816	0.000003	+
21.Horse Colic			
Bayes	0.041042	0.001190	+
C4.5	0.054880	0.000023	+
KNN	0.027723	0.011515	+
LogR	0.040803	0.008376	+
SVM	-0.019747	0.063515	
22.Statlog (Australian Credit Approval)			
Bayes	-0.006290	0.334724	
C4.5	0.052717	0.000000	+
KNN	-0.011413	0.023114	-
LogR	0.001077	0.202079	
SVM	0.000000	0.662790	
23.Statlog (Heart)			
Bayes	-0.257176	0.000000	-
C4.5	-0.165509	0.000000	-
KNN	-0.246245	0.000000	-
LogR	-0.276800	0.000000	-
SVM	-0.281970	0.000000	-
24.Teaching Assistant Evaluation			
Bayes	-0.020436	0.378772	
C4.5	-0.078382	0.019217	-
KNN	-0.045758	0.247333	
LogR	-0.000331	0.054958	
SVM	-0.017314	0.137910	

Table 8 paired t-test on accuracy - negMAL

DB	Δμ	p	sig
01.Blood Transfusion Service Center			
Bayes	0.003981	0.580559	
C4.5	0.043793	0.000019	+
KNN	-0.010972	0.014709	-
LogR	-0.011862	0.001224	-
SVM	-0.000510	0.256324	
02.Breast Cancer Wisconsin (Diagnostic)			
Bayes	-0.045712	0.000005	-
C4.5	-0.018762	0.123150	
KNN	-0.072356	0.000000	-
LogR	-0.059101	0.000000	-
SVM	-0.050755	0.000000	-
03.Breast Cancer Wisconsin (Original)			
Bayes	-0.014851	0.026773	
C4.5	0.015556	0.018600	+
KNN	-0.010446	0.132075	
LogR	-0.005020	0.425971	
SVM	-0.007639	0.278760	
04.Breast Cancer Wisconsin (Prognostic)			
Bayes	0.092498	0.000001	+
C4.5	0.106036	0.000000	+
KNN	-0.003521	0.325986	
LogR	-0.021750	0.088908	
SVM	-0.001749	0.212918	
05.Breast Tissue			
Bayes	0.005590	0.090736	
C4.5	-0.031756	0.130647	
KNN	-0.059721	0.008523	-
LogR	-0.043065	0.069523	
SVM	0.020143	0.373354	
06.Ecoli			
Bayes	0.323367	0.028424	
C4.5	-0.050415	0.195588	
KNN	-0.076024	0.165717	
LogR	-0.067158	0.103477	
SVM	-0.029955	0.420563	
07.Glass Identification			
Bayes	-0.030497	0.113094	
C4.5	-0.047175	0.005009	-
KNN	-0.060590	0.041355	
LogR	-0.029144	0.085093	
SVM	-0.017433	0.230660	
08.Haberman's Survival			
Bayes	-0.017197	0.028420	
C4.5	0.061407	0.000000	+
KNN	0.022928	0.004643	+
LogR	-0.011743	0.156730	
SVM	-0.002746	0.093389	
09.ILPD (Indian Liver Patient Dataset)			
Bayes	0.048205	0.000076	+
C4.5	0.036901	0.032855	
KNN	0.045357	0.000011	+
LogR	-0.005741	0.381775	
SVM	0.007720	0.082791	
10.Ionosphere			
Bayes	-0.217459	0.000000	-
C4.5	-0.263464	0.000000	-
KNN	-0.193642	0.000000	-
LogR	-0.215711	0.000000	-
SVM	-0.262120	0.000000	-
11.Iris			
Bayes	-0.094769	0.110640	
C4.5	-0.120123	0.003011	-
KNN	-0.127546	0.002777	-
LogR	-0.057099	0.002855	-
SVM	-0.120478	0.002568	-
12.Mammographic Mass			
Bayes	-0.269268	0.000000	-
C4.5	-0.215881	0.000000	-
KNN	-0.255181	0.000000	-
LogR	-0.278768	0.000000	-
SVM	-0.248589	0.000000	-
13.Pima Indians Diabetes			
Bayes	-0.098165	0.000000	-
C4.5	-0.039317	0.000004	-
KNN	-0.088171	0.000000	-
LogR	-0.121470	0.000000	-
SVM	-0.115714	0.000000	-
14.Seeds			
Bayes	-0.129431	0.000033	-
C4.5	-0.145315	0.000000	-
KNN	-0.154718	0.000000	-
LogR	-0.174592	0.000000	-
SVM	-0.144753	0.000005	-
15.Vertebral Column			
Bayes	-0.084834	0.000000	-
C4.5	-0.108557	0.000000	-
KNN	-0.114606	0.000000	-
LogR	-0.151075	0.000000	-
SVM	-0.157034	0.000000	-
16.Wine			
Bayes	-0.203871	0.000001	-
C4.5	-0.169209	0.027217	
KNN	-0.193104	0.000003	-
LogR	-0.203564	0.000000	-
SVM	-0.160474	0.082193	
17.Credit Approval			
Bayes	0.000785	0.511023	
C4.5	0.059853	0.000000	+
KNN	-0.008142	0.177818	
LogR	0.005888	0.360724	
SVM	0.000000	0.662790	
18.Cylinder Bands			
Bayes	-0.094367	0.000000	-
C4.5	-0.068004	0.000000	-
KNN	-0.076106	0.000000	-
LogR	-0.034375	0.000520	-
SVM	-0.083899	0.000000	-
19.Dermatology			
Bayes	-0.037533	0.091478	
C4.5	-0.024279	0.155556	
KNN	-0.049075	0.132351	
LogR	-0.027540	0.128958	
SVM	0.058783	0.055203	
20.Hepatitis			
Bayes	-0.015172	0.258235	
C4.5	0.048303	0.007376	+
KNN	0.012193	0.334489	
LogR	0.013244	0.434106	
SVM	0.058816	0.000003	+
21.Horse Colic			
Bayes	0.041042	0.001190	+
C4.5	0.054880	0.000023	+
KNN	0.027723	0.011515	+
LogR	0.040803	0.008376	+
SVM	-0.019747	0.063515	
22.Statlog (Australian Credit Approval)			
Bayes	-0.006290	0.334724	
C4.5	0.052717	0.000000	+
KNN	-0.011413	0.023114	-
LogR	0.001077	0.202079	
SVM	0.000000	0.662790	
23.Statlog (Heart)			
Bayes	-0.257124	0.000000	-
C4.5	-0.165458	0.000000	-
KNN	-0.246193	0.000000	-
LogR	-0.276749	0.000000	-
SVM	-0.281919	0.000000	-
24.Teaching Assistant Evaluation			
Bayes	-0.022158	0.339823	
C4.5	-0.080104	0.007047	-
KNN	-0.047481	0.095590	
LogR	-0.002053	0.096643	
SVM	-0.019036	0.317185	

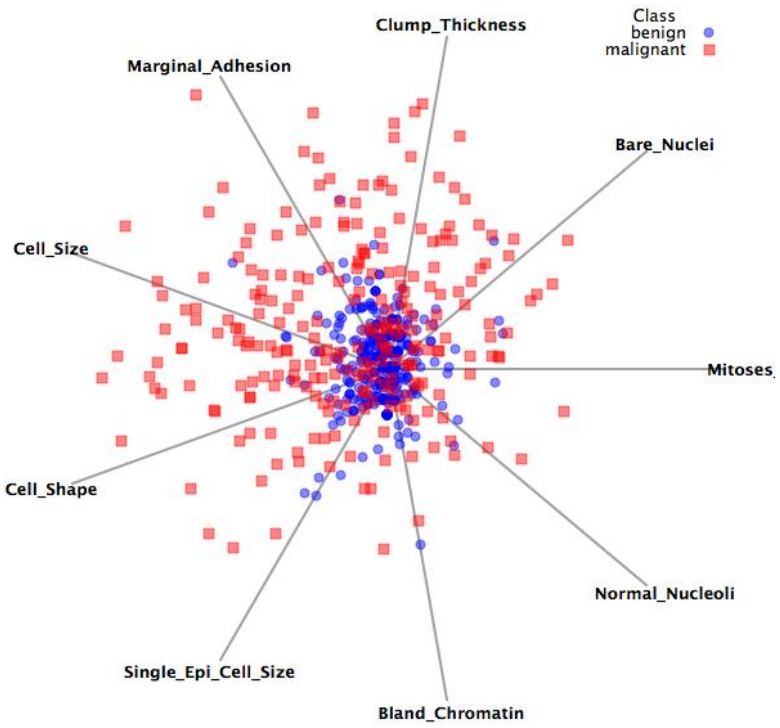

Figure 19 Linear projection - Breast Cancer Wisconsin (Original) dataset

First, we select the Breast Cancer Wisconsin (Original) dataset to discuss the result of posMAL. Even though the average p value of accuracy of naïve Bayes, k nearest neighbor, logistic regression, or SVM shown in Table 7 are indifferent from posMAL, we choose this dataset due to the fact that the result compared with C4.5 is significantly better. The linear projection of this dataset is shown in Figure 19 which the circles represents the positive instances and the squares represents the negative instances. It obviously displays a cluster in the data, where the positive instances stay closely in the center of the cluster; as a result, posMAL can effectively partition negative instances at the beginning phase.

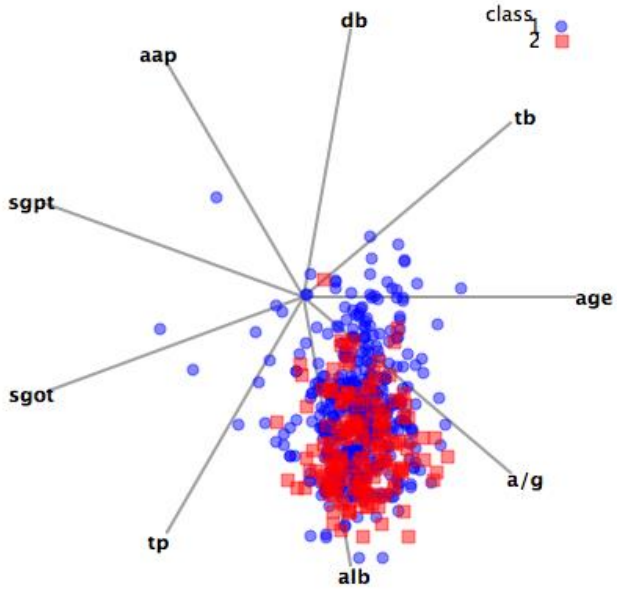

Figure 20 Linear projection - ILPD dataset (Indian Liver Patient Dataset)

Likewise, the ILPD dataset is selected from the Table 8 as the average p value of negMAL is significantly better than Bayes, and C4.5. The result is equivalence to k nearest neighbor, logistic regression, and SVM. The negative instances in Figure 20 assemble in the center of the cluster of the data. Therefore, negMAL which is used partition the positive instances in the first levels will also yield a desirable result.

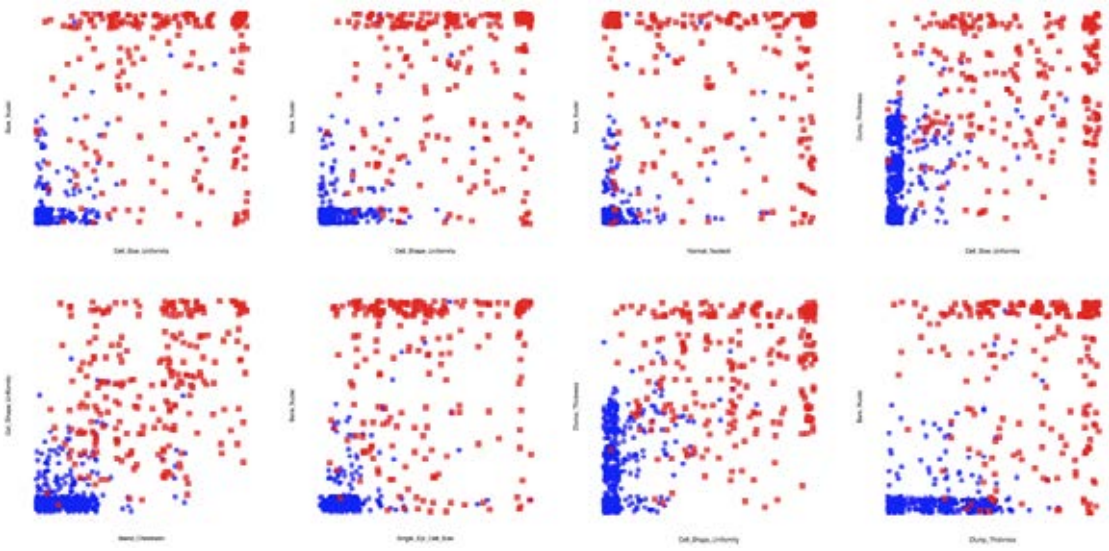

Figure 21 Scatter plots - Breast Cancer Wisconsin (Original) dataset

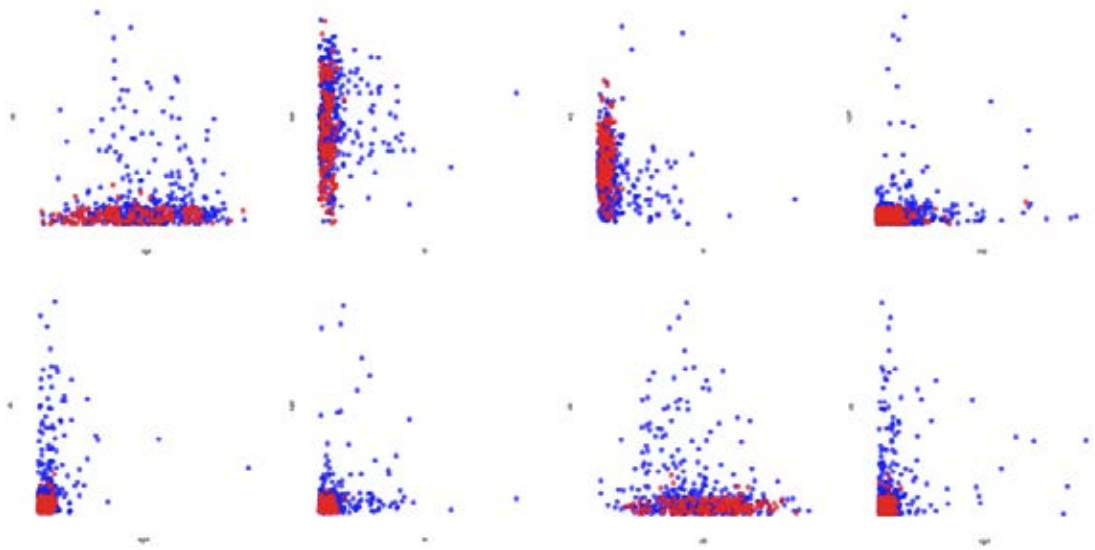

Figure 22 Scatter plots - ILPD dataset (Indian Liver Patient Dataset)

The scatter plots of the Breast Cancer Wisconsin (Original) dataset and the ILPD dataset are shown in Figure 21 and Figure 22, respectively. In each axis, we can clearly see that the positive instances of Breast Cancer Wisconsin (Original) dataset mostly locate close to the center. The negative instances of ILPD dataset stay closely in a group as well. Hence, posMAL and negMAL are able to partition these datasets well.

We notice that for the datasets that we select, Breast Cancer Wisconsin (Original) dataset and ILPD dataset have the similar pattern; where the data stays closely in a group.

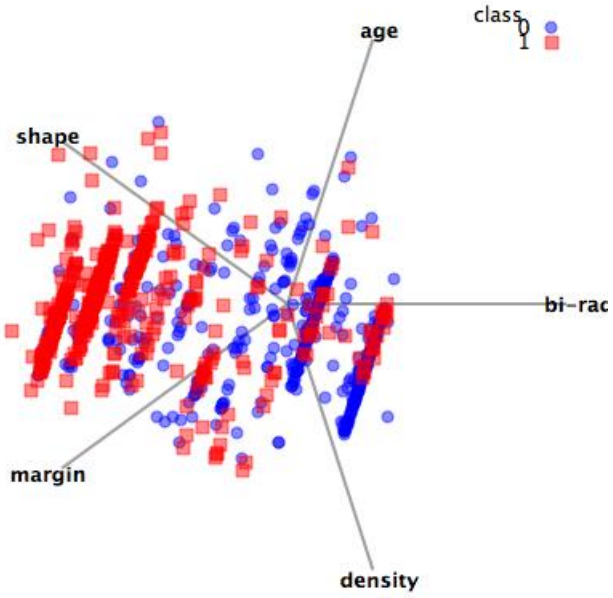

Figure 23 Linear projection - Mammographic Mass dataset

The Mammographic Mass dataset, on the other hand, demonstrates that the data scatters with no specific direction like the first two datasets. Also, we can see that the data do not stay together in a group. The linear projection is in line with the average p value of accuracy for both posMAL and negMAL in Table 7 and Table 8. In Table 7 of posMAL, it shows that the result of this dataset is inferior to naïve Bayes, C4.5, k nearest neighbor, logistic regression and SVM with the average p value of accuracy of 0. In addition, for negMAL in Table 8, it shows that the result of this dataset is inferior to naïve Bayes, C4.5, k nearest neighbor, logistic regression and SVM with the average p value of accuracy of 0.

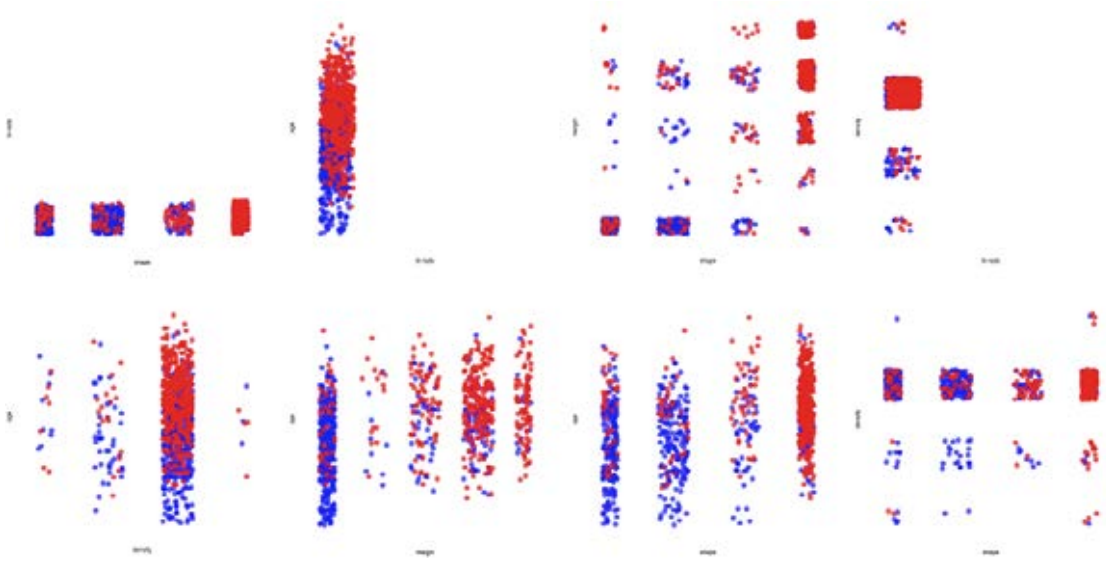

Figure 24 Scatter plots - Mammographic Mass dataset

Figure 24 is the scatter plots of the Mammographic Mass dataset. The data on each axis is different from that of the Breast Cancer Wisconsin (Original) dataset and the ILPD dataset that we mentioned earlier. In this dataset, the positive instances can hardly be separate from the negative instances as they are overlapped. Due to this fact, it is difficult for MAL to effectively partition the dataset.

Table 9 The average accuracy C4.5 and MAL

Average of CA	Learner		
DB	C4.5	NegMAL	PosMAL
01.Blood Transfusion Service Center	0.716333	0.760126	**0.762867**
02.Breast Cancer Wisconsin (Diagnostic)	**0.917288**	0.898526	0.904325
03.Breast Cancer Wisconsin (Original)	0.940744	0.9563	**0.964881**
04.Breast Cancer Wisconsin (Prognostic)	0.661255	**0.767291**	0.765372
05.Breast Tissue	**0.86928**	0.837524	0.83374
06.Ecoli	**0.937836**	0.887421	0.922881
07.Glass Identification	**0.880781**	0.833606	0.838817
08.Haberman's Survival	0.674966	0.736373	**0.743136**
09.ILPD (Indian Liver Patient Dataset)	0.675194	0.712095	**0.710479**
10.Ionosphere	**0.922461**	0.658998	0.669259
11.Iris	**0.958611**	0.838488	0.896466
12.Mammographic Mass	**0.763267**	0.547386	0.548045
13.Pima Indians Diabetes	**0.689179**	0.649862	0.647505
14.Seeds	**0.94494**	0.799625	0.836177
15.Vertebral Column	**0.791891**	0.683333	0.751859
16.Wine	**0.946733**	0.777525	0.770231
17.Credit Approval	0.796548	**0.856401**	**0.856401**
18.Cylinder Bands (Continuous attributes)	**0.661908**	0.593904	0.602418
19.Dermatology	**0.96275**	0.938471	0.938471
20.Hepatitis	0.78595	**0.834253**	**0.834253**
21.Horse Colic	0.76509	**0.81997**	**0.81997**
22.Statlog (Australian Credit Approval)	0.801198	**0.853915**	**0.853915**
23.Statlog (Heart)	**0.717078**	0.55162	0.551569
24.Teaching Assistant Evaluation	**0.744362**	0.664258	0.66598

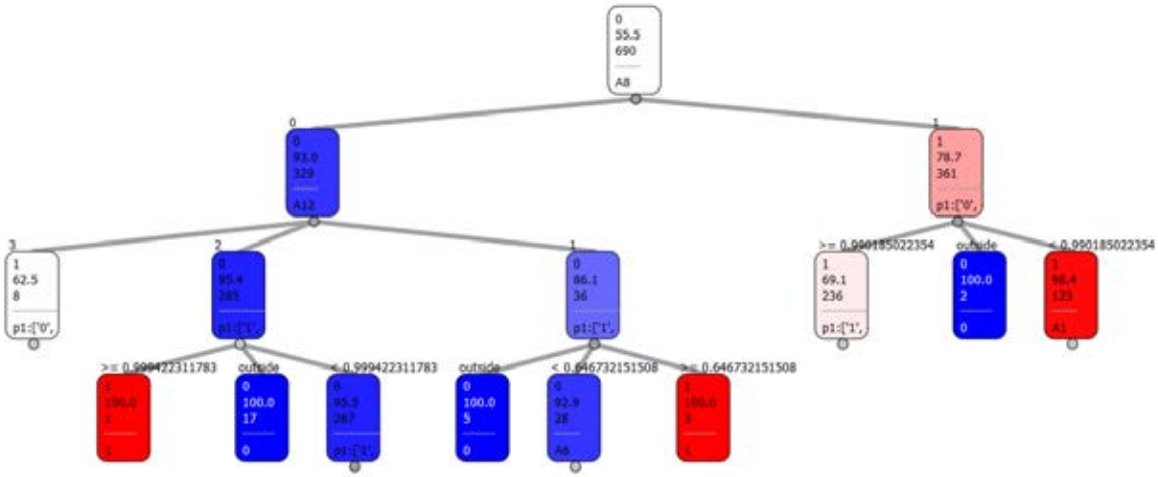

Figure 25 The tree of negMAL - Statlog (Australian Credit Approval) dataset

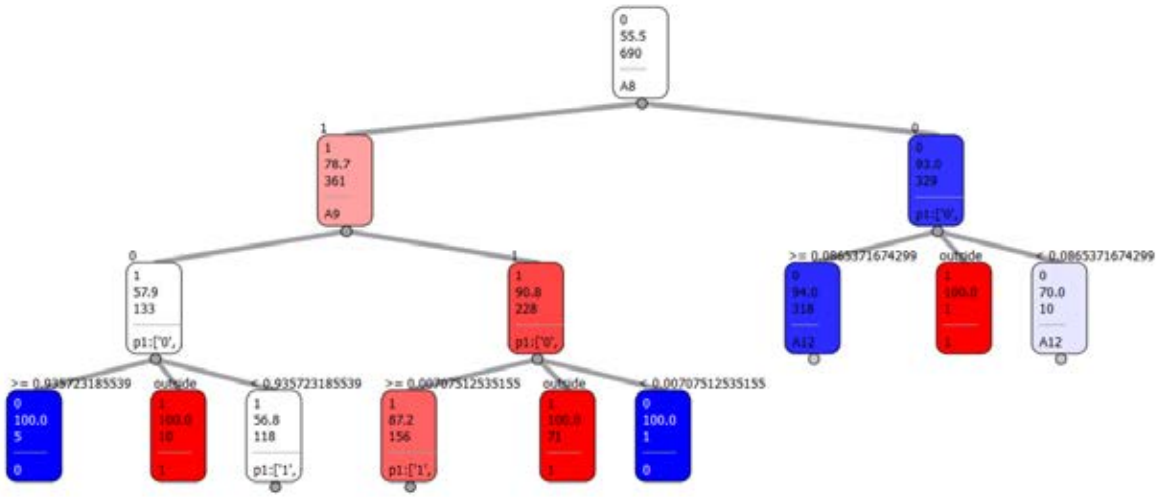

Figure 26 The tree of posMAL - Statlog (Australian Credit Approval) dataset

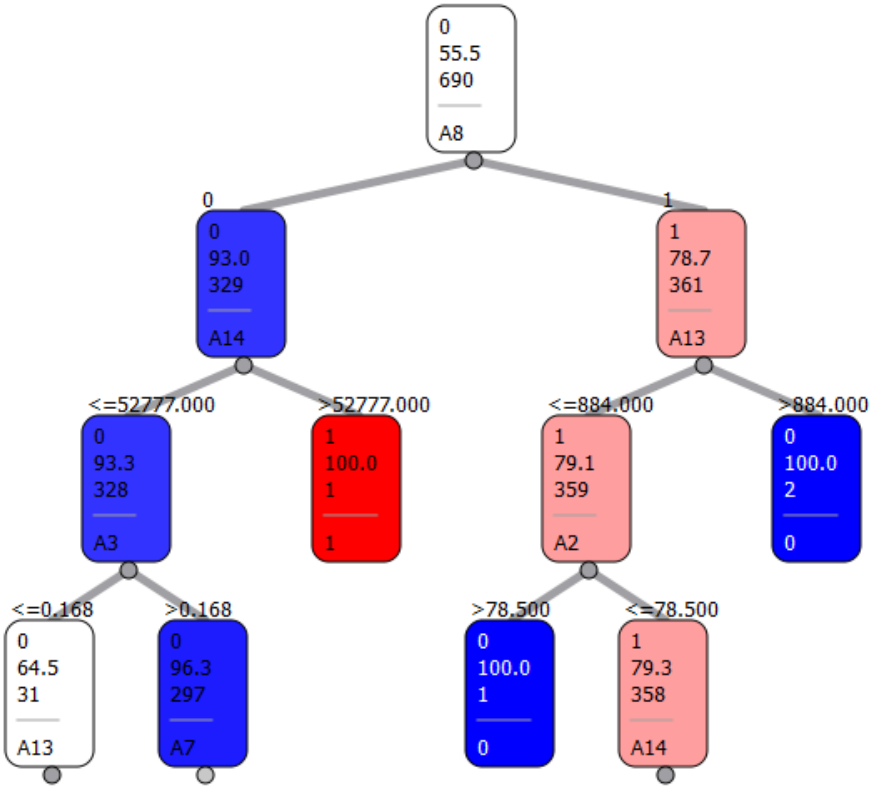

Figure 27 The tree of C4.5 - Statlog (Australian Credit Approval) dataset

- **The comparison of C4.5 and MAL**

Table 9 summarizes the value of the average accuracy of C4.5 and MAL, which the value in bold font represents the better outcome for each dataset. In order to further discuss the result between C4.5 and MAL, we select to focus on the Statlog (Australian Credit Approval) dataset, which the average accuracy value of both posMAL and negMAL of 0.853915 is better than that of C4.5 of 0.801198.

As we can see in at the first level of Figure 25, Figure 26, Figure 27, all of the trees use the attribute A8, which is the discrete attribute. For the latter levels, negMAL in Figure 25 and posMAL in Figure 26 give the result with the higher purity rate than Figure 27 that is split by C4.5.

In Figure 25, the root node contains the attribute A8 which splits instances into 361 and 329 instances. After that, the nodes at the deeper levels are split by

negMAL; for example, node with the attribute A8=1 is split to 236, 123 and 2 instances and node with the attribute A12=2 is split to 267, 17 and 1 instances.

In Figure 26, the root node contains the attribute A8 which also splits the instances into 361 and 329 instances. Then, the latter nodes are split by posMAL; for example, node with the attribute A9=0 is split to 118, 10 and 5 instances and node with the attribute A9=1 is split to 156, 71 and 1 instances.

In Figure 27 of C4.5, the root node contains attribute A8 which splits the instances in the same way as the first two algorithms. However, we can obviously see that C4.5 does not effectively split the instances at the next levels; for example, node with the attribute A8=0 is split to 328 and 1 instances and node with the attribute A8=1 is split to 359 and 2 instances. When we compare C4.5 against negMAL and posMAL, a number of instances split by C4.5 displays that C4.5 is only able to split a few instances at one level of the tree.

CHAPTER V

CONCLUSION

In this thesis, we introduced the new algorithm called multi-attributed lens (MAL). This algorithm uses two types of lens including positive lens (posMAL) and negative lens (negMAL). We performed a test based on the UCI datasets and compared the result against the other algorithms, using t-test as an indicator of the result.

From the experiment, we evaluated the performance of our algorithm compared with the other existing algorithms in terms of the average p value of accuracy of posMAL and negMAL. We found that, for posMAL, our algorithm is better than the other algorithms in 8 datasets which are Blood Transfusion Service Center, Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis, and Horse Colic.

For negMAL, our algorithm is better than the other algorithms in 7 datasets which are Breast Cancer Wisconsin (Original), Breast Cancer Wisconsin (Prognostic), Haberman's Survival, ILPD (Indian Liver Patient Dataset), Credit Approval, Hepatitis, and Horse Colic.

The observation throughout this thesis shows that there are some similarities in the characteristics of the datasets which the result from our algorithms is superior.

Firstly, the better outcome is displayed with the datasets having instances mostly clustered in the center of the group, enabling the lens to effectively partition the instances. In case of posMAL as shown in Figure 28, the positive instances are clustered in the center; therefore, the positive lens can leave out a high number of negative instances. For negMAL, if the negative instances are closely positioned; then, the negative lens can effectively partition these instances.

On the other hand, the datasets with the data that are widely spread show the inferior result. As shown in Figure 29, the positive and negative instances do not stay in a group and they scattered without the specific pattern. Hence, when we apply MAL, the lens can barely partition the dataset and only cut down a few numbers of instances.

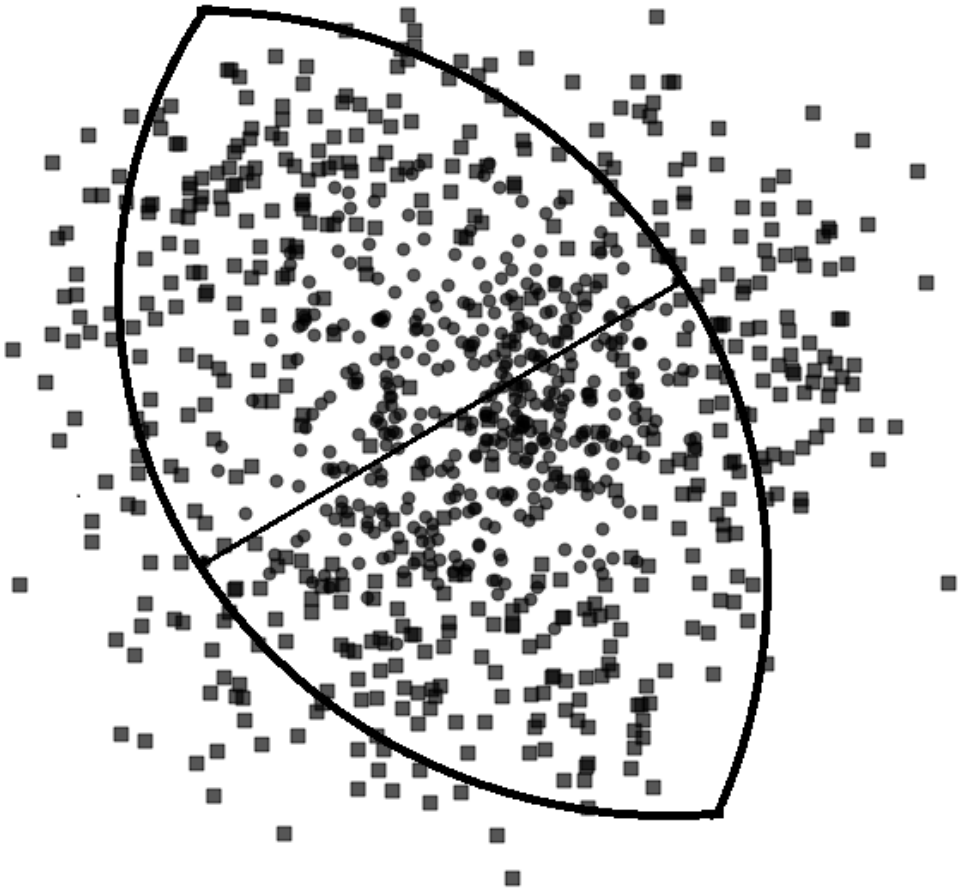

Figure 28 The positive instances stay in the center partitioning by posMAL

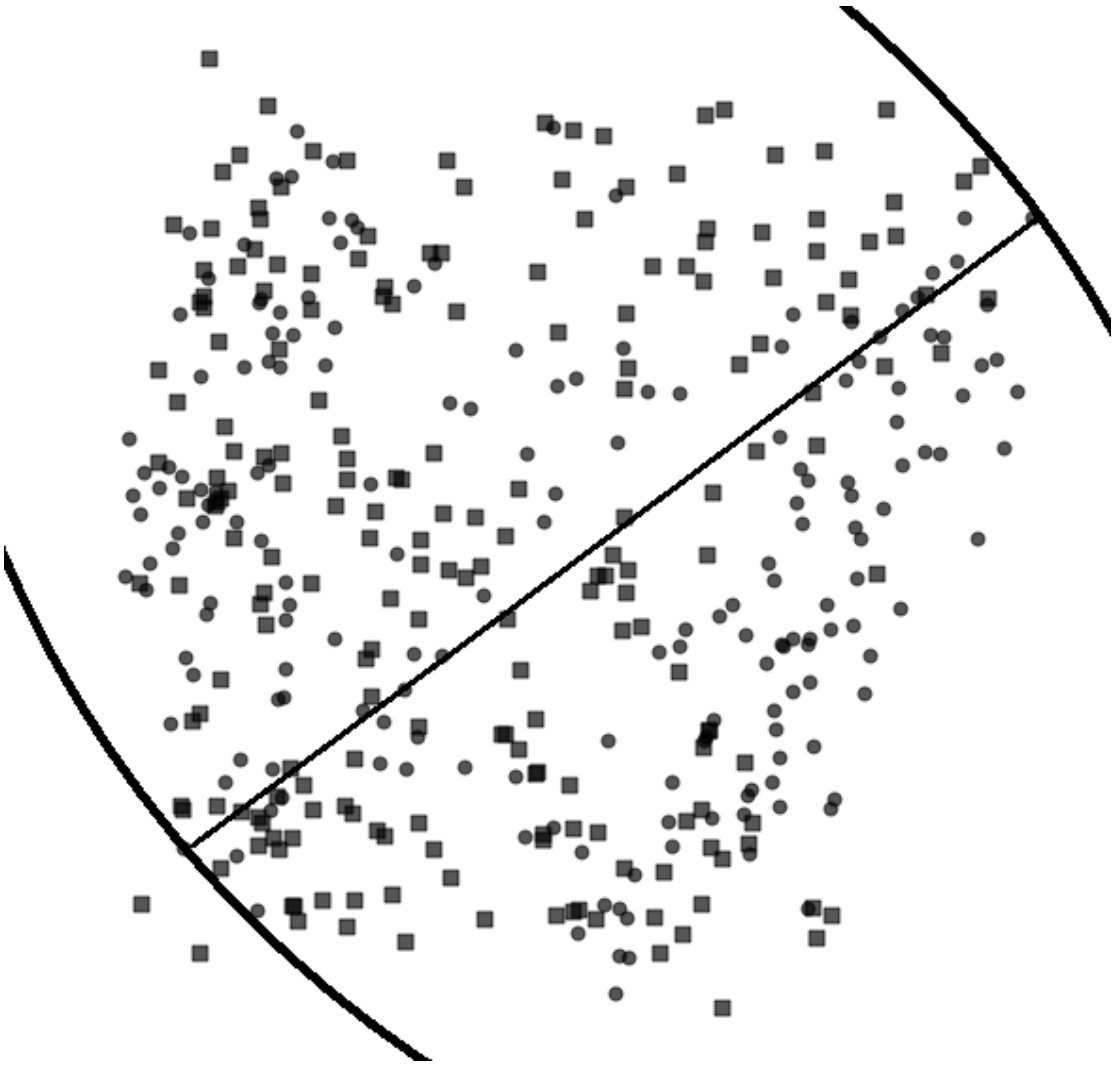

Figure 29 The scattering instances is split by posMAL

In addition, we compared accuracy between C4.5 and our algorithm which both are decision trees. In our experiment, the examples we chose, which MAL has higher accuracy than C4.5, demonstrated that the trees for the same dataset has the better outcome. That was due to the fact that MAL was able to split the nodes with greater purity rate than that using C4.5. If the tree can partition more instances at one level of the tree, it will lead to the more desirable result as the model will not be overfitted.

- **Future work**

Throughout the experiment, we found that our algorithm can partition the dataset at a high gain ratio at the first levels of the tree. However, at the deeper levels, the tree can hardly be partitioned, resulting in a high number of ineffective nodes which lead to the overfitted problem. This algorithm, hence, requires the solution to overcome such problems; for example, the criteria are needed to be set up to prune the tree. Then, the adaptive classifiers, such as k nearest neighbor, shall be used to further deal with the pruned node.

Also, the dataset that has the outliers may lead the tree to be tall; therefore, the outlier should be properly handled in order to deal with this problem.

Lastly, the positive cluster and negative cluster may locate in the dataset at the same time; mix lens is a suggestion to deal with this type of dataset. The algorithm of mix lens will compare between positive lens and negative lens, and select the lens that has higher gain ratio to be used for partitioning.

REFERENCES

[1] Reategui, E.B., and Campbell, J. A classification system for credit card transactions. Advances in Case-Based Reasoning (1995) : 280-291.

[2] Marais, M.L., Patell, J.M., and Wolfson, M.A. The experimental design of classification tests: the case of commercial bank loan classification. Journal of Accounting Research 22 (1984).

[3] Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. Gene selection for cancer classification using support vector machines. Machine learning 46 no.1-3 (2002) : 389-422.

[4] Qi, W., Gu, L., Jiang, H., Chen, X.-R., and Zhang, H.-J. Integrating visual, audio and text analysis for news video. vol. 3 (2000) : 520-523.

[5] Sheng, Y., and Rovnyak, S.M. Decision tree-based methodology for high impedance fault detection. Power Delivery, IEEE Transactions 19 no.2 (2004) : 533-536.

[6] Vlahou, A., Schorge, J.O., Gregory, B.W., and Coleman, R.L. Diagnosis of ovarian cancer using decision tree classification of mass spectral data. BioMed Research International 2003 no.5 (2003) : 308-314.

[7] Quinlan, J.R. Induction of decision trees. Machine learning 1 no.1 (1986) : 81-106.

[8] Quinlan, J.R. Improved use of continuous attributes in C4. 5. arXiv preprint cs/9603103 (1996).

[9] Jolliffe, I. Principal component analysis, Wiley Online Library (2005).

[10] Sun, W., Chen, J., and Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing 21 no.3

(2007) : 1300-1317.

[11] Krung Sinapiromsaran, and Narudom Techaval. Network intrusion detection using multi-attributed frame decision tree. (2012) : 203-207.

[12] Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. When is "nearest neighbor" meaningful?. Springer (1999) : 217-235.

[13] Chen, J., Huang, H., Tian, S., and Qu, Y. Feature selection for text classification with Naïve Bayes. Expert Systems with Applications 36 no.3 (2009) : 5432-5435.

[14] Joachims, T. Making large scale SVM learning practical (1999).

[15] Hosmer, D.W., and Lemeshow, S. Model-Building Strategies and Methods for Logistic Regression. Applied Logistic Regression, Second Edition (2000) : 91-142.

[16] Kononenko, I., and Bratko, I. Information-based evaluation criterion for classifier's performance. Machine Learning 6 (1991) : 67-80.

[17] Brier, G.W. Verification of forecasts expressed in terms of probability. Monthly weather review 78 (1950) : 1-3.

[18] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., and Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16 no.1 (2000) : 412-424.

[19] Breiman, L. Classification and regression trees. CRC press, 1993.

[20] Demšar, J., Zupan, B., Leban, G., and Curk, T. Orange: From experimental machine learning to interactive data mining. Springer (2004).

[21] Bache, K., and Lichman, M. UCI Machine Learning Repository [Online]. 2013. Available from : http://archive.ics.uci.edu/ml.

[22] Yeh, I., Yang, K.-J., Ting, T.-M. and others. Knowledge discovery on RFM model using Bernoulli sequence. Expert Systems with Applications 36 no. 3 (2009) :5866-5871.

[23] Street, W. N., Wolberg, W. H. and Mangasarian, O. L. Nuclear feature extraction for breast tumor diagnosis. International Society for Optics and Photonics (1993) :861-870.

[24] Mangasarian, O. L., Street, W. N. and Wolberg, W. H. Breast cancer diagnosis and prognosis via linear programming. Operations Research 43 no. 4, (1995) :570-577.

[25] Wolberg, W. H., Street, W. N. and Mangasarian, O. L. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. Cancer letters 77 no. 2 (1994) :163-171.

[26] Wolberg, W. H., Street, W. N. and Mangasarian, O. L. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Analytical and Quantitative cytology and histology 17 no. 2 (1995) 77-87.

[27] Wolberg, W. H., Street, W. N., Heisey, D. M. and Mangasarian, O. L. Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. Archives of Surgery 130 no. 5 (1995) :511.

[28] Wolberg, W. H., Street, W. N., Heisey, D. M. and Mangasarian, O. L. Computer-derived nuclear features distinguish malignant from benign breast cytology. Human Pathology 26 no. 7 (1995) :792-796.

[29] Wolberg, W. H. and Mangasarian, O. L. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. ConferenceProceedings of the national academy of sciences 87 no. 23 (1990) :9193-9196.

[30] Street, W. N., Mangasarian, O. L. and Wolberg, W. H. An inductive learning approach to prognostic prediction. (1995) :522-530.

[31] Wolberg, W. H., Street, W. N., Heisey, D. M. and Mangasarian, O. L. Computer-derived nuclear" grade" and breast cancer prognosis. Analytical and Quantitative Cytology and Histology 17 no. 4 (1995) :257-264.

[32] Jossinet, J. Variability of impedivity in normal and pathological breast tissue. Medical and Biological Engineering and Computing 34 no. 5 (1996) :346-350.

[33] da Silva, J. E., de Sá, J. M. and Jossinet, J. Classification of breast tissue by electrical impedance spectroscopy. Medical and Biological Engineering and Computing 38 no. 1 (2000) :26-30.

[34] Horton, P. and Nakai, K. A probabilistic classification system for predicting the cellular localization sites of proteins. 4 (1996) :109-115.

[35] Evett, I. W. and Spiehler, E. J. Rule induction in forensic science. KBS in Goverment, Online Publications (1987) :107-118.

[36] Haberman, S. J. Generalized residuals for log-linear models. (1976) :104-122.

[37] Landwehr, J. M., Pregibon, D. and Shoemaker, A. C. Graphical methods for assessing logistic regression models. Journal of the American Statistical Association 79 no. 385 (1984) :61-71.

[38] Loh, W.-Y. Logistic regression tree analysis. Springer (2006) :537-549.

[39] Ramana, B. V., Babu, M. and Venkateswarlu, N. B. A critical comparative study of liver patients from USA and INDIA: An Exploratory Analysis. International Journal of Computer Science 9 (2012).

[40] Ramana, B. V., Babu, M. and Venkateswarlu, N. B. A critical study of selected classification algorithms for liver disease diagnosis. International Journal of Database Management Systems 3 no. 2 (2011) :101-114.

[41] Sigillito, V. and Baker, K. B. Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest 10 (1989) :262-266.

[42] Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of eugenics 7 no. 2 (1936) :179-188.

[43] Duda, R. O., Hart, P. E. and others. Pattern classification and scene analysis. Wiley New York 3 (1973).

[44] Dasarathy, B. V. Nosing around the neighborhood: A new system structure and classification rule for recognition in partially exposed environments. Pattern Analysis and Machine Intelligence, IEEE Transactions on no. 1 (1980) :67-71.

[45] Gates, G. W. The Reduced Nearest Neighbor Rule. IEEE Transactions on Information Theory (1972) :431-432.

[46] Elter M., Schulz-Wendtland R. and Wittenberg, T. The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Medical Physics 34 (2007) :4164.

[47] Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C. and Johannes, R. S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. American Medical Informatics Association (1988) :261.

[48] Charytanowicz, M., Niewczas, J., Kulczycki, P., Kowalski, P. A., Łukasik, S. and Zak, S. Complete Gradient Clustering Algorithm for Features Analysis of X-Ray Images. Springer (2010) :15-24.

[49] Berthonnaud, E., Dimnet, J., Roussouly, P. and Labelle, H. Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters. Journal of spinal disorders & techniques 18 no. 1 (2005) :40-47.

[50] Neto, A. R. and Barreto, G. A. WCI 04 On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis. Latin America Transactions, IEEE (Revista IEEE America Latina) 7 no. 4 (2009) :487-496.

[51] da Rocha Neto, A. R., Sousa, R., Barreto, G. and Cardoso, J. S. Diagnostic of pathology on the vertebral column with embedded reject option. Springer (2011) :588-595.

[52] Aeberhard, S., Coomans, D. and De Vel, O. Comparison of classifiers in high dimensional settings. *Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland,* Tech. Rep 92-02 (1992).

[53] Aeberhard, S., Coomans, D. and De Vel, O. The classification performance of RDA. *Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland,* Tech. Rep 92-01 (1992).

[54] Quinlan, J. R. Simplifying decision trees. International journal of man-machine studies 27 no. 3 (1987) :221-234.

[55] Quinlan, J. R. C4. 5: programs for machine learning. Morgan kaufmann 1 (1993).

[56] Evans, B. and Fisher, D. Overcoming process delays with decision tree induction. IEEE expert 9 no. 1 (1994) :60-66.

[57] Güvenir, H. A., Demiröz, G. and Ilter, N. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artificial Intelligence in Medicine 13 no. 3 (1998) :147-165.

[58] Diaconis, P. and Efron, B. Computer-intensive methods in statistics. Scientific American 248 no. 5 (1983) :116-130.

[59] Cestnik, B., Kononenko, I. and Bratko, I. A Knowledge-Elicitation Tool for

Sophisticated Users ASSISTANT 86 (1987) :31-45.

[60] Loh, W.-Y. and Shih, Y.-S. Split selection methods for classification trees. Statistica sinica 7 no. 4 (1997) :815-840.

[61] Lim, T.-S., Loh, W.-Y. and Shih, Y.-S. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine learning 40 no. 3 (2000) :203-228.

APPENDICES

APPENDIX A AVERAGE VALUE OF PERFORMANCE MEASURES

Table 10 Average value of accuracy for each dataset and classifier

Accuracy	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.75615	0.71633	0.77110	0.77199	0.76013	0.76287	0.76064
02.Breast Cancer Wisconsin (Diagnostic)	0.94424	0.91729	0.97088	0.95763	0.89853	0.90433	0.94928
03.Breast Cancer Wisconsin (Original)	0.97115	0.94074	0.96675	0.96132	0.95630	0.96488	0.96394
04.Breast Cancer Wisconsin (Prognostic)	0.67479	0.66126	0.77081	0.78904	0.76729	0.76537	0.76904
05.Breast Tissue	0.83193	0.86928	0.89725	0.88059	0.83752	0.83374	0.81738
06.Ecoli	0.56405	0.93784	0.96345	0.95458	0.88742	0.92288	0.91738
07.Glass Identification	0.86410	0.88078	0.89420	0.86275	0.83361	0.83882	0.85104
08.Haberman's Survival	0.75357	0.67497	0.71345	0.74812	0.73637	0.74314	0.73912
09.ILPD (Indian Liver Patient Dataset)	0.66389	0.67519	0.66674	0.71784	0.71210	0.71048	0.70437
10.Ionosphere	0.87646	0.92246	0.85264	0.87471	0.65900	0.66926	0.92112
11.Iris	0.93326	0.95861	0.96603	0.89559	0.83849	0.89647	0.95897
12.Mammographic Mass	0.81665	0.76327	0.80257	0.82615	0.54739	0.54805	0.79597
13.Pima Indians Diabetes	0.74803	0.68918	0.73803	0.77133	0.64986	0.64751	0.76558
14.Seeds	0.92906	0.94494	0.95434	0.97422	0.79963	0.83618	0.94438
15.Vertebral Column	0.76817	0.79189	0.79794	0.83441	0.68333	0.75186	0.84037
16.Wine	0.98140	0.94673	0.97063	0.98109	0.77752	0.77023	0.93800
17.Credit Approval	0.85562	0.79655	0.86454	0.85051	0.85640	0.85640	0.85640
18.Cylinder Bands (Continuous attributes)	0.68827	0.66191	0.67001	0.62828	0.59390	0.60242	0.67780
19.Dermatology	0.97600	0.96275	0.98755	0.96601	0.93847	0.93847	0.87969
20.Hepatitis	0.84943	0.78595	0.82206	0.82101	0.83425	0.83425	0.77544
21.Horse Colic	0.77893	0.76509	0.79225	0.77917	0.81997	0.81997	0.83972
22.Statlog (Australian Credit Approval)	0.86021	0.80120	0.86533	0.85284	0.85392	0.85392	0.85392
23.Statlog (Heart)	0.80874	0.71708	0.79781	0.82837	0.55162	0.55157	0.83354
24.Teaching Assistant Evaluation	0.68642	0.74436	0.71174	0.66631	0.66426	0.66598	0.68329

Table 11 Average value of sensitivity for each dataset and classifier

Average of Sensitivity	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.39449	0.30893	0.30635	0.12949	0.00415	0.03300	0.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.96225	0.92234	0.98943	0.98903	0.97148	0.97817	0.97103
03.Breast Cancer Wisconsin (Original)	0.97786	0.89755	0.94879	0.93557	0.96148	0.97851	0.93667
04.Breast Cancer Wisconsin (Prognostic)	0.48028	0.41917	0.33929	0.50903	0.01681	0.08851	0.00000
05.Breast Tissue	0.82034	0.91836	0.95504	0.96247	0.98496	0.98650	1.00000
06.Ecoli	0.54881	0.96168	0.98036	0.97807	0.90505	0.95905	0.99459
07.Glass Identification	0.87227	0.91941	0.93677	0.94308	0.98078	0.98737	0.96046
08.Haberman's Survival	0.20443	0.36076	0.20911	0.19696	0.00200	0.16476	0.00000
09.ILPD (Indian Liver Patient Dataset)	0.69868	0.43997	0.26164	0.26545	0.00100	0.00033	0.04749
10.Ionosphere	0.89371	0.93336	0.97916	0.93780	0.89286	0.99492	0.97077
11.Iris	0.91093	0.96903	0.97366	0.93477	0.87356	0.93486	0.96877
12.Mammographic Mass	0.86908	0.70883	0.80512	0.82735	0.03064	0.05962	0.81278
13.Pima Indians Diabetes	0.66316	0.56810	0.52552	0.56586	0.00388	0.00521	0.53938
14.Seeds	0.91491	0.96065	0.96776	0.98108	0.94876	0.84302	0.95908
15.Vertebral Column	0.84715	0.67292	0.68493	0.68153	0.02257	0.90965	0.73340
16.Wine	0.98591	0.95533	0.97914	0.98276	0.94041	0.96668	0.98902
17.Credit Approval	0.89038	0.82483	0.88210	0.84531	0.80153	0.80153	0.80153
18.Cylinder Bands (Continuous attributes)	0.78731	0.71009	0.67831	0.74900	0.99011	0.99674	0.81096
19.Dermatology	0.98572	0.97569	0.99453	0.96861	0.98829	0.98829	0.99994
20.Hepatitis	0.89749	0.84531	0.90861	0.90201	0.95203	0.95203	1.00000
21.Horse Colic	0.76169	0.67494	0.67595	0.70934	0.80909	0.80909	0.72093
22.Statlog (Australian Credit Approval)	0.82404	0.77905	0.84999	0.85896	0.92456	0.92456	0.92456
23.Statlog (Heart)	0.85741	0.71481	0.76412	0.78964	0.00775	0.00208	0.79363
24.Teaching Assistant Evaluation	0.84939	0.79497	0.82746	0.85371	0.97814	0.99206	0.92280

Table 12 Average value of specificity for each dataset and classifier

Average of Specificity	Classifier						
Dataset	Bayes	C4.5	KNN	LogReg	NegMAL	PosMAL	SVM
01.Blood Transfusion Service Center	0.86994	0.84455	0.91735	0.97418	0.99803	0.99256	1.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.91453	0.90899	0.94019	0.90581	0.77814	0.78245	0.91338
03.Breast Cancer Wisconsin (Original)	0.96764	0.96337	0.97614	0.97482	0.95361	0.95775	0.97822
04.Breast Cancer Wisconsin (Prognostic)	0.73392	0.73455	0.90141	0.87301	0.99297	0.96886	1.00000
05.Breast Tissue	0.89049	0.63234	0.62347	0.48274	0.15327	0.12437	0.00000
06.Ecoli	0.97113	0.53732	0.61927	0.59072	0.22560	0.37115	0.11912
07.Glass Identification	0.69957	0.60244	0.54854	0.40578	0.09357	0.11898	0.14107
08.Haberman's Survival	0.94735	0.78591	0.89139	0.94259	0.99557	0.94734	1.00000
09.ILPD (Indian Liver Patient Dataset)	0.64981	0.77032	0.83063	0.90084	0.99973	0.99773	0.97013
10.Ionosphere	0.84565	0.90291	0.62611	0.76170	0.24045	0.08644	0.83222
11.Iris	0.97792	0.93778	0.95079	0.81722	0.76833	0.81968	0.93935
12.Mammographic Mass	0.77133	0.81033	0.80036	0.82512	0.99412	0.97031	0.78145
13.Pima Indians Diabetes	0.79375	0.75442	0.85251	0.88204	0.99789	0.99354	0.88744
14.Seeds	0.95734	0.91353	0.92751	0.96048	0.50136	0.82249	0.91498
15.Vertebral Column	0.73056	0.84854	0.85175	0.90721	0.99798	0.67672	0.89130
16.Wine	0.97727	0.92896	0.96366	0.98021	0.43118	0.37294	0.80868
17.Credit Approval	0.81267	0.76165	0.84284	0.85699	0.92426	0.92426	0.92426
18.Cylinder Bands (Continuous attributes)	0.55419	0.59673	0.65883	0.46483	0.05731	0.06844	0.49751
19.Dermatology	0.91881	0.89365	0.94973	0.95037	0.66933	0.66933	0.16538
20.Hepatitis	0.68558	0.58178	0.52407	0.54342	0.42863	0.42863	0.00000
21.Horse Colic	0.78885	0.81639	0.85840	0.81896	0.82603	0.82603	0.90730
22.Statlog (Australian Credit Approval)	0.88885	0.81875	0.87752	0.84802	0.79802	0.79802	0.79802
23.Statlog (Heart)	0.76981	0.71889	0.82477	0.85935	0.98671	0.99116	0.86546
24.Teaching Assistant Evaluation	0.36545	0.64367	0.47994	0.29926	0.04565	0.02580	0.20608

Table 13 Average value of area under ROC curve for each dataset and classifier

Average of AUC	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.72953	0.58507	0.70755	0.75172	0.49694	0.51125	0.50000
02.Breast Cancer Wisconsin (Diagnostic)	0.98136	0.91567	0.99201	0.99079	0.88056	0.91300	0.98936
03.Breast Cancer Wisconsin (Original)	0.99303	0.93046	0.99135	0.99371	0.96436	0.97626	0.99367
04.Breast Cancer Wisconsin (Prognostic)	0.66177	0.57686	0.64271	0.80852	0.50476	0.53567	0.50000
05.Breast Tissue	0.91120	0.77535	0.91707	0.86173	0.71343	0.74620	0.50000
06.Ecoli	0.88754	0.74950	0.84929	0.82539	0.68250	0.72206	0.62364
07.Glass Identification	0.89046	0.76092	0.88759	0.85449	0.63127	0.63977	0.59987
08.Haberman's Survival	0.66516	0.56729	0.62399	0.64206	0.49009	0.55598	0.50000
09.ILPD (Indian Liver Patient Dataset)	0.74255	0.63986	0.64977	0.74691	0.53749	0.61698	0.66239
10.Ionosphere	0.93012	0.91814	0.92822	0.85921	0.61348	0.53718	0.95538
11.Iris	0.98481	0.95340	0.99568	0.92845	0.89171	0.93430	0.98692
12.Mammographic Mass	0.89915	0.77883	0.86438	0.89577	0.53987	0.55316	0.86852
13.Pima Indians Diabetes	0.82592	0.66228	0.79435	0.82542	0.49985	0.53652	0.82899
14.Seeds	0.98404	0.93709	0.99328	0.99610	0.81745	0.88201	0.98706
15.Vertebral Column	0.86579	0.76073	0.87196	0.89919	0.61882	0.80806	0.91060
16.Wine	0.99879	0.94215	0.99861	0.99750	0.79131	0.82555	0.91499
17.Credit Approval	0.91184	0.79471	0.91451	0.91430	0.86290	0.86290	0.92320
18.Cylinder Bands (Continuous attributes)	0.75259	0.66142	0.72104	0.65996	0.52165	0.53009	0.72881
19.Dermatology	0.95983	0.94293	0.99875	0.98328	0.89660	0.89705	0.58333
20.Hepatitis	0.89075	0.71354	0.80969	0.81669	0.69319	0.69319	0.50000
21.Horse Colic	0.82011	0.74875	0.82927	0.81355	0.81756	0.81756	0.86938
22.Statlog (Australian Credit Approval)	0.91863	0.80082	0.91507	0.92181	0.86129	0.86129	0.92203
23.Statlog (Heart)	0.89640	0.71685	0.87311	0.89982	0.50255	0.49467	0.90196
24.Teaching Assistant Evaluation	0.68968	0.72682	0.74019	0.67314	0.54575	0.53250	0.68852

Table 14 Average value of information score for each dataset and classifier

Average of Information Score	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.07383	0.05247	0.08421	0.07652	0.00407	0.01515	- 0.00034
02.Breast Cancer Wisconsin (Diagnostic)	0.84013	0.78346	0.84362	0.75379	1.99619	2.13226	0.80889
03.Breast Cancer Wisconsin (Original)	0.86534	0.80106	0.82777	0.82769	4.23712	2.37325	0.81936
04.Breast Cancer Wisconsin (Prognostic)	- 0.04392	- 0.06668	- 0.02894	0.08495	0.12401	0.10245	- 0.04793
05.Breast Tissue	0.18724	0.31027	0.33246	0.19356	1.72157	2.78347	- 0.01788
06.Ecoli	- 2.05750	0.25568	0.27411	0.26536	1.89234	4.88058	0.06604
07.Glass Identification	0.10679	0.25491	0.21565	0.07826	0.60117	1.72636	0.03973
08.Haberman's Survival	0.06882	0.02231	0.02606	0.04497	0.01595	0.04359	- 0.01841
09.ILPD (Indian Liver Patient Dataset)	0.08987	0.13098	0.00789	0.10371	0.30287	0.15043	0.01896
10.Ionosphere	0.67401	0.77698	0.59860	0.65567	0.39862	0.15496	0.67901
11.Iris	0.75977	0.82849	0.83381	0.65345	3.66952	4.85523	0.74704
12.Mammographic Mass	0.60513	0.51035	0.46901	0.49171	0.22755	0.31220	0.44708
13.Pima Indians Diabetes	0.33871	0.27010	0.26504	0.30477	0.01802	0.04021	0.29237
14.Seeds	0.75525	0.79882	0.79763	0.68133	1.86879	3.79594	0.73150
15.Vertebral Column	0.37875	0.45054	0.36929	0.24093	0.95642	0.66875	0.42505
16.Wine	0.86634	0.79843	0.83070	0.86736	1.75427	3.16456	0.68595
17.Credit Approval	0.66112	0.58148	0.59636	0.61562	0.55396	0.55396	0.57481
18.Cylinder Bands (Continuous attributes)	0.27044	0.29131	0.16559	0.12088	0.03713	0.15187	0.15944
19.Dermatology	0.55825	0.52208	0.55461	0.52458	3.80435	3.80799	0.12439
20.Hepatitis	0.38354	0.22739	0.22122	0.28102	0.21357	0.21357	0.01538
21.Horse Colic	0.46580	0.44635	0.36889	0.41250	0.39007	0.39007	0.43858
22.Statlog (Australian Credit Approval)	0.66845	0.58825	0.59962	0.61321	0.54920	0.54920	0.56813
23.Statlog (Heart)	0.54980	0.42016	0.50258	0.55752	0.04407	- 0.00385	0.52153
24.Teaching Assistant Evaluation	0.11705	0.36999	0.21391	0.08823	0.19900	0.17543	0.07746

Table 15 Average value of F-measure for each dataset and classifier

Average of F-measure	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.43496	0.34087	0.38842	0.21059	0.00789	0.06066	0.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.95553	0.93257	0.97699	0.96686	0.92274	0.92739	0.95981
03.Breast Cancer Wisconsin (Original)	0.95899	0.91206	0.95139	0.94314	0.93792	0.95056	0.94674
04.Breast Cancer Wisconsin (Prognostic)	0.40097	0.35892	0.39728	0.52048	0.02753	0.13961	0.00000
05.Breast Tissue	0.88298	0.91936	0.93889	0.93053	0.90853	0.90655	0.89898
06.Ecoli	0.59547	0.96011	0.97735	0.97204	0.90554	0.94950	0.95314
07.Glass Identification	0.90055	0.91838	0.92780	0.91066	0.90122	0.90462	0.90607
08.Haberman's Survival	0.29611	0.36123	0.26864	0.28295	0.00328	0.22319	0.00000
09.ILPD (Indian Liver Patient Dataset)	0.54476	0.42951	0.30981	0.34671	0.00194	0.00056	0.06876
10.Ionosphere	0.90251	0.93899	0.89529	0.90588	0.76058	0.79434	0.94057
11.Iris	0.94587	0.96888	0.97447	0.92458	0.86990	0.92334	0.96913
12.Mammographic Mass	0.81480	0.73439	0.79083	0.81521	0.05592	0.07949	0.78691
13.Pima Indians Diabetes	0.64793	0.56058	0.58223	0.63282	0.00741	0.00972	0.61568
14.Seeds	0.94421	0.95876	0.96587	0.98066	0.86778	0.86709	0.95824
15.Vertebral Column	0.70265	0.67307	0.68478	0.72199	0.04185	0.67894	0.74597
16.Wine	0.98548	0.95888	0.97727	0.98513	0.84930	0.84887	0.96077
17.Credit Approval	0.87221	0.81730	0.87801	0.86216	0.86022	0.86022	0.86022
18.Cylinder Bands (Continuous attributes)	0.74337	0.70597	0.70171	0.69756	0.73723	0.74258	0.74250
19.Dermatology	0.98562	0.97741	0.99248	0.97907	0.96509	0.96509	0.93444
20.Hepatitis	0.90183	0.85836	0.88717	0.88587	0.89924	0.89924	0.87343
21.Horse Colic	0.71448	0.67440	0.70071	0.69913	0.76551	0.76551	0.76411
22.Statlog (Australian Credit Approval)	0.83886	0.77550	0.84778	0.83762	0.84844	0.84844	0.84844
23.Statlog (Heart)	0.79959	0.69135	0.76996	0.80311	0.01435	0.00368	0.80863
24.Teaching Assistant Evaluation	0.78077	0.80265	0.79023	0.77120	0.79375	0.79717	0.79353

Table 16 Average value of precision for each dataset and classifier

Average of Precision	Classifier						
Dataset	Bayes	C4.5	KNN	LogReg	NegMAL	PosMAL	SVM
01.Blood Transfusion Service Center	0.49562	0.38678	0.54935	0.61785	0.10917	0.50256	0.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.94937	0.94412	0.96512	0.94601	0.87978	0.88216	0.94934
03.Breast Cancer Wisconsin (Original)	0.94151	0.92881	0.95490	0.95209	0.91661	0.92483	0.95821
04.Breast Cancer Wisconsin (Prognostic)	0.35435	0.32526	0.51599	0.56928	0.10139	0.43968	0.00000
05.Breast Tissue	0.97285	0.92465	0.92585	0.90395	0.84561	0.84054	0.81738
06.Ecoli	0.68083	0.95947	0.97486	0.96659	0.94063	0.94562	0.91860
07.Glass Identification	0.93779	0.91954	0.92163	0.88395	0.84272	0.84293	0.86062
08.Haberman's Survival	0.61026	0.37702	0.39891	0.58001	0.01000	0.46723	0.00000
09.ILPD (Indian Liver Patient Dataset)	0.44840	0.43720	0.38972	0.52554	0.04167	0.00185	0.22606
10.Ionosphere	0.91365	0.94656	0.82546	0.87767	0.69245	0.66126	0.91305
11.Iris	0.98835	0.97053	0.97641	0.91724	0.91749	0.92292	0.97102
12.Mammographic Mass	0.76889	0.76433	0.77914	0.80523	0.76273	0.74026	0.76441
13.Pima Indians Diabetes	0.63711	0.55721	0.65964	0.72532	0.11953	0.09099	0.72522
14.Seeds	0.97783	0.95852	0.96555	0.98096	0.81386	0.91779	0.95905
15.Vertebral Column	0.60337	0.68271	0.69336	0.78425	0.37847	0.54345	0.76723
16.Wine	0.98589	0.96453	0.97738	0.98826	0.78983	0.76619	0.94033
17.Credit Approval	0.85609	0.81163	0.87520	0.88134	0.92963	0.92963	0.92963
18.Cylinder Bands (Continuous attributes)	0.70615	0.70666	0.73012	0.65469	0.58732	0.59175	0.68725
19.Dermatology	0.98589	0.97960	0.99061	0.99055	0.94526	0.94526	0.87956
20.Hepatitis	0.91020	0.87654	0.86976	0.87504	0.85349	0.85349	0.77544
21.Horse Colic	0.67776	0.68105	0.73517	0.69563	0.72951	0.72951	0.82143
22.Statlog (Australian Credit Approval)	0.85612	0.77476	0.84765	0.81874	0.78459	0.78459	0.78459
23.Statlog (Heart)	0.75303	0.67450	0.78263	0.82345	0.12290	0.01722	0.82970
24.Teaching Assistant Evaluation	0.72634	0.81861	0.76213	0.70862	0.66855	0.66663	0.70074

Table 17 Average value of recall for each dataset and classifier

Average of Recall	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.39449	0.30893	0.30635	0.12949	0.00415	0.03300	0.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.96225	0.92234	0.98943	0.98903	0.97148	0.97817	0.97103
03.Breast Cancer Wisconsin (Original)	0.97786	0.89755	0.94879	0.93557	0.96148	0.97851	0.93667
04.Breast Cancer Wisconsin (Prognostic)	0.48028	0.41917	0.33929	0.50903	0.01681	0.08851	0.00000
05.Breast Tissue	0.82034	0.91836	0.95504	0.96247	0.98496	0.98650	1.00000
06.Ecoli	0.54881	0.96168	0.98036	0.97807	0.90505	0.95905	0.99459
07.Glass Identification	0.87227	0.91941	0.93677	0.94308	0.98078	0.98737	0.96046
08.Haberman's Survival	0.20443	0.36076	0.20911	0.19696	0.00200	0.16476	0.00000
09.ILPD (Indian Liver Patient Dataset)	0.69868	0.43997	0.26164	0.26545	0.00100	0.00033	0.04749
10.Ionosphere	0.89371	0.93336	0.97916	0.93780	0.89286	0.99492	0.97077
11.Iris	0.91093	0.96903	0.97366	0.93477	0.87356	0.93486	0.96877
12.Mammographic Mass	0.86908	0.70883	0.80512	0.82735	0.03064	0.05962	0.81278
13.Pima Indians Diabetes	0.66316	0.56810	0.52552	0.56586	0.00388	0.00521	0.53938
14.Seeds	0.91491	0.96065	0.96776	0.98108	0.94876	0.84302	0.95908
15.Vertebral Column	0.84715	0.67292	0.68493	0.68153	0.02257	0.90965	0.73340
16.Wine	0.98591	0.95533	0.97914	0.98276	0.94041	0.96668	0.98902
17.Credit Approval	0.89038	0.82483	0.88210	0.84531	0.80153	0.80153	0.80153
18.Cylinder Bands (Continuous attributes)	0.78731	0.71009	0.67831	0.74900	0.99011	0.99674	0.81096
19.Dermatology	0.98572	0.97569	0.99453	0.96861	0.98829	0.98829	0.99994
20.Hepatitis	0.89749	0.84531	0.90861	0.90201	0.95203	0.95203	1.00000
21.Horse Colic	0.76169	0.67494	0.67595	0.70934	0.80909	0.80909	0.72093
22.Statlog (Australian Credit Approval)	0.82404	0.77905	0.84999	0.85896	0.92456	0.92456	0.92456
23.Statlog (Heart)	0.85741	0.71481	0.76412	0.78964	0.00775	0.00208	0.79363
24.Teaching Assistant Evaluation	0.84939	0.79497	0.82746	0.85371	0.97814	0.99206	0.92280

Table 18 Average value of Brier score for each dataset and classifier

Average of Brier score	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.32908	0.54321	0.33341	0.31016	0.39457	0.40101	0.36414
02.Breast Cancer Wisconsin (Diagnostic)	0.10289	0.16542	0.05388	0.07633	1,716.27	1,989.71	0.07024
03.Breast Cancer Wisconsin (Original)	0.05071	0.11851	0.05142	0.05724	30,149.6	7,175.90	0.05393
04.Breast Cancer Wisconsin (Prognostic)	0.53359	0.67749	0.36491	0.30079	6.07295	4.38777	0.35564
05.Breast Tissue	0.29008	0.26144	0.15305	0.17780	439.0865	854.1442	0.29703
06.Ecoli	0.79864	0.12433	0.06170	0.06904	4,592.68	30,081.2	0.13930
07.Glass Identification	0.19824	0.23844	0.15263	0.19071	674.9035	2,611.27	0.22171
08.Haberman's Survival	0.36185	0.64837	0.40281	0.37366	1.83564	1.35370	0.38574
09.ILPD (Indian Liver Patient Dataset)	0.46991	0.58852	0.40628	0.34981	60.77842	45.97653	0.38913
10.Ionosphere	0.21836	0.15508	0.22752	0.21777	26.92106	3.52444	0.12732
11.Iris	0.09835	0.08278	0.04937	0.13682	1,479.09	2,596.77	0.07095
12.Mammographic Mass	0.28268	0.43937	0.29489	0.25178	22.93055	52.71337	0.28946
13.Pima Indians Diabetes	0.34659	0.61913	0.34956	0.31976	2.14627	6.18186	0.31884
14.Seeds	0.11390	0.11012	0.06057	0.07657	682.7313	3,247.35	0.08333
15.Vertebral Column	0.34699	0.41622	0.27523	0.28210	300.9584	412.4655	0.22629
16.Wine	0.03029	0.10653	0.04304	0.03081	406.6937	1,590.49	0.09785
17.Credit Approval	0.23300	0.40522	0.21786	0.22057	0.23754	0.23754	0.22550
18.Cylinder Bands (Continuous attributes)	0.41272	0.64642	0.43052	0.45514	1.31169	5.03086	0.41698
19.Dermatology	0.04366	0.07050	0.02103	0.05869	21,359.3	21,360.5	0.20206
20.Hepatitis	0.26333	0.42810	0.28119	0.29469	3.00743	3.00743	0.34894
21.Horse Colic	0.36376	0.46610	0.30833	0.34220	0.28576	0.28576	0.25041
22.Statlog (Australian Credit Approval)	0.22325	0.39546	0.21343	0.20972	0.23949	0.23949	0.22799
23.Statlog (Heart)	0.27476	0.56584	0.29861	0.25522	1.12466	0.60432	0.24726
24.Teaching Assistant Evaluation	0.41651	0.49291	0.39436	0.42421	5.37952	4.70358	0.41074

Table 19 Average value of MCC for each dataset and classifier

Average of MCC	Classifier						
Dataset	**Bayes**	**C4.5**	**KNN**	**LogReg**	**NegMAL**	**PosMAL**	**SVM**
01.Blood Transfusion Service Center	0.28822	0.16684	0.28132	0.20105	0.00954	0.08670	0.00000
02.Breast Cancer Wisconsin (Diagnostic)	0.88165	0.82729	0.93840	0.91060	0.78651	0.79858	0.89237
03.Breast Cancer Wisconsin (Original)	0.93758	0.86861	0.92663	0.91460	0.90556	0.92469	0.92025
04.Breast Cancer Wisconsin (Prognostic)	0.19584	0.14268	0.28087	0.39997	0.01897	0.12069	0.00000
05.Breast Tissue	0.62090	0.55441	0.60555	0.47721	0.15820	0.13516	0.00000
06.Ecoli	0.41217	0.48262	0.60403	0.57394	0.13053	0.36090	0.11376
07.Glass Identification	0.52696	0.52793	0.51426	0.37868	0.08766	0.14453	0.11488
08.Haberman's Survival	0.23573	0.14967	0.12817	0.21611	- 0.01033	0.15803	0.00000
09.ILPD (Indian Liver Patient Dataset)	0.31864	0.21046	0.10716	0.21362	0.00377	- 0.00898	0.02828
10.Ionosphere	0.73698	0.83536	0.68296	0.72601	0.22616	0.19632	0.82896
11.Iris	0.86990	0.91013	0.92578	0.75841	0.63958	0.78063	0.91057
12.Mammographic Mass	0.64142	0.52392	0.60591	0.65278	0.08810	0.09064	0.59405
13.Pima Indians Diabetes	0.45415	0.32224	0.40276	0.48049	0.00745	- 0.00708	0.46490
14.Seeds	0.85328	0.87891	0.89984	0.94339	0.46756	0.66257	0.87791
15.Vertebral Column	0.54520	0.52517	0.54063	0.61449	0.06793	0.55286	0.63396
16.Wine	0.96057	0.88416	0.93883	0.95967	0.39900	0.38820	0.79818
17.Credit Approval	0.70870	0.58955	0.72702	0.70109	0.72339	0.72339	0.72339
18.Cylinder Bands (Continuous attributes)	0.35436	0.30995	0.33597	0.22515	0.12164	0.18006	0.33046
19.Dermatology	0.90909	0.86443	0.95199	0.88612	0.66244	0.66244	0.16564
20.Hepatitis	0.58324	0.41963	0.47272	0.47484	0.46131	0.46131	0.00000
21.Horse Colic	0.54122	0.49428	0.54649	0.52841	0.62446	0.62446	0.65014
22.Statlog (Australian Credit Approval)	0.71708	0.59896	0.72847	0.70503	0.71900	0.71900	0.71900
23.Statlog (Heart)	0.62691	0.43477	0.59401	0.65556	- 0.01428	- 0.01812	0.66535
24.Teaching Assistant Evaluation	0.25007	0.44251	0.33159	0.17637	0.04599	0.04476	0.16256

APPENDIX B P VALUE OF PERFORMANCE MEASURES (posMAL)

Table 20 p value of accuracy for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.005	0.497		Bayes	0.047	0.000	+	Bayes	0.001	0.589	
C4.5	0.045	0.000	+	C4.5	0.036	0.037		C4.5	0.060	0.000	+
KNN	-0.010	0.035		KNN	0.044	0.000	+	KNN	-0.008	0.267	
LogReg	-0.011	0.018	-	LogReg	-0.007	0.385		LogReg	0.006	0.445	
SVM	0.000	0.235		SVM	0.007	0.140		SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.044	0.000	-	Bayes	-0.210	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.017	0.130		C4.5	-0.256	0.000	-	C4.5	-0.062	0.000	-
KNN	-0.070	0.000	-	KNN	-0.187	0.000	-	KNN	-0.070	0.000	-
LogReg	-0.057	0.000	-	LogReg	-0.209	0.000	-	LogReg	-0.029	0.003	-
SVM	-0.049	0.000	-	SVM	-0.255	0.000	-	SVM	-0.078	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.009	0.092		Bayes	-0.050	0.129		Bayes	-0.038	0.095	
C4.5	0.021	0.009	+	C4.5	-0.075	0.005	-	C4.5	-0.024	0.222	
KNN	-0.005	0.389		KNN	-0.083	0.003	-	KNN	-0.049	0.134	
LogReg	0.000	0.432		LogReg	-0.012	0.003	-	LogReg	-0.028	0.152	
SVM	-0.002	0.498		SVM	-0.076	0.008	-	SVM	0.059	0.055	
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.091	0.000	+	Bayes	-0.269	0.000	-	Bayes	-0.015	0.310	
C4.5	0.105	0.000	+	C4.5	-0.215	0.000	-	C4.5	0.048	0.013	+
KNN	-0.005	0.458		KNN	-0.255	0.000	-	KNN	0.012	0.361	
LogReg	-0.023	0.102		LogReg	-0.278	0.000	-	LogReg	0.013	0.447	
SVM	-0.003	0.412		SVM	-0.248	0.000	-	SVM	0.059	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.004	0.097		Bayes	-0.099	0.000	-	Bayes	0.041	0.001	+
C4.5	-0.033	0.102		C4.5	-0.040	0.000	-	C4.5	0.055	0.000	+
KNN	-0.061	0.016	-	KNN	-0.089	0.000	-	KNN	0.028	0.017	+
LogReg	-0.045	0.103		LogReg	-0.122	0.000	-	LogReg	0.041	0.011	+
SVM	0.018	0.305		SVM	-0.117	0.000	-	SVM	-0.020	0.096	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.347	0.050		Bayes	-0.090	0.005	-	Bayes	-0.006	0.371	
C4.5	-0.027	0.272		C4.5	-0.106	0.012	-	C4.5	0.053	0.000	+
KNN	-0.053	0.304		KNN	-0.115	0.000	-	KNN	-0.011	0.044	
LogReg	-0.044	0.172		LogReg	-0.135	0.000	-	LogReg	0.001	0.305	
SVM	-0.007	0.390		SVM	-0.105	0.009	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.027	0.111		Bayes	-0.039	0.212		Bayes	-0.258	0.000	-
C4.5	-0.043	0.034		C4.5	-0.063	0.041		C4.5	-0.166	0.000	-
KNN	-0.057	0.084		KNN	-0.069	0.011	-	KNN	-0.247	0.000	-
LogReg	-0.025	0.092		LogReg	-0.105	0.000	-	LogReg	-0.277	0.000	-
SVM	-0.014	0.215		SVM	-0.111	0.000	-	SVM	-0.282	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.015	0.120		Bayes	-0.208	0.000	-	Bayes	-0.021	0.356	
C4.5	0.064	0.000	+	C4.5	-0.173	0.022	-	C4.5	-0.079	0.014	-
KNN	0.026	0.004	+	KNN	-0.197	0.000	-	KNN	-0.046	0.202	
LogReg	-0.009	0.231		LogReg	-0.207	0.000	-	LogReg	-0.001	0.073	
SVM	0.000	0.333		SVM	-0.164	0.038		SVM	-0.018	0.197	

Table 21 p value of sensitivity for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.381	0.000	-	Bayes	-0.698	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.295	0.000	-	C4.5	-0.439	0.000	-	C4.5	-0.023	0.067	
KNN	-0.293	0.000	-	KNN	-0.261	0.000	-	KNN	-0.081	0.000	-
LogReg	-0.116	0.000	-	LogReg	-0.265	0.000	-	LogReg	-0.044	0.000	-
SVM	0.014	0.069		SVM	-0.047	0.003	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	0.011	0.214		Bayes	0.067	0.240		Bayes	0.207	0.000	+
C4.5	0.051	0.000	+	C4.5	0.027	0.093		C4.5	0.284	0.000	+
KNN	-0.016	0.013	-	KNN	-0.018	0.010	-	KNN	0.316	0.000	+
LogReg	-0.015	0.010	-	LogReg	0.023	0.085		LogReg	0.246	0.000	+
SVM	0.003	0.382		SVM	-0.010	0.013	-	SVM	0.184	0.000	+
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.006	0.545		Bayes	-0.004	0.338		Bayes	0.003	0.080	
C4.5	0.074	0.000	+	C4.5	-0.062	0.102		C4.5	0.013	0.488	
KNN	0.023	0.098		KNN	-0.066	0.107		KNN	-0.006	0.055	
LogReg	0.036	0.015	+	LogReg	-0.027	0.240		LogReg	0.020	0.173	
SVM	0.035	0.052		SVM	-0.061	0.112		SVM	-0.012	0.016	-
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.420	0.000	-	Bayes	-0.819	0.000	-	Bayes	0.055	0.001	+
C4.5	-0.359	0.000	-	C4.5	-0.659	0.000	-	C4.5	0.107	0.000	+
KNN	-0.279	0.000	-	KNN	-0.755	0.000	-	KNN	0.043	0.004	+
LogReg	-0.449	0.000	-	LogReg	-0.777	0.000	-	LogReg	0.050	0.024	+
SVM	0.060	0.019	+	SVM	-0.763	0.000	-	SVM	-0.048	0.000	-
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.164	0.022	+	Bayes	-0.659	0.000	-	Bayes	0.047	0.069	
C4.5	0.066	0.103		C4.5	-0.563	0.000	-	C4.5	0.134	0.000	+
KNN	0.029	0.208		KNN	-0.521	0.000	-	KNN	0.133	0.000	+
LogReg	0.021	0.177		LogReg	-0.561	0.000	-	LogReg	0.100	0.000	+
SVM	-0.016	0.114		SVM	-0.535	0.000	-	SVM	0.088	0.000	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.390	0.034		Bayes	-0.016	0.217		Bayes	0.101	0.000	+
C4.5	-0.022	0.075		C4.5	-0.062	0.110		C4.5	0.146	0.000	+
KNN	-0.041	0.116		KNN	-0.069	0.096		KNN	0.075	0.000	+
LogReg	-0.039	0.080		LogReg	-0.082	0.028		LogReg	0.066	0.000	+
SVM	-0.055	0.168		SVM	-0.060	0.090		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.109	0.066		Bayes	-0.233	0.183		Bayes	-0.851	0.000	-
C4.5	0.062	0.097		C4.5	-0.059	0.004	-	C4.5	-0.709	0.000	-
KNN	0.045	0.113		KNN	-0.071	0.007	-	KNN	-0.758	0.000	-
LogReg	0.038	0.133		LogReg	-0.068	0.007	-	LogReg	-0.783	0.000	-
SVM	0.021	0.072		SVM	-0.119	0.026		SVM	-0.787	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.138	0.060		Bayes	-0.040	0.170		Bayes	0.137	0.000	+
C4.5	-0.294	0.000	-	C4.5	-0.009	0.260		C4.5	0.192	0.000	+
KNN	-0.143	0.064		KNN	-0.033	0.117		KNN	0.159	0.000	+
LogReg	-0.130	0.055		LogReg	-0.036	0.196		LogReg	0.133	0.000	+
SVM	0.067	0.117		SVM	-0.043	0.131		SVM	0.064	0.007	+

Table 22 p value of specificity for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.126	0.000	+	Bayes	0.349	0.000	+	Bayes	0.112	0.000	+
C4.5	0.152	0.000	+	C4.5	0.228	0.000	+	C4.5	0.163	0.000	+
KNN	0.079	0.000	+	KNN	0.168	0.000	+	KNN	0.081	0.000	+
LogReg	0.022	0.000	+	LogReg	0.098	0.000	+	LogReg	0.067	0.000	+
SVM	-0.004	0.017	-	SVM	0.028	0.005	+	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.135	0.000	-	Bayes	-0.707	0.000	-	Bayes	-0.489	0.000	-
C4.5	-0.129	0.000	-	C4.5	-0.764	0.000	-	C4.5	-0.532	0.000	-
KNN	-0.161	0.000	-	KNN	-0.488	0.000	-	KNN	-0.594	0.000	-
LogReg	-0.126	0.000	-	LogReg	-0.623	0.000	-	LogReg	-0.400	0.000	-
SVM	-0.134	0.000	-	SVM	-0.694	0.000	-	SVM	-0.433	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.011	0.054		Bayes	-0.142	0.093		Bayes	-0.249	0.064	
C4.5	-0.007	0.176		C4.5	-0.102	0.145		C4.5	-0.224	0.180	
KNN	-0.020	0.002	-	KNN	-0.115	0.163		KNN	-0.280	0.095	
LogReg	-0.019	0.005	-	LogReg	0.018	0.061		LogReg	-0.281	0.177	
SVM	-0.022	0.001	-	SVM	-0.104	0.186		SVM	0.504	0.006	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.245	0.000	+	Bayes	0.207	0.000	+	Bayes	-0.257	0.000	-
C4.5	0.244	0.000	+	C4.5	0.168	0.001	+	C4.5	-0.153	0.004	-
KNN	0.077	0.000	+	KNN	0.178	0.001	+	KNN	-0.095	0.039	
LogReg	0.106	0.000	+	LogReg	0.153	0.002	+	LogReg	-0.115	0.098	
SVM	-0.021	0.025	-	SVM	0.197	0.000	+	SVM	0.429	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.742	0.000	-	Bayes	0.202	0.000	+	Bayes	0.037	0.086	
C4.5	-0.484	0.000	-	C4.5	0.242	0.000	+	C4.5	0.010	0.384	
KNN	-0.475	0.000	-	KNN	0.144	0.000	+	KNN	-0.032	0.053	
LogReg	-0.334	0.074		LogReg	0.114	0.000	+	LogReg	0.007	0.421	
SVM	0.148	0.159		SVM	0.109	0.000	+	SVM	-0.081	0.000	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.649	0.042		Bayes	-0.237	0.002	-	Bayes	-0.091	0.000	-
C4.5	-0.215	0.117		C4.5	-0.193	0.048		C4.5	-0.021	0.159	
KNN	-0.297	0.127		KNN	-0.207	0.078		KNN	-0.080	0.000	-
LogReg	-0.268	0.136		LogReg	-0.240	0.002	-	LogReg	-0.050	0.000	-
SVM	0.203	0.206		SVM	-0.195	0.066		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.573	0.002	-	Bayes	0.053	0.026		Bayes	0.217	0.000	+
C4.5	-0.476	0.000	-	C4.5	-0.065	0.000	-	C4.5	0.268	0.000	+
KNN	-0.422	0.026		KNN	-0.068	0.000	-	KNN	0.162	0.000	+
LogReg	-0.279	0.061		LogReg	-0.123	0.000	-	LogReg	0.128	0.000	+
SVM	-0.015	0.116		SVM	-0.107	0.000	-	SVM	0.122	0.000	+
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	0.029	0.181		Bayes	-0.562	0.003	-	Bayes	-0.331	0.000	-
C4.5	0.191	0.000	+	C4.5	-0.514	0.043		C4.5	-0.609	0.000	-
KNN	0.085	0.000	+	KNN	-0.549	0.001	-	KNN	-0.445	0.000	-
LogReg	0.034	0.173		LogReg	-0.565	0.001	-	LogReg	-0.265	0.000	-
SVM	-0.024	0.061		SVM	-0.394	0.040		SVM	-0.172	0.024	-

Table 23 p value of area under ROC curve for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.228	0.000	-	Bayes	-0.155	0.000	-	Bayes	-0.049	0.000	-
C4.5	-0.083	0.000	-	C4.5	-0.052	0.127		C4.5	0.068	0.000	+
KNN	-0.206	0.000	-	KNN	-0.062	0.025	-	KNN	-0.052	0.000	-
LogReg	-0.250	0.000	-	LogReg	-0.159	0.000	-	LogReg	-0.051	0.000	-
SVM	0.002	0.092		SVM	-0.075	0.022	-	SVM	-0.060	0.000	-
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.090	0.000	-	Bayes	-0.367	0.000	-	Bayes	-0.225	0.000	-
C4.5	-0.024	0.086		C4.5	-0.355	0.000	-	C4.5	-0.134	0.000	-
KNN	-0.101	0.000	-	KNN	-0.365	0.000	-	KNN	-0.194	0.000	-
LogReg	-0.099	0.000	-	LogReg	-0.296	0.000	-	LogReg	-0.133	0.000	-
SVM	-0.098	0.000	-	SVM	-0.392	0.000	-	SVM	-0.202	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.021	0.000	-	Bayes	-0.060	0.011	-	Bayes	-0.063	0.002	-
C4.5	0.041	0.000	+	C4.5	-0.028	0.207		C4.5	-0.046	0.242	
KNN	-0.020	0.000	-	KNN	-0.070	0.003	-	KNN	-0.102	0.017	-
LogReg	-0.022	0.000	-	LogReg	-0.003	0.028		LogReg	-0.087	0.203	
SVM	-0.022	0.000	-	SVM	-0.062	0.009	-	SVM	0.313	0.000	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.137	0.000	-	Bayes	-0.351	0.000	-	Bayes	-0.198	0.000	-
C4.5	-0.053	0.032		C4.5	-0.230	0.000	-	C4.5	-0.020	0.446	
KNN	-0.118	0.000	-	KNN	-0.316	0.000	-	KNN	-0.117	0.000	-
LogReg	-0.284	0.000	-	LogReg	-0.347	0.000	-	LogReg	-0.123	0.000	-
SVM	0.024	0.190		SVM	-0.320	0.000	-	SVM	0.193	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.179	0.010	-	Bayes	-0.313	0.000	-	Bayes	-0.003	0.586	
C4.5	-0.043	0.053		C4.5	-0.149	0.000	-	C4.5	0.069	0.000	+
KNN	-0.185	0.032		KNN	-0.281	0.000	-	KNN	-0.012	0.300	
LogReg	-0.130	0.031		LogReg	-0.312	0.000	-	LogReg	0.004	0.639	
SVM	0.232	0.000	+	SVM	-0.316	0.000	-	SVM	-0.052	0.001	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.171	0.027		Bayes	-0.118	0.001	-	Bayes	-0.057	0.000	-
C4.5	-0.033	0.188		C4.5	-0.071	0.166		C4.5	0.060	0.000	+
KNN	-0.133	0.019	-	KNN	-0.127	0.000	-	KNN	-0.054	0.000	-
LogReg	-0.109	0.046		LogReg	-0.130	0.000	-	LogReg	-0.061	0.000	-
SVM	0.093	0.027		SVM	-0.121	0.000	-	SVM	-0.061	0.000	-
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.243	0.000	-	Bayes	-0.121	0.001	-	Bayes	-0.397	0.000	-
C4.5	-0.114	0.039		C4.5	-0.016	0.003	-	C4.5	-0.218	0.000	-
KNN	-0.240	0.003	-	KNN	-0.127	0.000	-	KNN	-0.374	0.000	-
LogReg	-0.207	0.009	-	LogReg	-0.154	0.000	-	LogReg	-0.401	0.000	-
SVM	0.048	0.004	+	SVM	-0.166	0.000	-	SVM	-0.403	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.148	0.000	-	Bayes	-0.189	0.000	-	Bayes	-0.152	0.003	-
C4.5	-0.050	0.123		C4.5	-0.132	0.061		C4.5	-0.189	0.000	-
KNN	-0.107	0.007	-	KNN	-0.189	0.000	-	KNN	-0.202	0.000	-
LogReg	-0.125	0.001	-	LogReg	-0.188	0.000	-	LogReg	-0.135	0.046	
SVM	0.017	0.175		SVM	-0.105	0.000	-	SVM	-0.151	0.039	

Table 24 p value of information score for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.066	0.000	-	Bayes	0.124	0.181		Bayes	-0.107	0.000	-
C4.5	-0.045	0.010	-	C4.5	0.083	0.167		C4.5	-0.028	0.112	
KNN	-0.076	0.000	-	KNN	0.206	0.020	+	KNN	-0.042	0.001	-
LogReg	-0.069	0.000	-	LogReg	0.110	0.189		LogReg	-0.062	0.000	-
SVM	0.008	0.141		SVM	0.195	0.031		SVM	-0.021	0.054	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	1.201	0.000	+	Bayes	-0.435	0.000	-	Bayes	-0.157	0.000	-
C4.5	1.258	0.000	+	C4.5	-0.538	0.000	-	C4.5	-0.178	0.000	-
KNN	1.198	0.000	+	KNN	-0.359	0.000	-	KNN	-0.052	0.253	
LogReg	1.288	0.000	+	LogReg	-0.416	0.000	-	LogReg	-0.007	0.044	
SVM	1.233	0.000	+	SVM	-0.440	0.000	-	SVM	-0.046	0.201	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	2.234	0.000	+	Bayes	3.589	0.000	+	Bayes	3.247	0.031	
C4.5	2.299	0.000	+	C4.5	3.521	0.000	+	C4.5	3.283	0.052	
KNN	2.272	0.000	+	KNN	3.515	0.000	+	KNN	3.251	0.026	
LogReg	2.272	0.000	+	LogReg	3.696	0.000	+	LogReg	3.281	0.048	
SVM	2.280	0.000	+	SVM	3.602	0.000	+	SVM	3.681	0.017	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.156	0.002	+	Bayes	-0.320	0.000	-	Bayes	-0.170	0.005	-
C4.5	0.179	0.010	+	C4.5	-0.226	0.000	-	C4.5	-0.014	0.627	
KNN	0.141	0.000	+	KNN	-0.184	0.000	-	KNN	-0.008	0.422	
LogReg	0.027	0.396		LogReg	-0.207	0.000	-	LogReg	-0.067	0.178	
SVM	0.160	0.000	+	SVM	-0.162	0.000	-	SVM	0.198	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	2.116	0.040		Bayes	-0.311	0.000	-	Bayes	-0.076	0.014	-
C4.5	1.993	0.056		C4.5	-0.243	0.000	-	C4.5	-0.056	0.052	
KNN	1.971	0.056		KNN	-0.238	0.000	-	KNN	0.021	0.170	
LogReg	2.110	0.020	+	LogReg	-0.277	0.000	-	LogReg	-0.022	0.248	
SVM	2.321	0.004	+	SVM	-0.265	0.000	-	SVM	-0.049	0.015	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	5.861	0.006	+	Bayes	2.166	0.041		Bayes	-0.119	0.000	-
C4.5	3.548	0.042		C4.5	2.123	0.031		C4.5	-0.039	0.088	
KNN	3.529	0.037		KNN	2.124	0.035		KNN	-0.050	0.000	-
LogReg	3.538	0.047		LogReg	2.240	0.033		LogReg	-0.064	0.005	-
SVM	3.737	0.013	+	SVM	2.190	0.048		SVM	-0.019	0.064	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	1.266	0.097		Bayes	0.386	0.037		Bayes	-0.523	0.000	-
C4.5	1.118	0.081		C4.5	0.314	0.065		C4.5	-0.394	0.000	-
KNN	1.158	0.080		KNN	0.395	0.044		KNN	-0.476	0.000	-
LogReg	1.295	0.054		LogReg	0.524	0.005	+	LogReg	-0.531	0.000	-
SVM	1.334	0.043		SVM	0.340	0.047		SVM	-0.495	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.044	0.105		Bayes	1.477	0.003	+	Bayes	0.069	0.105	
C4.5	0.002	0.436		C4.5	1.545	0.003	+	C4.5	-0.184	0.052	
KNN	-0.002	0.494		KNN	1.513	0.001	+	KNN	-0.028	0.064	
LogReg	-0.020	0.257		LogReg	1.476	0.005	+	LogReg	0.098	0.137	
SVM	0.043	0.002	+	SVM	1.657	0.002	+	SVM	0.109	0.075	

Table 25 p value of F-measure for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.409	0.000	-	Bayes	-0.543	0.000	-	Bayes	-0.012	0.139	
C4.5	-0.315	0.000	-	C4.5	-0.428	0.000	-	C4.5	0.043	0.000	+
KNN	-0.363	0.000	-	KNN	-0.308	0.000	-	KNN	-0.018	0.020	-
LogReg	-0.185	0.000	-	LogReg	-0.345	0.000	-	LogReg	-0.002	0.743	
SVM	0.025	0.069		SVM	-0.067	0.002	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.031	0.000	-	Bayes	-0.119	0.000	-	Bayes	-0.003	0.360	
C4.5	-0.008	0.077		C4.5	-0.156	0.000	-	C4.5	0.035	0.001	+
KNN	-0.053	0.000	-	KNN	-0.112	0.000	-	KNN	0.039	0.005	+
LogReg	-0.043	0.000	-	LogReg	-0.123	0.000	-	LogReg	0.043	0.000	+
SVM	-0.036	0.000	-	SVM	-0.157	0.000	-	SVM	-0.002	0.353	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.013	0.099		Bayes	-0.037	0.154		Bayes	-0.021	0.094	
C4.5	0.034	0.004	+	C4.5	-0.060	0.004	-	C4.5	-0.012	0.224	
KNN	-0.006	0.460		KNN	-0.066	0.003	-	KNN	-0.027	0.136	
LogReg	0.003	0.343		LogReg	-0.016	0.004	-	LogReg	-0.014	0.150	
SVM	-0.001	0.436		SVM	-0.060	0.005	-	SVM	0.031	0.023	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.306	0.000	-	Bayes	-0.743	0.000	-	Bayes	-0.003	0.630	
C4.5	-0.264	0.000	-	C4.5	-0.663	0.000	-	C4.5	0.041	0.002	+
KNN	-0.302	0.000	-	KNN	-0.719	0.000	-	KNN	0.012	0.161	
LogReg	-0.425	0.000	-	LogReg	-0.744	0.000	-	LogReg	0.013	0.285	
SVM	0.095	0.020	+	SVM	-0.715	0.000	-	SVM	0.026	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.024	0.084		Bayes	-0.639	0.000	-	Bayes	0.051	0.002	+
C4.5	-0.012	0.075		C4.5	-0.552	0.000	-	C4.5	0.091	0.000	+
KNN	-0.031	0.034		KNN	-0.574	0.000	-	KNN	0.065	0.000	+
LogReg	-0.023	0.086		LogReg	-0.624	0.000	-	LogReg	0.066	0.001	+
SVM	0.008	0.317		SVM	-0.607	0.000	-	SVM	0.001	0.688	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.339	0.030		Bayes	-0.062	0.010	-	Bayes	0.010	0.132	
C4.5	-0.026	0.292		C4.5	-0.077	0.013	-	C4.5	0.073	0.000	+
KNN	-0.043	0.309		KNN	-0.084	0.000	-	KNN	0.001	0.467	
LogReg	-0.038	0.173		LogReg	-0.099	0.000	-	LogReg	0.011	0.221	
SVM	-0.019	0.402		SVM	-0.076	0.008	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.003	0.191		Bayes	-0.236	0.267		Bayes	-0.788	0.000	-
C4.5	-0.015	0.094		C4.5	-0.206	0.267		C4.5	-0.680	0.000	-
KNN	-0.025	0.080		KNN	-0.218	0.234		KNN	-0.759	0.000	-
LogReg	-0.007	0.053		LogReg	-0.255	0.268		LogReg	-0.792	0.000	-
SVM	-0.003	0.288		SVM	-0.279	0.119		SVM	-0.797	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.205	0.063		Bayes	-0.138	0.000	-	Bayes	0.015	0.161	
C4.5	-0.270	0.011	-	C4.5	-0.111	0.029		C4.5	-0.007	0.076	
KNN	-0.177	0.052		KNN	-0.129	0.000	-	KNN	0.006	0.204	
LogReg	-0.191	0.040		LogReg	-0.137	0.000	-	LogReg	0.025	0.202	
SVM	0.092	0.114		SVM	-0.113	0.061		SVM	0.002	0.075	

Table 26 p value of precision for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.255	0.186		Bayes	-0.425	0.000	-	Bayes	0.074	0.000	+
C4.5	-0.146	0.077		C4.5	-0.414	0.000	-	C4.5	0.118	0.000	+
KNN	-0.309	0.079		KNN	-0.366	0.000	-	KNN	0.054	0.000	+
LogReg	-0.378	0.078		LogReg	-0.502	0.000	-	LogReg	0.048	0.000	+
SVM	0.240	0.073		SVM	-0.203	0.021	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.069	0.000	-	Bayes	-0.242	0.000	-	Bayes	-0.116	0.000	-
C4.5	-0.064	0.000	-	C4.5	-0.275	0.000	-	C4.5	-0.116	0.000	-
KNN	-0.085	0.000	-	KNN	-0.154	0.000	-	KNN	-0.140	0.000	-
LogReg	-0.065	0.000	-	LogReg	-0.206	0.000	-	LogReg	-0.064	0.000	-
SVM	-0.069	0.000	-	SVM	-0.241	0.000	-	SVM	-0.097	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.020	0.052		Bayes	-0.054	0.095		Bayes	-0.041	0.062	
C4.5	-0.007	0.475		C4.5	-0.036	0.146		C4.5	-0.034	0.181	
KNN	-0.033	0.003	-	KNN	-0.042	0.163		KNN	-0.045	0.093	
LogReg	-0.030	0.008	-	LogReg	0.017	0.076		LogReg	-0.045	0.177	
SVM	-0.036	0.001	-	SVM	-0.037	0.184		SVM	0.066	0.008	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.051	0.338		Bayes	-0.019	0.033		Bayes	-0.057	0.001	-
C4.5	-0.022	0.300		C4.5	-0.015	0.024	-	C4.5	-0.023	0.087	
KNN	-0.212	0.129		KNN	-0.030	0.097		KNN	-0.016	0.128	
LogReg	-0.266	0.062		LogReg	-0.056	0.264		LogReg	-0.022	0.127	
SVM	0.304	0.024	+	SVM	-0.015	0.042		SVM	0.078	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.128	0.000	-	Bayes	-0.525	0.000	-	Bayes	0.052	0.014	+
C4.5	-0.080	0.000	-	C4.5	-0.445	0.000	-	C4.5	0.048	0.006	+
KNN	-0.081	0.005	-	KNN	-0.547	0.000	-	KNN	-0.006	0.468	
LogReg	-0.060	0.050		LogReg	-0.613	0.000	-	LogReg	0.034	0.186	
SVM	0.027	0.366		SVM	-0.613	0.000	-	SVM	-0.092	0.000	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.264	0.039		Bayes	-0.093	0.002	-	Bayes	-0.072	0.000	-
C4.5	-0.015	0.218		C4.5	-0.073	0.031		C4.5	0.010	0.389	
KNN	-0.030	0.305		KNN	-0.080	0.038		KNN	-0.063	0.000	-
LogReg	-0.022	0.224		LogReg	-0.096	0.000	-	LogReg	-0.034	0.001	-
SVM	0.026	0.451		SVM	-0.074	0.036		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.093	0.002	-	Bayes	-0.115	0.127		Bayes	-0.663	0.000	-
C4.5	-0.074	0.001	-	C4.5	-0.194	0.002	-	C4.5	-0.584	0.000	-
KNN	-0.076	0.035		KNN	-0.205	0.002	-	KNN	-0.693	0.000	-
LogReg	-0.039	0.119		LogReg	-0.296	0.000	-	LogReg	-0.733	0.000	-
SVM	-0.015	0.135		SVM	-0.279	0.000	-	SVM	-0.740	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.394	0.025	-	Bayes	-0.204	0.003	-	Bayes	-0.059	0.000	-
C4.5	-0.161	0.139		C4.5	-0.182	0.057		C4.5	-0.151	0.000	-
KNN	-0.183	0.166		KNN	-0.195	0.001	-	KNN	-0.095	0.000	-
LogReg	-0.364	0.065		LogReg	-0.206	0.001	-	LogReg	-0.041	0.165	
SVM	0.216	0.113		SVM	-0.158	0.072		SVM	-0.033	0.218	

Table 27 p value of recall for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.381	0.000	-	Bayes	-0.698	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.295	0.000	-	C4.5	-0.439	0.000	-	C4.5	-0.023	0.067	
KNN	-0.293	0.000	-	KNN	-0.261	0.000	-	KNN	-0.081	0.000	-
LogReg	-0.116	0.000	-	LogReg	-0.265	0.000	-	LogReg	-0.044	0.000	-
SVM	0.014	0.069		SVM	-0.047	0.003	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	0.011	0.214		Bayes	0.067	0.240		Bayes	0.207	0.000	+
C4.5	0.051	0.000	+	C4.5	0.027	0.093		C4.5	0.284	0.000	+
KNN	-0.016	0.013	-	KNN	-0.018	0.010	-	KNN	0.316	0.000	+
LogReg	-0.015	0.010	-	LogReg	0.023	0.085		LogReg	0.246	0.000	+
SVM	0.003	0.382		SVM	-0.010	0.013	-	SVM	0.184	0.000	+
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.006	0.545		Bayes	-0.004	0.338		Bayes	0.003	0.080	
C4.5	0.074	0.000	+	C4.5	-0.062	0.102		C4.5	0.013	0.488	
KNN	0.023	0.098		KNN	-0.066	0.107		KNN	-0.006	0.055	
LogReg	0.036	0.015	+	LogReg	-0.027	0.240		LogReg	0.020	0.173	
SVM	0.035	0.052		SVM	-0.061	0.112		SVM	-0.012	0.016	-
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.420	0.000	-	Bayes	-0.819	0.000	-	Bayes	0.055	0.001	+
C4.5	-0.359	0.000	-	C4.5	-0.659	0.000	-	C4.5	0.107	0.000	+
KNN	-0.279	0.000	-	KNN	-0.755	0.000	-	KNN	0.043	0.004	+
LogReg	-0.449	0.000	-	LogReg	-0.777	0.000	-	LogReg	0.050	0.024	+
SVM	0.060	0.019	+	SVM	-0.763	0.000	-	SVM	-0.048	0.000	-
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.164	0.022	+	Bayes	-0.659	0.000	-	Bayes	0.047	0.069	
C4.5	0.066	0.103		C4.5	-0.563	0.000	-	C4.5	0.134	0.000	+
KNN	0.029	0.208		KNN	-0.521	0.000	-	KNN	0.133	0.000	+
LogReg	0.021	0.177		LogReg	-0.561	0.000	-	LogReg	0.100	0.000	+
SVM	-0.016	0.114		SVM	-0.535	0.000	-	SVM	0.088	0.000	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.390	0.034		Bayes	-0.016	0.217		Bayes	0.101	0.000	+
C4.5	-0.022	0.075		C4.5	-0.062	0.110		C4.5	0.146	0.000	+
KNN	-0.041	0.116		KNN	-0.069	0.096		KNN	0.075	0.000	+
LogReg	-0.039	0.080		LogReg	-0.082	0.028		LogReg	0.066	0.000	+
SVM	-0.055	0.168		SVM	-0.060	0.090		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.109	0.066		Bayes	-0.233	0.183		Bayes	-0.851	0.000	-
C4.5	0.062	0.097		C4.5	-0.059	0.004	-	C4.5	-0.709	0.000	-
KNN	0.045	0.113		KNN	-0.071	0.007	-	KNN	-0.758	0.000	-
LogReg	0.038	0.133		LogReg	-0.068	0.007	-	LogReg	-0.783	0.000	-
SVM	0.021	0.072		SVM	-0.119	0.026		SVM	-0.787	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.138	0.060		Bayes	-0.040	0.170		Bayes	0.137	0.000	+
C4.5	-0.294	0.000	-	C4.5	-0.009	0.260		C4.5	0.192	0.000	+
KNN	-0.143	0.064		KNN	-0.033	0.117		KNN	0.159	0.000	+
LogReg	-0.130	0.055		LogReg	-0.036	0.196		LogReg	0.133	0.000	+
SVM	0.067	0.117		SVM	-0.043	0.131		SVM	0.064	0.007	+

Table 28 p value of Brier score for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.068	0.000	+	Bayes	56.025	0.040		Bayes	0.005	0.679	
C4.5	-0.146	0.000	-	C4.5	55.906	0.040		C4.5	-0.168	0.000	-
KNN	0.063	0.001	+	KNN	56.088	0.039		KNN	0.020	0.041	
LogReg	0.087	0.000	+	LogReg	56.145	0.039		LogReg	0.017	0.110	
SVM	0.033	0.057		SVM	56.105	0.039		SVM	0.012	0.202	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	1807.317	0.000	+	Bayes	11.265	0.003	+	Bayes	3.378	0.037	
C4.5	1807.255	0.000	+	C4.5	11.328	0.003	+	C4.5	3.145	0.060	
KNN	1807.366	0.000	+	KNN	11.255	0.003	+	KNN	3.361	0.039	
LogReg	1807.344	0.000	+	LogReg	11.265	0.003	+	LogReg	3.336	0.041	
SVM	1807.350	0.000	+	SVM	11.356	0.003	+	SVM	3.374	0.038	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	16261.283	0.028		Bayes	2173.510	0.000	+	Bayes	21359.722	0.061	
C4.5	16261.215	0.028		C4.5	2173.525	0.000	+	C4.5	21359.695	0.064	
KNN	16261.282	0.028		KNN	2173.559	0.000	+	KNN	21359.745	0.060	
LogReg	16261.276	0.028		LogReg	2173.471	0.000	+	LogReg	21359.707	0.062	
SVM	16261.280	0.028		SVM	2173.537	0.000	+	SVM	21359.564	0.065	
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	4.540	0.038		Bayes	42.940	0.000	+	Bayes	2.744	0.470	
C4.5	4.396	0.044		C4.5	42.783	0.000	+	C4.5	2.579	0.082	
KNN	4.709	0.034		KNN	42.927	0.000	+	KNN	2.726	0.684	
LogReg	4.773	0.033		LogReg	42.970	0.000	+	LogReg	2.713	0.497	
SVM	4.718	0.034		SVM	42.933	0.000	+	SVM	2.658	0.081	
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	676.766	0.019	+	Bayes	3.624	0.148		Bayes	-0.078	0.000	-
C4.5	676.794	0.019	+	C4.5	3.351	0.189		C4.5	-0.180	0.000	-
KNN	676.903	0.019	+	KNN	3.621	0.148		KNN	-0.023	0.068	
LogReg	676.878	0.019	+	LogReg	3.651	0.144		LogReg	-0.056	0.005	-
SVM	676.759	0.019	+	SVM	3.652	0.144		SVM	0.035	0.024	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	21550.521	0.032		Bayes	2211.389	0.034		Bayes	0.016	0.225	
C4.5	21551.195	0.033		C4.5	2211.393	0.032		C4.5	-0.156	0.000	-
KNN	21551.258	0.009	+	KNN	2211.442	0.032		KNN	0.026	0.002	+
LogReg	21551.250	0.009	+	LogReg	2211.426	0.033		LogReg	0.030	0.002	+
SVM	21551.180	0.013	+	SVM	2211.419	0.033		SVM	0.011	0.161	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	2043.266	0.041		Bayes	374.950	0.018	+	Bayes	0.676	0.000	+
C4.5	2043.226	0.037		C4.5	374.880	0.018	+	C4.5	0.385	0.181	
KNN	2043.312	0.037		KNN	375.021	0.178		KNN	0.652	0.000	+
LogReg	2043.274	0.037		LogReg	375.014	0.159		LogReg	0.696	0.000	+
SVM	2043.243	0.051		SVM	375.070	0.018	+	SVM	0.704	0.000	+
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	1.280	0.116		Bayes	904.115	0.000	+	Bayes	4.609	0.016	+
C4.5	0.994	0.361		C4.5	904.039	0.000	+	C4.5	4.532	0.023	+
KNN	1.239	0.211		KNN	904.103	0.000	+	KNN	4.631	0.012	+
LogReg	1.268	0.137		LogReg	904.115	0.000	+	LogReg	4.601	0.017	+
SVM	1.256	0.156		SVM	904.048	0.000	+	SVM	4.614	0.014	+

Table 29 p value of MCC for each dataset and classifier (posMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.253	0.000	-	Bayes	-0.322	0.000	-	Bayes	0.015	0.363	
C4.5	-0.132	0.000	-	C4.5	-0.214	0.000	-	C4.5	0.134	0.000	+
KNN	-0.246	0.000	-	KNN	-0.111	0.000	-	KNN	-0.004	0.522	
LogReg	-0.166	0.000	-	LogReg	-0.217	0.000	-	LogReg	0.022	0.131	
SVM	0.035	0.188		SVM	-0.032	0.156		SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.091	0.000	-	Bayes	-0.530	0.000	-	Bayes	-0.194	0.000	-
C4.5	-0.037	0.151		C4.5	-0.628	0.000	-	C4.5	-0.149	0.000	-
KNN	-0.148	0.000	-	KNN	-0.476	0.000	-	KNN	-0.175	0.000	-
LogReg	-0.120	0.000	-	LogReg	-0.519	0.000	-	LogReg	-0.065	0.030	
SVM	-0.102	0.000	-	SVM	-0.622	0.000	-	SVM	-0.170	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.020	0.099		Bayes	-0.121	0.090		Bayes	-0.247	0.108	
C4.5	0.049	0.006	+	C4.5	-0.161	0.008	-	C4.5	-0.202	0.208	
KNN	-0.009	0.448		KNN	-0.176	0.005	-	KNN	-0.290	0.135	
LogReg	0.003	0.373		LogReg	-0.009	0.016	-	LogReg	-0.224	0.156	
SVM	-0.003	0.473		SVM	-0.161	0.019	-	SVM	0.497	0.010	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.116	0.119		Bayes	-0.551	0.000	-	Bayes	-0.122	0.057	
C4.5	-0.062	0.303		C4.5	-0.434	0.000	-	C4.5	0.042	0.491	
KNN	-0.201	0.001	-	KNN	-0.516	0.000	-	KNN	-0.011	0.733	
LogReg	-0.320	0.000	-	LogReg	-0.563	0.000	-	LogReg	-0.014	0.748	
SVM	0.080	0.108		SVM	-0.504	0.000	-	SVM	0.461	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.465	0.014	-	Bayes	-0.452	0.000	-	Bayes	0.083	0.001	+
C4.5	-0.399	0.001	-	C4.5	-0.320	0.000	-	C4.5	0.130	0.000	+
KNN	-0.450	0.006	-	KNN	-0.401	0.000	-	KNN	0.078	0.002	+
LogReg	-0.322	0.067		LogReg	-0.478	0.000	-	LogReg	0.096	0.003	+
SVM	0.156	0.244		SVM	-0.463	0.000	-	SVM	-0.026	0.226	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.135	0.098		Bayes	-0.230	0.003	-	Bayes	0.002	0.272	
C4.5	-0.206	0.175		C4.5	-0.256	0.011	-	C4.5	0.120	0.000	+
KNN	-0.327	0.081		KNN	-0.277	0.000	-	KNN	-0.009	0.510	
LogReg	-0.297	0.062		LogReg	-0.320	0.000	-	LogReg	0.014	0.184	
SVM	0.163	0.137		SVM	-0.255	0.009	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.392	0.019	-	Bayes	-0.154	0.272		Bayes	-0.644	0.000	-
C4.5	-0.393	0.001	-	C4.5	-0.134	0.146		C4.5	-0.452	0.000	-
KNN	-0.379	0.044		KNN	-0.149	0.212		KNN	-0.611	0.000	-
LogReg	-0.244	0.150		LogReg	-0.223	0.127		LogReg	-0.673	0.000	-
SVM	0.020	0.120		SVM	-0.243	0.072		SVM	-0.682	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.175	0.069		Bayes	-0.566	0.000	-	Bayes	-0.202	0.001	-
C4.5	-0.089	0.149		C4.5	-0.489	0.026		C4.5	-0.395	0.000	-
KNN	-0.068	0.156		KNN	-0.544	0.000	-	KNN	-0.284	0.000	-
LogReg	-0.156	0.068		LogReg	-0.565	0.000	-	LogReg	-0.129	0.084	
SVM	0.060	0.108		SVM	-0.403	0.043		SVM	-0.115	0.286	

APPENDIX C P VALUE OF PERFORMANCE MEASURES (negMAL)

Table 30 p value of accuracy for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.005	0.497		Bayes	0.047	0.000	+	Bayes	0.001	0.589	
C4.5	0.045	0.000	+	C4.5	0.036	0.037		C4.5	0.060	0.000	+
KNN	-0.010	0.035		KNN	0.044	0.000	+	KNN	-0.008	0.267	
LogReg	-0.011	0.018	-	LogReg	-0.007	0.385		LogReg	0.006	0.445	
SVM	0.000	0.235		SVM	0.007	0.140		SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.044	0.000	-	Bayes	-0.210	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.017	0.130		C4.5	-0.256	0.000	-	C4.5	-0.062	0.000	-
KNN	-0.070	0.000	-	KNN	-0.187	0.000	-	KNN	-0.070	0.000	-
LogReg	-0.057	0.000	-	LogReg	-0.209	0.000	-	LogReg	-0.029	0.003	-
SVM	-0.049	0.000	-	SVM	-0.255	0.000	-	SVM	-0.078	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.009	0.092		Bayes	-0.050	0.129		Bayes	-0.038	0.095	
C4.5	0.021	0.009	+	C4.5	-0.075	0.005	-	C4.5	-0.024	0.222	
KNN	-0.005	0.389		KNN	-0.083	0.003	-	KNN	-0.049	0.134	
LogReg	0.000	0.432		LogReg	-0.012	0.003	-	LogReg	-0.028	0.152	
SVM	-0.002	0.498		SVM	-0.076	0.008	-	SVM	0.059	0.055	
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.091	0.000	+	Bayes	-0.269	0.000	-	Bayes	-0.015	0.310	
C4.5	0.105	0.000	+	C4.5	-0.215	0.000	-	C4.5	0.048	0.013	+
KNN	-0.005	0.458		KNN	-0.255	0.000	-	KNN	0.012	0.361	
LogReg	-0.023	0.102		LogReg	-0.278	0.000	-	LogReg	0.013	0.447	
SVM	-0.003	0.412		SVM	-0.248	0.000	-	SVM	0.059	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.004	0.097		Bayes	-0.099	0.000	-	Bayes	0.041	0.001	+
C4.5	-0.033	0.102		C4.5	-0.040	0.000	-	C4.5	0.055	0.000	+
KNN	-0.061	0.016	-	KNN	-0.089	0.000	-	KNN	0.028	0.017	+
LogReg	-0.045	0.103		LogReg	-0.122	0.000	-	LogReg	0.041	0.011	+
SVM	0.018	0.305		SVM	-0.117	0.000	-	SVM	-0.020	0.096	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.347	0.050		Bayes	-0.090	0.005	-	Bayes	-0.006	0.371	
C4.5	-0.027	0.272		C4.5	-0.106	0.012	-	C4.5	0.053	0.000	+
KNN	-0.053	0.304		KNN	-0.115	0.000	-	KNN	-0.011	0.044	
LogReg	-0.044	0.172		LogReg	-0.135	0.000	-	LogReg	0.001	0.305	
SVM	-0.007	0.390		SVM	-0.105	0.009	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.027	0.111		Bayes	-0.039	0.212		Bayes	-0.258	0.000	-
C4.5	-0.043	0.034		C4.5	-0.063	0.041		C4.5	-0.166	0.000	-
KNN	-0.057	0.084		KNN	-0.069	0.011	-	KNN	-0.247	0.000	-
LogReg	-0.025	0.092		LogReg	-0.105	0.000	-	LogReg	-0.277	0.000	-
SVM	-0.014	0.215		SVM	-0.111	0.000	-	SVM	-0.282	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.015	0.120		Bayes	-0.208	0.000	-	Bayes	-0.021	0.356	
C4.5	0.064	0.000	+	C4.5	-0.173	0.022	-	C4.5	-0.079	0.014	-
KNN	0.026	0.004	+	KNN	-0.197	0.000	-	KNN	-0.046	0.202	
LogReg	-0.009	0.231		LogReg	-0.207	0.000	-	LogReg	-0.001	0.073	
SVM	0.000	0.333		SVM	-0.164	0.038		SVM	-0.018	0.197	

Table 31 p value of sensitivity for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.381	0.000	-	Bayes	-0.698	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.295	0.000	-	C4.5	-0.439	0.000	-	C4.5	-0.023	0.067	
KNN	-0.293	0.000	-	KNN	-0.261	0.000	-	KNN	-0.081	0.000	-
LogReg	-0.116	0.000	-	LogReg	-0.265	0.000	-	LogReg	-0.044	0.000	-
SVM	0.014	0.069		SVM	-0.047	0.003	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	0.011	0.214		Bayes	0.067	0.240		Bayes	0.207	0.000	+
C4.5	0.051	0.000	+	C4.5	0.027	0.093		C4.5	0.284	0.000	+
KNN	-0.016	0.013	-	KNN	-0.018	0.010	-	KNN	0.316	0.000	+
LogReg	-0.015	0.010	-	LogReg	0.023	0.085		LogReg	0.246	0.000	+
SVM	0.003	0.382		SVM	-0.010	0.013	-	SVM	0.184	0.000	+
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.006	0.545		Bayes	-0.004	0.338		Bayes	0.003	0.080	
C4.5	0.074	0.000	+	C4.5	-0.062	0.102		C4.5	0.013	0.488	
KNN	0.023	0.098		KNN	-0.066	0.107		KNN	-0.006	0.055	
LogReg	0.036	0.015	+	LogReg	-0.027	0.240		LogReg	0.020	0.173	
SVM	0.035	0.052		SVM	-0.061	0.112		SVM	-0.012	0.016	-
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.420	0.000	-	Bayes	-0.819	0.000	-	Bayes	0.055	0.001	+
C4.5	-0.359	0.000	-	C4.5	-0.659	0.000	-	C4.5	0.107	0.000	+
KNN	-0.279	0.000	-	KNN	-0.755	0.000	-	KNN	0.043	0.004	+
LogReg	-0.449	0.000	-	LogReg	-0.777	0.000	-	LogReg	0.050	0.024	+
SVM	0.060	0.019	+	SVM	-0.763	0.000	-	SVM	-0.048	0.000	-
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.164	0.022	+	Bayes	-0.659	0.000	-	Bayes	0.047	0.069	
C4.5	0.066	0.103		C4.5	-0.563	0.000	-	C4.5	0.134	0.000	+
KNN	0.029	0.208		KNN	-0.521	0.000	-	KNN	0.133	0.000	+
LogReg	0.021	0.177		LogReg	-0.561	0.000	-	LogReg	0.100	0.000	+
SVM	-0.016	0.114		SVM	-0.535	0.000	-	SVM	0.088	0.000	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.390	0.034		Bayes	-0.016	0.217		Bayes	0.101	0.000	+
C4.5	-0.022	0.075		C4.5	-0.062	0.110		C4.5	0.146	0.000	+
KNN	-0.041	0.116		KNN	-0.069	0.096		KNN	0.075	0.000	+
LogReg	-0.039	0.080		LogReg	-0.082	0.028		LogReg	0.066	0.000	+
SVM	-0.055	0.168		SVM	-0.060	0.090		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.109	0.066		Bayes	-0.233	0.183		Bayes	-0.851	0.000	-
C4.5	0.062	0.097		C4.5	-0.059	0.004	-	C4.5	-0.709	0.000	-
KNN	0.045	0.113		KNN	-0.071	0.007	-	KNN	-0.758	0.000	-
LogReg	0.038	0.133		LogReg	-0.068	0.007	-	LogReg	-0.783	0.000	-
SVM	0.021	0.072		SVM	-0.119	0.026		SVM	-0.787	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.138	0.060		Bayes	-0.040	0.170		Bayes	0.137	0.000	+
C4.5	-0.294	0.000	-	C4.5	-0.009	0.260		C4.5	0.192	0.000	+
KNN	-0.143	0.064		KNN	-0.033	0.117		KNN	0.159	0.000	+
LogReg	-0.130	0.055		LogReg	-0.036	0.196		LogReg	0.133	0.000	+
SVM	0.067	0.117		SVM	-0.043	0.131		SVM	0.064	0.007	+

Table 32 p value of specificity for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.126	0.000	+	Bayes	0.349	0.000	+	Bayes	0.112	0.000	+
C4.5	0.152	0.000	+	C4.5	0.228	0.000	+	C4.5	0.163	0.000	+
KNN	0.079	0.000	+	KNN	0.168	0.000	+	KNN	0.081	0.000	+
LogReg	0.022	0.000	+	LogReg	0.098	0.000	+	LogReg	0.067	0.000	+
SVM	-0.004	0.017	-	SVM	0.028	0.005	+	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.135	0.000	-	Bayes	-0.707	0.000	-	Bayes	-0.489	0.000	-
C4.5	-0.129	0.000	-	C4.5	-0.764	0.000	-	C4.5	-0.532	0.000	-
KNN	-0.161	0.000	-	KNN	-0.488	0.000	-	KNN	-0.594	0.000	-
LogReg	-0.126	0.000	-	LogReg	-0.623	0.000	-	LogReg	-0.400	0.000	-
SVM	-0.134	0.000	-	SVM	-0.694	0.000	-	SVM	-0.433	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.011	0.054		Bayes	-0.142	0.093		Bayes	-0.249	0.064	
C4.5	-0.007	0.176		C4.5	-0.102	0.145		C4.5	-0.224	0.180	
KNN	-0.020	0.002	-	KNN	-0.115	0.163		KNN	-0.280	0.095	
LogReg	-0.019	0.005	-	LogReg	0.018	0.061		LogReg	-0.281	0.177	
SVM	-0.022	0.001	-	SVM	-0.104	0.186		SVM	0.504	0.006	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.245	0.000	+	Bayes	0.207	0.000	+	Bayes	-0.257	0.000	-
C4.5	0.244	0.000	+	C4.5	0.168	0.001	+	C4.5	-0.153	0.004	-
KNN	0.077	0.000	+	KNN	0.178	0.001	+	KNN	-0.095	0.039	
LogReg	0.106	0.000	+	LogReg	0.153	0.002	+	LogReg	-0.115	0.098	
SVM	-0.021	0.025	-	SVM	0.197	0.000	+	SVM	0.429	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.742	0.000	-	Bayes	0.202	0.000	+	Bayes	0.037	0.086	
C4.5	-0.484	0.000	-	C4.5	0.242	0.000	+	C4.5	0.010	0.384	
KNN	-0.475	0.000	-	KNN	0.144	0.000	+	KNN	-0.032	0.053	
LogReg	-0.334	0.074		LogReg	0.114	0.000	+	LogReg	0.007	0.421	
SVM	0.148	0.159		SVM	0.109	0.000	+	SVM	-0.081	0.000	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.649	0.042		Bayes	-0.237	0.002	-	Bayes	-0.091	0.000	-
C4.5	-0.215	0.117		C4.5	-0.193	0.048		C4.5	-0.021	0.159	
KNN	-0.297	0.127		KNN	-0.207	0.078		KNN	-0.080	0.000	-
LogReg	-0.268	0.136		LogReg	-0.240	0.002	-	LogReg	-0.050	0.000	-
SVM	0.203	0.206		SVM	-0.195	0.066		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.573	0.002	-	Bayes	0.053	0.026		Bayes	0.217	0.000	+
C4.5	-0.476	0.000	-	C4.5	-0.065	0.000	-	C4.5	0.268	0.000	+
KNN	-0.422	0.026		KNN	-0.068	0.000	-	KNN	0.162	0.000	+
LogReg	-0.279	0.061		LogReg	-0.123	0.000	-	LogReg	0.128	0.000	+
SVM	-0.015	0.116		SVM	-0.107	0.000	-	SVM	0.122	0.000	+
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	0.029	0.181		Bayes	-0.562	0.003	-	Bayes	-0.331	0.000	-
C4.5	0.191	0.000	+	C4.5	-0.514	0.043		C4.5	-0.609	0.000	-
KNN	0.085	0.000	+	KNN	-0.549	0.001	-	KNN	-0.445	0.000	-
LogReg	0.034	0.173		LogReg	-0.565	0.001	-	LogReg	-0.265	0.000	-
SVM	-0.024	0.061		SVM	-0.394	0.040		SVM	-0.172	0.024	-

Table 33 p value of area under ROC curve for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.228	0.000	-	Bayes	-0.155	0.000	-	Bayes	-0.049	0.000	-
C4.5	-0.083	0.000	-	C4.5	-0.052	0.127		C4.5	0.068	0.000	+
KNN	-0.206	0.000	-	KNN	-0.062	0.025	-	KNN	-0.052	0.000	-
LogReg	-0.250	0.000	-	LogReg	-0.159	0.000	-	LogReg	-0.051	0.000	-
SVM	0.002	0.092		SVM	-0.075	0.022	-	SVM	-0.060	0.000	-
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.090	0.000	-	Bayes	-0.367	0.000	-	Bayes	-0.225	0.000	-
C4.5	-0.024	0.086		C4.5	-0.355	0.000	-	C4.5	-0.134	0.000	-
KNN	-0.101	0.000	-	KNN	-0.365	0.000	-	KNN	-0.194	0.000	-
LogReg	-0.099	0.000	-	LogReg	-0.296	0.000	-	LogReg	-0.133	0.000	-
SVM	-0.098	0.000	-	SVM	-0.392	0.000	-	SVM	-0.202	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.021	0.000	-	Bayes	-0.060	0.011	-	Bayes	-0.063	0.002	-
C4.5	0.041	0.000	+	C4.5	-0.028	0.207		C4.5	-0.046	0.242	
KNN	-0.020	0.000	-	KNN	-0.070	0.003	-	KNN	-0.102	0.017	-
LogReg	-0.022	0.000	-	LogReg	-0.003	0.028		LogReg	-0.087	0.203	
SVM	-0.022	0.000	-	SVM	-0.062	0.009	-	SVM	0.313	0.000	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.137	0.000	-	Bayes	-0.351	0.000	-	Bayes	-0.198	0.000	-
C4.5	-0.053	0.032		C4.5	-0.230	0.000	-	C4.5	-0.020	0.446	
KNN	-0.118	0.000	-	KNN	-0.316	0.000	-	KNN	-0.117	0.000	-
LogReg	-0.284	0.000	-	LogReg	-0.347	0.000	-	LogReg	-0.123	0.000	-
SVM	0.024	0.190		SVM	-0.320	0.000	-	SVM	0.193	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.179	0.010	-	Bayes	-0.313	0.000	-	Bayes	-0.003	0.586	
C4.5	-0.043	0.053		C4.5	-0.149	0.000	-	C4.5	0.069	0.000	+
KNN	-0.185	0.032		KNN	-0.281	0.000	-	KNN	-0.012	0.300	
LogReg	-0.130	0.031		LogReg	-0.312	0.000	-	LogReg	0.004	0.639	
SVM	0.232	0.000	+	SVM	-0.316	0.000	-	SVM	-0.052	0.001	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.171	0.027		Bayes	-0.118	0.001	-	Bayes	-0.057	0.000	-
C4.5	-0.033	0.188		C4.5	-0.071	0.166		C4.5	0.060	0.000	+
KNN	-0.133	0.019	-	KNN	-0.127	0.000	-	KNN	-0.054	0.000	-
LogReg	-0.109	0.046		LogReg	-0.130	0.000	-	LogReg	-0.061	0.000	-
SVM	0.093	0.027		SVM	-0.121	0.000	-	SVM	-0.061	0.000	-
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.243	0.000	-	Bayes	-0.121	0.001	-	Bayes	-0.397	0.000	-
C4.5	-0.114	0.039		C4.5	-0.016	0.003	-	C4.5	-0.218	0.000	-
KNN	-0.240	0.003	-	KNN	-0.127	0.000	-	KNN	-0.374	0.000	-
LogReg	-0.207	0.009	-	LogReg	-0.154	0.000	-	LogReg	-0.401	0.000	-
SVM	0.048	0.004	+	SVM	-0.166	0.000	-	SVM	-0.403	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.148	0.000	-	Bayes	-0.189	0.000	-	Bayes	-0.152	0.003	-
C4.5	-0.050	0.123		C4.5	-0.132	0.061		C4.5	-0.189	0.000	-
KNN	-0.107	0.007	-	KNN	-0.189	0.000	-	KNN	-0.202	0.000	-
LogReg	-0.125	0.001	-	LogReg	-0.188	0.000	-	LogReg	-0.135	0.046	
SVM	0.017	0.175		SVM	-0.105	0.000	-	SVM	-0.151	0.039	

Table 34 p value of information score for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.066	0.000	-	Bayes	0.124	0.181		Bayes	-0.107	0.000	-
C4.5	-0.045	0.010	-	C4.5	0.083	0.167		C4.5	-0.028	0.112	
KNN	-0.076	0.000	-	KNN	0.206	0.020	+	KNN	-0.042	0.001	-
LogReg	-0.069	0.000	-	LogReg	0.110	0.189		LogReg	-0.062	0.000	-
SVM	0.008	0.141		SVM	0.195	0.031		SVM	-0.021	0.054	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	1.201	0.000	+	Bayes	-0.435	0.000	-	Bayes	-0.157	0.000	-
C4.5	1.258	0.000	+	C4.5	-0.538	0.000	-	C4.5	-0.178	0.000	-
KNN	1.198	0.000	+	KNN	-0.359	0.000	-	KNN	-0.052	0.253	
LogReg	1.288	0.000	+	LogReg	-0.416	0.000	-	LogReg	-0.007	0.044	
SVM	1.233	0.000	+	SVM	-0.440	0.000	-	SVM	-0.046	0.201	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	2.234	0.000	+	Bayes	3.589	0.000	+	Bayes	3.247	0.031	
C4.5	2.299	0.000	+	C4.5	3.521	0.000	+	C4.5	3.283	0.052	
KNN	2.272	0.000	+	KNN	3.515	0.000	+	KNN	3.251	0.026	
LogReg	2.272	0.000	+	LogReg	3.696	0.000	+	LogReg	3.281	0.048	
SVM	2.280	0.000	+	SVM	3.602	0.000	+	SVM	3.681	0.017	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	0.156	0.002	+	Bayes	-0.320	0.000	-	Bayes	-0.170	0.005	-
C4.5	0.179	0.010	+	C4.5	-0.226	0.000	-	C4.5	-0.014	0.627	
KNN	0.141	0.000	+	KNN	-0.184	0.000	-	KNN	-0.008	0.422	
LogReg	0.027	0.396		LogReg	-0.207	0.000	-	LogReg	-0.067	0.178	
SVM	0.160	0.000	+	SVM	-0.162	0.000	-	SVM	0.198	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	2.116	0.040		Bayes	-0.311	0.000	-	Bayes	-0.076	0.014	-
C4.5	1.993	0.056		C4.5	-0.243	0.000	-	C4.5	-0.056	0.052	
KNN	1.971	0.056		KNN	-0.238	0.000	-	KNN	0.021	0.170	
LogReg	2.110	0.020	+	LogReg	-0.277	0.000	-	LogReg	-0.022	0.248	
SVM	2.321	0.004	+	SVM	-0.265	0.000	-	SVM	-0.049	0.015	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	5.861	0.006	+	Bayes	2.166	0.041		Bayes	-0.119	0.000	-
C4.5	3.548	0.042		C4.5	2.123	0.031		C4.5	-0.039	0.088	
KNN	3.529	0.037		KNN	2.124	0.035		KNN	-0.050	0.000	-
LogReg	3.538	0.047		LogReg	2.240	0.033		LogReg	-0.064	0.005	-
SVM	3.737	0.013	+	SVM	2.190	0.048		SVM	-0.019	0.064	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	1.266	0.097		Bayes	0.386	0.037		Bayes	-0.523	0.000	-
C4.5	1.118	0.081		C4.5	0.314	0.065		C4.5	-0.394	0.000	-
KNN	1.158	0.080		KNN	0.395	0.044		KNN	-0.476	0.000	-
LogReg	1.295	0.054		LogReg	0.524	0.005	+	LogReg	-0.531	0.000	-
SVM	1.334	0.043		SVM	0.340	0.047		SVM	-0.495	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.044	0.105		Bayes	1.477	0.003	+	Bayes	0.069	0.105	
C4.5	0.002	0.436		C4.5	1.545	0.003	+	C4.5	-0.184	0.052	
KNN	-0.002	0.494		KNN	1.513	0.001	+	KNN	-0.028	0.064	
LogReg	-0.020	0.257		LogReg	1.476	0.005	+	LogReg	0.098	0.137	
SVM	0.043	0.002	+	SVM	1.657	0.002	+	SVM	0.109	0.075	

Table 35 p value of F-measure for each dataset and classifier (negMAL)

Dataset	$\Delta\mu$	p	Sig	Dataset	$\Delta\mu$	p	Sig	Dataset	$\Delta\mu$	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.409	0.000	-	Bayes	-0.543	0.000	-	Bayes	-0.012	0.139	
C4.5	-0.315	0.000	-	C4.5	-0.428	0.000	-	C4.5	0.043	0.000	+
KNN	-0.363	0.000	-	KNN	-0.308	0.000	-	KNN	-0.018	0.020	-
LogReg	-0.185	0.000	-	LogReg	-0.345	0.000	-	LogReg	-0.002	0.743	
SVM	0.025	0.069		SVM	-0.067	0.002	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.031	0.000	-	Bayes	-0.119	0.000	-	Bayes	-0.003	0.360	
C4.5	-0.008	0.077		C4.5	-0.156	0.000	-	C4.5	0.035	0.001	+
KNN	-0.053	0.000	-	KNN	-0.112	0.000	-	KNN	0.039	0.005	+
LogReg	-0.043	0.000	-	LogReg	-0.123	0.000	-	LogReg	0.043	0.000	+
SVM	-0.036	0.000	-	SVM	-0.157	0.000	-	SVM	-0.002	0.353	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.013	0.099		Bayes	-0.037	0.154		Bayes	-0.021	0.094	
C4.5	0.034	0.004	+	C4.5	-0.060	0.004	-	C4.5	-0.012	0.224	
KNN	-0.006	0.460		KNN	-0.066	0.003	-	KNN	-0.027	0.136	
LogReg	0.003	0.343		LogReg	-0.016	0.004	-	LogReg	-0.014	0.150	
SVM	-0.001	0.436		SVM	-0.060	0.005	-	SVM	0.031	0.023	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.306	0.000	-	Bayes	-0.743	0.000	-	Bayes	-0.003	0.630	
C4.5	-0.264	0.000	-	C4.5	-0.663	0.000	-	C4.5	0.041	0.002	+
KNN	-0.302	0.000	-	KNN	-0.719	0.000	-	KNN	0.012	0.161	
LogReg	-0.425	0.000	-	LogReg	-0.744	0.000	-	LogReg	0.013	0.285	
SVM	0.095	0.020	+	SVM	-0.715	0.000	-	SVM	0.026	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.024	0.084		Bayes	-0.639	0.000	-	Bayes	0.051	0.002	+
C4.5	-0.012	0.075		C4.5	-0.552	0.000	-	C4.5	0.091	0.000	+
KNN	-0.031	0.034		KNN	-0.574	0.000	-	KNN	0.065	0.000	+
LogReg	-0.023	0.086		LogReg	-0.624	0.000	-	LogReg	0.066	0.001	+
SVM	0.008	0.317		SVM	-0.607	0.000	-	SVM	0.001	0.688	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.339	0.030		Bayes	-0.062	0.010	-	Bayes	0.010	0.132	
C4.5	-0.026	0.292		C4.5	-0.077	0.013	-	C4.5	0.073	0.000	+
KNN	-0.043	0.309		KNN	-0.084	0.000	-	KNN	0.001	0.467	
LogReg	-0.038	0.173		LogReg	-0.099	0.000	-	LogReg	0.011	0.221	
SVM	-0.019	0.402		SVM	-0.076	0.008	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.003	0.191		Bayes	-0.236	0.267		Bayes	-0.788	0.000	-
C4.5	-0.015	0.094		C4.5	-0.206	0.267		C4.5	-0.680	0.000	-
KNN	-0.025	0.080		KNN	-0.218	0.234		KNN	-0.759	0.000	-
LogReg	-0.007	0.053		LogReg	-0.255	0.268		LogReg	-0.792	0.000	-
SVM	-0.003	0.288		SVM	-0.279	0.119		SVM	-0.797	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.205	0.063		Bayes	-0.138	0.000	-	Bayes	0.015	0.161	
C4.5	-0.270	0.011	-	C4.5	-0.111	0.029		C4.5	-0.007	0.076	
KNN	-0.177	0.052		KNN	-0.129	0.000	-	KNN	0.006	0.204	
LogReg	-0.191	0.040		LogReg	-0.137	0.000	-	LogReg	0.025	0.202	
SVM	0.092	0.114		SVM	-0.113	0.061		SVM	0.002	0.075	

Table 36 p value of precision for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.255	0.186		Bayes	-0.425	0.000	-	Bayes	0.074	0.000	+
C4.5	-0.146	0.077		C4.5	-0.414	0.000	-	C4.5	0.118	0.000	+
KNN	-0.309	0.079		KNN	-0.366	0.000	-	KNN	0.054	0.000	+
LogReg	-0.378	0.078		LogReg	-0.502	0.000	-	LogReg	0.048	0.000	+
SVM	0.240	0.073		SVM	-0.203	0.021	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.069	0.000	-	Bayes	-0.242	0.000	-	Bayes	-0.116	0.000	-
C4.5	-0.064	0.000	-	C4.5	-0.275	0.000	-	C4.5	-0.116	0.000	-
KNN	-0.085	0.000	-	KNN	-0.154	0.000	-	KNN	-0.140	0.000	-
LogReg	-0.065	0.000	-	LogReg	-0.206	0.000	-	LogReg	-0.064	0.000	-
SVM	-0.069	0.000	-	SVM	-0.241	0.000	-	SVM	-0.097	0.000	-
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.020	0.052		Bayes	-0.054	0.095		Bayes	-0.041	0.062	
C4.5	-0.007	0.475		C4.5	-0.036	0.146		C4.5	-0.034	0.181	
KNN	-0.033	0.003	-	KNN	-0.042	0.163		KNN	-0.045	0.093	
LogReg	-0.030	0.008	-	LogReg	0.017	0.076		LogReg	-0.045	0.177	
SVM	-0.036	0.001	-	SVM	-0.037	0.184		SVM	0.066	0.008	+
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.051	0.338		Bayes	-0.019	0.033		Bayes	-0.057	0.001	-
C4.5	-0.022	0.300		C4.5	-0.015	0.024	-	C4.5	-0.023	0.087	
KNN	-0.212	0.129		KNN	-0.030	0.097		KNN	-0.016	0.128	
LogReg	-0.266	0.062		LogReg	-0.056	0.264		LogReg	-0.022	0.127	
SVM	0.304	0.024	+	SVM	-0.015	0.042		SVM	0.078	0.000	+
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.128	0.000	-	Bayes	-0.525	0.000	-	Bayes	0.052	0.014	+
C4.5	-0.080	0.000	-	C4.5	-0.445	0.000	-	C4.5	0.048	0.006	+
KNN	-0.081	0.005	-	KNN	-0.547	0.000	-	KNN	-0.006	0.468	
LogReg	-0.060	0.050		LogReg	-0.613	0.000	-	LogReg	0.034	0.186	
SVM	0.027	0.366		SVM	-0.613	0.000	-	SVM	-0.092	0.000	-
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.264	0.039		Bayes	-0.093	0.002	-	Bayes	-0.072	0.000	-
C4.5	-0.015	0.218		C4.5	-0.073	0.031		C4.5	0.010	0.389	
KNN	-0.030	0.305		KNN	-0.080	0.038		KNN	-0.063	0.000	-
LogReg	-0.022	0.224		LogReg	-0.096	0.000	-	LogReg	-0.034	0.001	-
SVM	0.026	0.451		SVM	-0.074	0.036		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.093	0.002	-	Bayes	-0.115	0.127		Bayes	-0.663	0.000	-
C4.5	-0.074	0.001	-	C4.5	-0.194	0.002	-	C4.5	-0.584	0.000	-
KNN	-0.076	0.035		KNN	-0.205	0.002	-	KNN	-0.693	0.000	-
LogReg	-0.039	0.119		LogReg	-0.296	0.000	-	LogReg	-0.733	0.000	-
SVM	-0.015	0.135		SVM	-0.279	0.000	-	SVM	-0.740	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.394	0.025	-	Bayes	-0.204	0.003	-	Bayes	-0.059	0.000	-
C4.5	-0.161	0.139		C4.5	-0.182	0.057		C4.5	-0.151	0.000	-
KNN	-0.183	0.166		KNN	-0.195	0.001	-	KNN	-0.095	0.000	-
LogReg	-0.364	0.065		LogReg	-0.206	0.001	-	LogReg	-0.041	0.165	
SVM	0.216	0.113		SVM	-0.158	0.072		SVM	-0.033	0.218	

Table 37 p value of recall for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.381	0.000	-	Bayes	-0.698	0.000	-	Bayes	-0.089	0.000	-
C4.5	-0.295	0.000	-	C4.5	-0.439	0.000	-	C4.5	-0.023	0.067	
KNN	-0.293	0.000	-	KNN	-0.261	0.000	-	KNN	-0.081	0.000	-
LogReg	-0.116	0.000	-	LogReg	-0.265	0.000	-	LogReg	-0.044	0.000	-
SVM	0.014	0.069		SVM	-0.047	0.003	-	SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	0.011	0.214		Bayes	0.067	0.240		Bayes	0.207	0.000	+
C4.5	0.051	0.000	+	C4.5	0.027	0.093		C4.5	0.284	0.000	+
KNN	-0.016	0.013	-	KNN	-0.018	0.010	-	KNN	0.316	0.000	+
LogReg	-0.015	0.010	-	LogReg	0.023	0.085		LogReg	0.246	0.000	+
SVM	0.003	0.382		SVM	-0.010	0.013	-	SVM	0.184	0.000	+
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.006	0.545		Bayes	-0.004	0.338		Bayes	0.003	0.080	
C4.5	0.074	0.000	+	C4.5	-0.062	0.102		C4.5	0.013	0.488	
KNN	0.023	0.098		KNN	-0.066	0.107		KNN	-0.006	0.055	
LogReg	0.036	0.015	+	LogReg	-0.027	0.240		LogReg	0.020	0.173	
SVM	0.035	0.052		SVM	-0.061	0.112		SVM	-0.012	0.016	-
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.420	0.000	-	Bayes	-0.819	0.000	-	Bayes	0.055	0.001	+
C4.5	-0.359	0.000	-	C4.5	-0.659	0.000	-	C4.5	0.107	0.000	+
KNN	-0.279	0.000	-	KNN	-0.755	0.000	-	KNN	0.043	0.004	+
LogReg	-0.449	0.000	-	LogReg	-0.777	0.000	-	LogReg	0.050	0.024	+
SVM	0.060	0.019	+	SVM	-0.763	0.000	-	SVM	-0.048	0.000	-
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	0.164	0.022	+	Bayes	-0.659	0.000	-	Bayes	0.047	0.069	
C4.5	0.066	0.103		C4.5	-0.563	0.000	-	C4.5	0.134	0.000	+
KNN	0.029	0.208		KNN	-0.521	0.000	-	KNN	0.133	0.000	+
LogReg	0.021	0.177		LogReg	-0.561	0.000	-	LogReg	0.100	0.000	+
SVM	-0.016	0.114		SVM	-0.535	0.000	-	SVM	0.088	0.000	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	0.390	0.034		Bayes	-0.016	0.217		Bayes	0.101	0.000	+
C4.5	-0.022	0.075		C4.5	-0.062	0.110		C4.5	0.146	0.000	+
KNN	-0.041	0.116		KNN	-0.069	0.096		KNN	0.075	0.000	+
LogReg	-0.039	0.080		LogReg	-0.082	0.028		LogReg	0.066	0.000	+
SVM	-0.055	0.168		SVM	-0.060	0.090		SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	0.109	0.066		Bayes	-0.233	0.183		Bayes	-0.851	0.000	-
C4.5	0.062	0.097		C4.5	-0.059	0.004	-	C4.5	-0.709	0.000	-
KNN	0.045	0.113		KNN	-0.071	0.007	-	KNN	-0.758	0.000	-
LogReg	0.038	0.133		LogReg	-0.068	0.007	-	LogReg	-0.783	0.000	-
SVM	0.021	0.072		SVM	-0.119	0.026		SVM	-0.787	0.000	-
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.138	0.060		Bayes	-0.040	0.170		Bayes	0.137	0.000	+
C4.5	-0.294	0.000	-	C4.5	-0.009	0.260		C4.5	0.192	0.000	+
KNN	-0.143	0.064		KNN	-0.033	0.117		KNN	0.159	0.000	+
LogReg	-0.130	0.055		LogReg	-0.036	0.196		LogReg	0.133	0.000	+
SVM	0.067	0.117		SVM	-0.043	0.131		SVM	0.064	0.007	+

Table 38 p value of Brier score for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	0.068	0.000	+	Bayes	56.025	0.040		Bayes	0.005	0.679	
C4.5	-0.146	0.000	-	C4.5	55.906	0.040		C4.5	-0.168	0.000	-
KNN	0.063	0.001	+	KNN	56.088	0.039		KNN	0.020	0.041	
LogReg	0.087	0.000	+	LogReg	56.145	0.039		LogReg	0.017	0.110	
SVM	0.033	0.057		SVM	56.105	0.039		SVM	0.012	0.202	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	1807.317	0.000	+	Bayes	11.265	0.003	+	Bayes	3.378	0.037	
C4.5	1807.255	0.000	+	C4.5	11.328	0.003	+	C4.5	3.145	0.060	
KNN	1807.366	0.000	+	KNN	11.255	0.003	+	KNN	3.361	0.039	
LogReg	1807.344	0.000	+	LogReg	11.265	0.003	+	LogReg	3.336	0.041	
SVM	1807.350	0.000	+	SVM	11.356	0.003	+	SVM	3.374	0.038	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	16261.283	0.028		Bayes	2173.510	0.000	+	Bayes	21359.722	0.061	
C4.5	16261.215	0.028		C4.5	2173.525	0.000	+	C4.5	21359.695	0.064	
KNN	16261.282	0.028		KNN	2173.559	0.000	+	KNN	21359.745	0.060	
LogReg	16261.276	0.028		LogReg	2173.471	0.000	+	LogReg	21359.707	0.062	
SVM	16261.280	0.028		SVM	2173.537	0.000	+	SVM	21359.564	0.065	
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	4.540	0.038		Bayes	42.940	0.000	+	Bayes	2.744	0.470	
C4.5	4.396	0.044		C4.5	42.783	0.000	+	C4.5	2.579	0.082	
KNN	4.709	0.034		KNN	42.927	0.000	+	KNN	2.726	0.684	
LogReg	4.773	0.033		LogReg	42.970	0.000	+	LogReg	2.713	0.497	
SVM	4.718	0.034		SVM	42.933	0.000	+	SVM	2.658	0.081	
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	676.766	0.019	+	Bayes	3.624	0.148		Bayes	-0.078	0.000	-
C4.5	676.794	0.019	+	C4.5	3.351	0.189		C4.5	-0.180	0.000	-
KNN	676.903	0.019	+	KNN	3.621	0.148		KNN	-0.023	0.068	
LogReg	676.878	0.019	+	LogReg	3.651	0.144		LogReg	-0.056	0.005	-
SVM	676.759	0.019	+	SVM	3.652	0.144		SVM	0.035	0.024	+
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	21550.521	0.032		Bayes	2211.389	0.034		Bayes	0.016	0.225	
C4.5	21551.195	0.033		C4.5	2211.393	0.032		C4.5	-0.156	0.000	-
KNN	21551.258	0.009	+	KNN	2211.442	0.032		KNN	0.026	0.002	+
LogReg	21551.250	0.009	+	LogReg	2211.426	0.033		LogReg	0.030	0.002	+
SVM	21551.180	0.013	+	SVM	2211.419	0.033		SVM	0.011	0.161	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	2043.266	0.041		Bayes	374.950	0.018	+	Bayes	0.676	0.000	+
C4.5	2043.226	0.037		C4.5	374.880	0.018	+	C4.5	0.385	0.181	
KNN	2043.312	0.037		KNN	375.021	0.178		KNN	0.652	0.000	+
LogReg	2043.274	0.037		LogReg	375.014	0.159		LogReg	0.696	0.000	+
SVM	2043.243	0.051		SVM	375.070	0.018	+	SVM	0.704	0.000	+
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	1.280	0.116		Bayes	904.115	0.000	+	Bayes	4.609	0.016	+
C4.5	0.994	0.361		C4.5	904.039	0.000	+	C4.5	4.532	0.023	+
KNN	1.239	0.211		KNN	904.103	0.000	+	KNN	4.631	0.012	+
LogReg	1.268	0.137		LogReg	904.115	0.000	+	LogReg	4.601	0.017	+
SVM	1.256	0.156		SVM	904.048	0.000	+	SVM	4.614	0.014	+

Table 39 p value of MCC for each dataset and classifier (negMAL)

Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig	Dataset	Δμ	p	Sig
01.Blood Transfusion Service Center				09.ILPD (Indian Liver Patient Dataset)				17.Credit Approval			
Bayes	-0.253	0.000	-	Bayes	-0.322	0.000	-	Bayes	0.015	0.363	
C4.5	-0.132	0.000	-	C4.5	-0.214	0.000	-	C4.5	0.134	0.000	
KNN	-0.246	0.000	-	KNN	-0.111	0.000	-	KNN	-0.004	0.522	
LogReg	-0.166	0.000	-	LogReg	-0.217	0.000	-	LogReg	0.022	0.131	
SVM	0.035	0.188		SVM	-0.032	0.156		SVM	0.000	1.000	
02.Breast Cancer Wisconsin (Diagnostic)				10.Ionosphere				18.Cylinder Bands (Continuous attributes)			
Bayes	-0.091	0.000	-	Bayes	-0.530	0.000	-	Bayes	-0.194	0.000	
C4.5	-0.037	0.151		C4.5	-0.628	0.000	-	C4.5	-0.149	0.000	
KNN	-0.148	0.000	-	KNN	-0.476	0.000	-	KNN	-0.175	0.000	
LogReg	-0.120	0.000	-	LogReg	-0.519	0.000	-	LogReg	-0.065	0.030	
SVM	-0.102	0.000	-	SVM	-0.622	0.000	-	SVM	-0.170	0.000	
03.Breast Cancer Wisconsin (Original)				11.Iris				19.Dermatology			
Bayes	-0.020	0.099		Bayes	-0.121	0.090		Bayes	-0.247	0.108	
C4.5	0.049	0.006	+	C4.5	-0.161	0.008	-	C4.5	-0.202	0.208	
KNN	-0.009	0.448		KNN	-0.176	0.005	-	KNN	-0.290	0.135	
LogReg	0.003	0.373		LogReg	-0.009	0.016	-	LogReg	-0.224	0.156	
SVM	-0.003	0.473		SVM	-0.161	0.019	-	SVM	0.497	0.010	
04.Breast Cancer Wisconsin (Prognostic)				12.Mammographic Mass				20.Hepatitis			
Bayes	-0.116	0.119		Bayes	-0.551	0.000	-	Bayes	-0.122	0.057	
C4.5	-0.062	0.303		C4.5	-0.434	0.000	-	C4.5	0.042	0.491	
KNN	-0.201	0.001	-	KNN	-0.516	0.000	-	KNN	-0.011	0.733	
LogReg	-0.320	0.000	-	LogReg	-0.563	0.000	-	LogReg	-0.014	0.748	
SVM	0.080	0.108		SVM	-0.504	0.000	-	SVM	0.461	0.000	
05.Breast Tissue				13.Pima Indians Diabetes				21.Horse Colic			
Bayes	-0.465	0.014	-	Bayes	-0.452	0.000	-	Bayes	0.083	0.001	
C4.5	-0.399	0.001	-	C4.5	-0.320	0.000	-	C4.5	0.130	0.000	
KNN	-0.450	0.006	-	KNN	-0.401	0.000	-	KNN	0.078	0.002	
LogReg	-0.322	0.067		LogReg	-0.478	0.000	-	LogReg	0.096	0.003	
SVM	0.156	0.244		SVM	-0.463	0.000	-	SVM	-0.026	0.226	
06.Ecoli				14.Seeds				22.Statlog (Australian Credit Approval)			
Bayes	-0.135	0.098		Bayes	-0.230	0.003	-	Bayes	0.002	0.272	
C4.5	-0.206	0.175		C4.5	-0.256	0.011	-	C4.5	0.120	0.000	
KNN	-0.327	0.081		KNN	-0.277	0.000	-	KNN	-0.009	0.510	
LogReg	-0.297	0.062		LogReg	-0.320	0.000	-	LogReg	0.014	0.184	
SVM	0.163	0.137		SVM	-0.255	0.009	-	SVM	0.000	1.000	
07.Glass Identification				15.Vertebral Column				23.Statlog (Heart)			
Bayes	-0.392	0.019	-	Bayes	-0.154	0.272		Bayes	-0.644	0.000	
C4.5	-0.393	0.001	-	C4.5	-0.134	0.146		C4.5	-0.452	0.000	
KNN	-0.379	0.044		KNN	-0.149	0.212		KNN	-0.611	0.000	
LogReg	-0.244	0.150		LogReg	-0.223	0.127		LogReg	-0.673	0.000	
SVM	0.020	0.120		SVM	-0.243	0.072		SVM	-0.682	0.000	
08.Haberman's Survival				16.Wine				24.Teaching Assistant Evaluation			
Bayes	-0.175	0.069		Bayes	-0.566	0.000	-	Bayes	-0.202	0.001	
C4.5	-0.089	0.149		C4.5	-0.489	0.026		C4.5	-0.395	0.000	
KNN	-0.068	0.156		KNN	-0.544	0.000	-	KNN	-0.284	0.000	
LogReg	-0.156	0.068		LogReg	-0.565	0.000	-	LogReg	-0.129	0.084	
SVM	0.060	0.108		SVM	-0.403	0.043		SVM	-0.115	0.286	

APPENDIX D LINEAR PROJECTIONS OF THE DATASETS

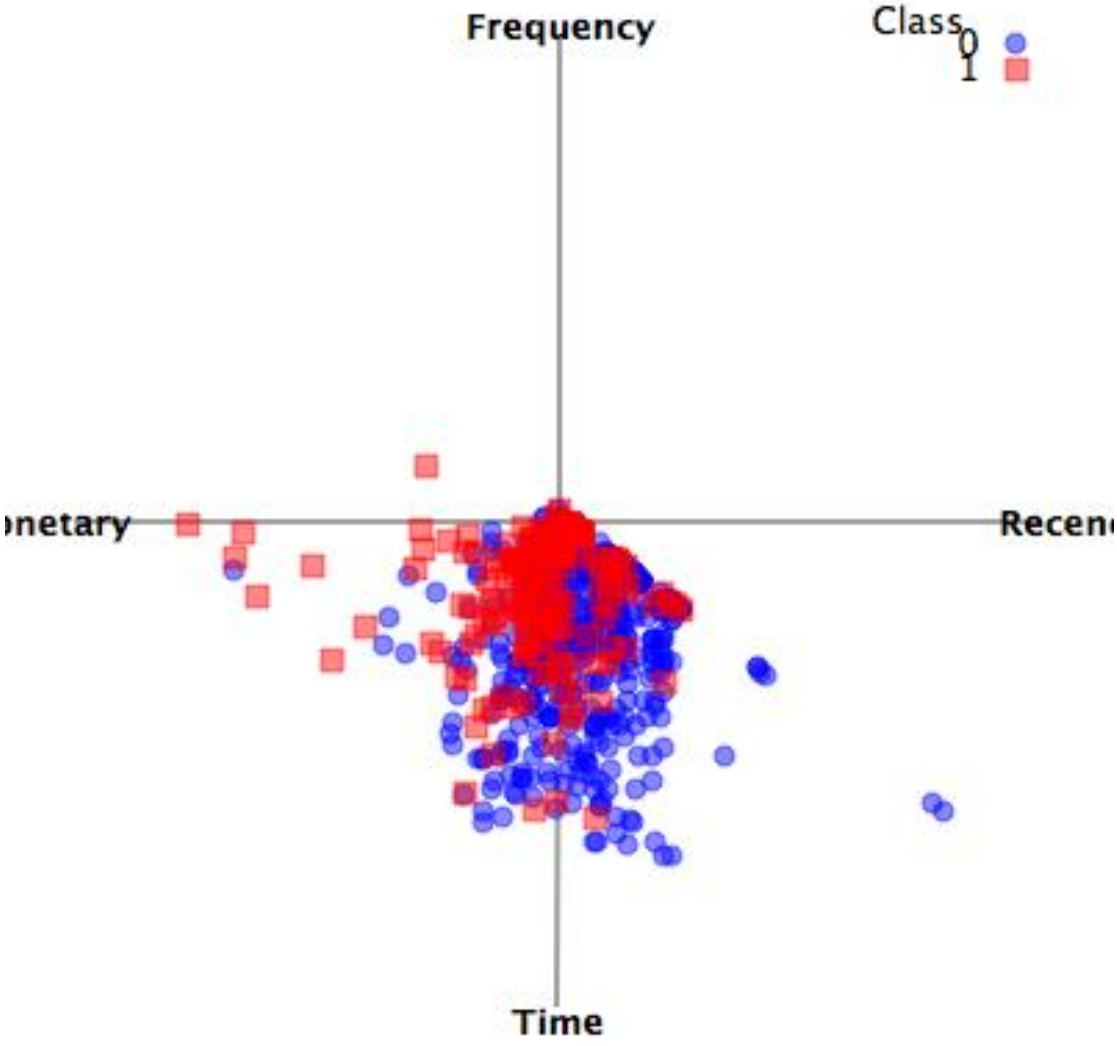

Figure 30 Linear projection of Blood Transfusion Service Center

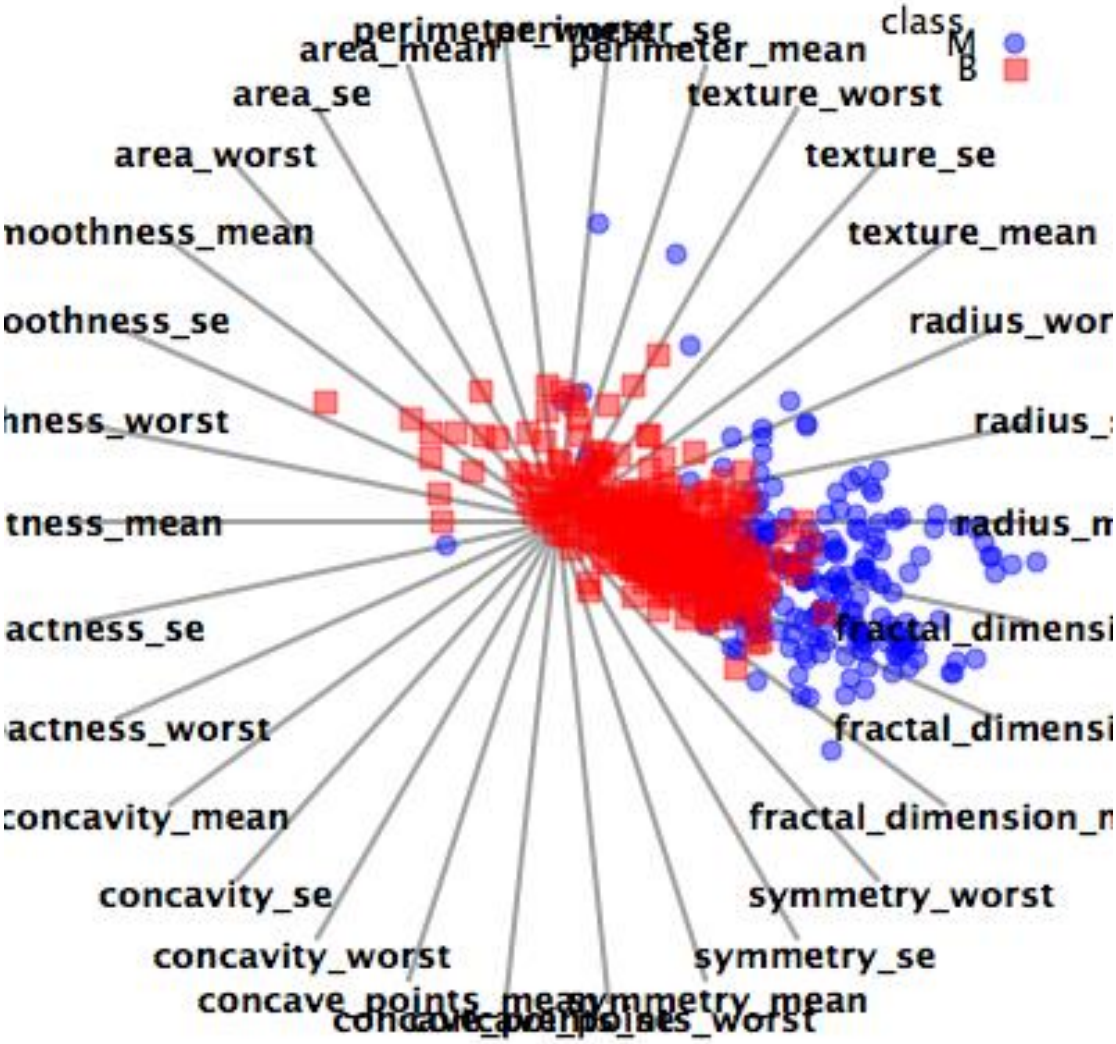

Figure 31 Linear projection of Breast Cancer Wisconsin (Diagnostic)

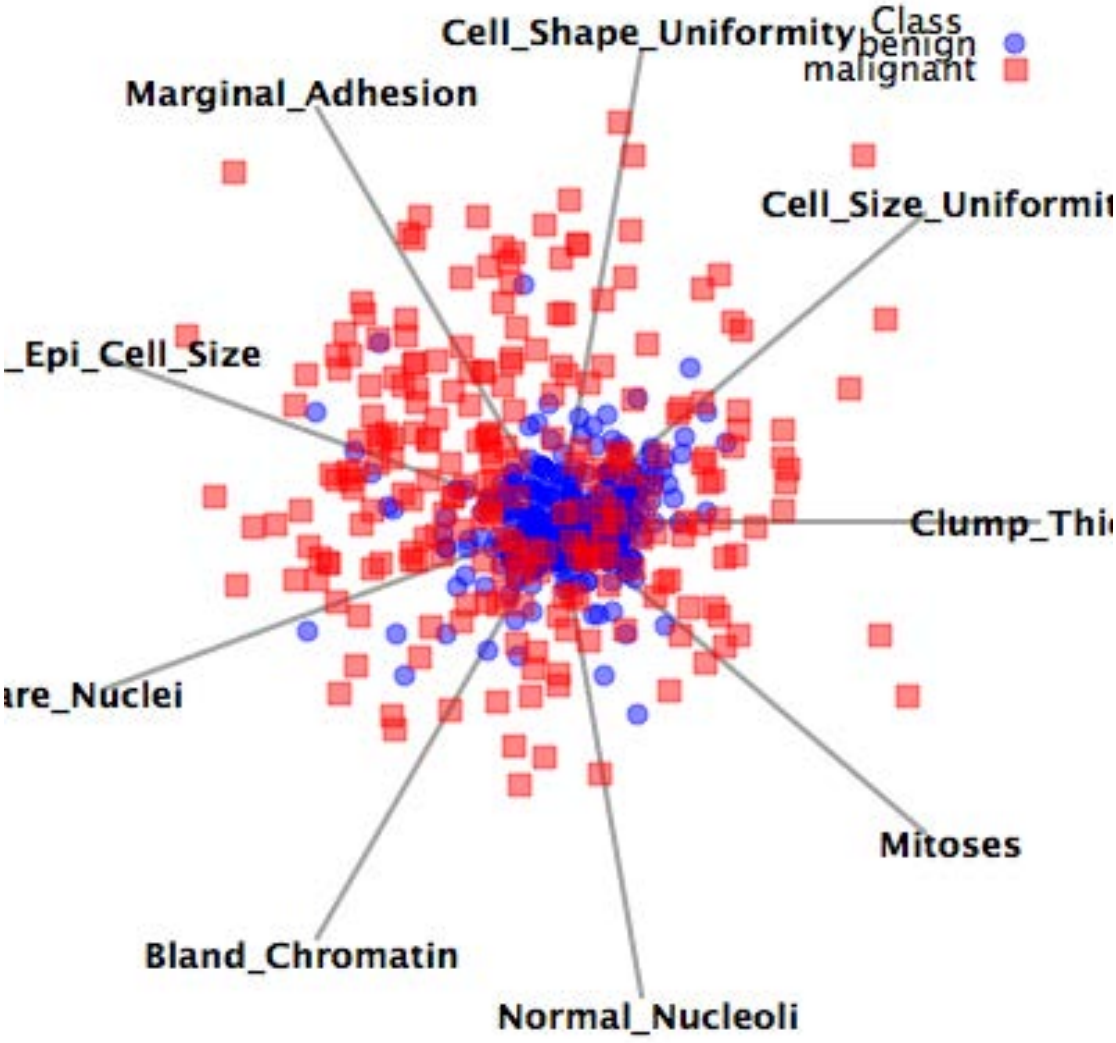

Figure 32 Linear projection of Breast Cancer Wisconsin (Original)

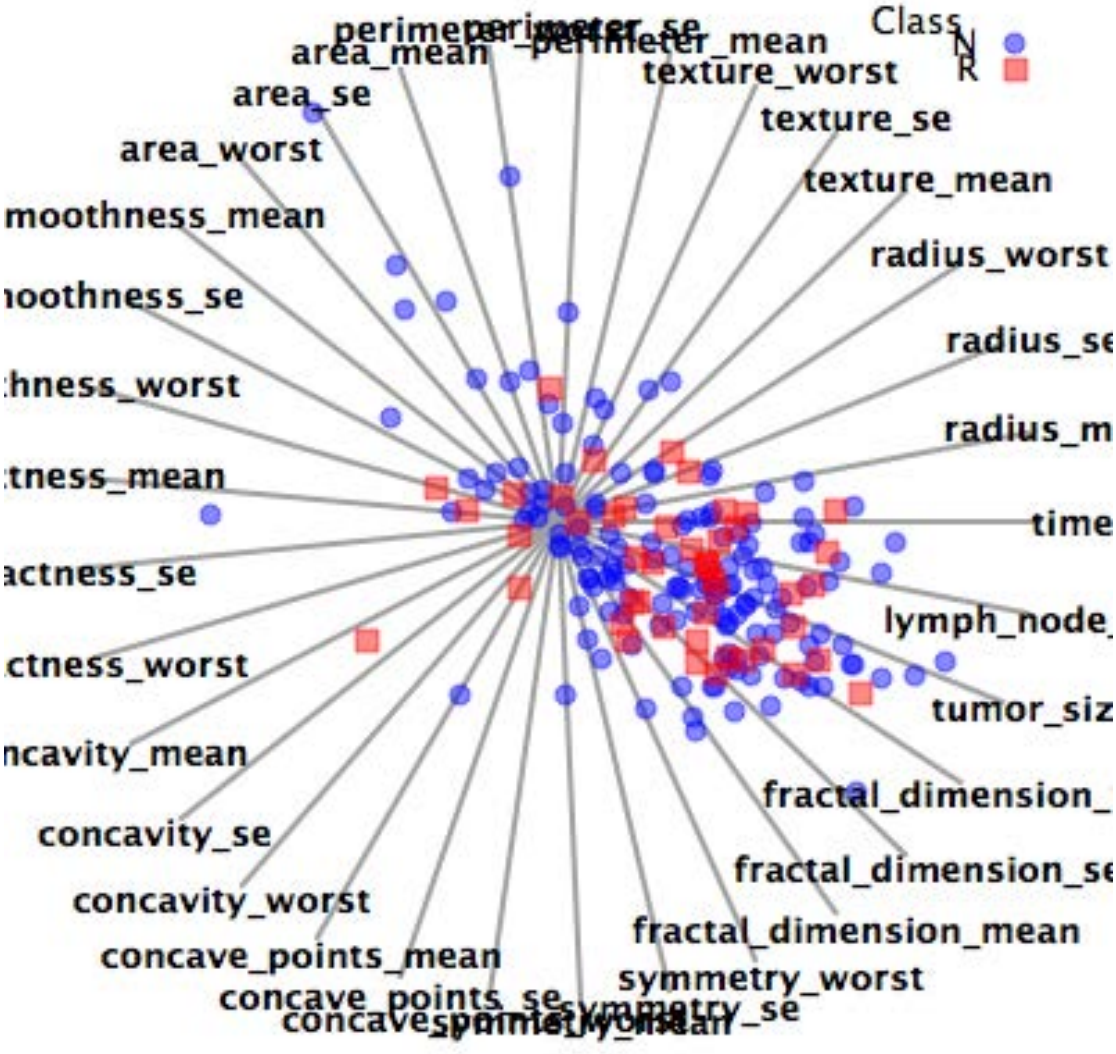

Figure 33 Linear projection of Breast Cancer Wisconsin (Prognostic)

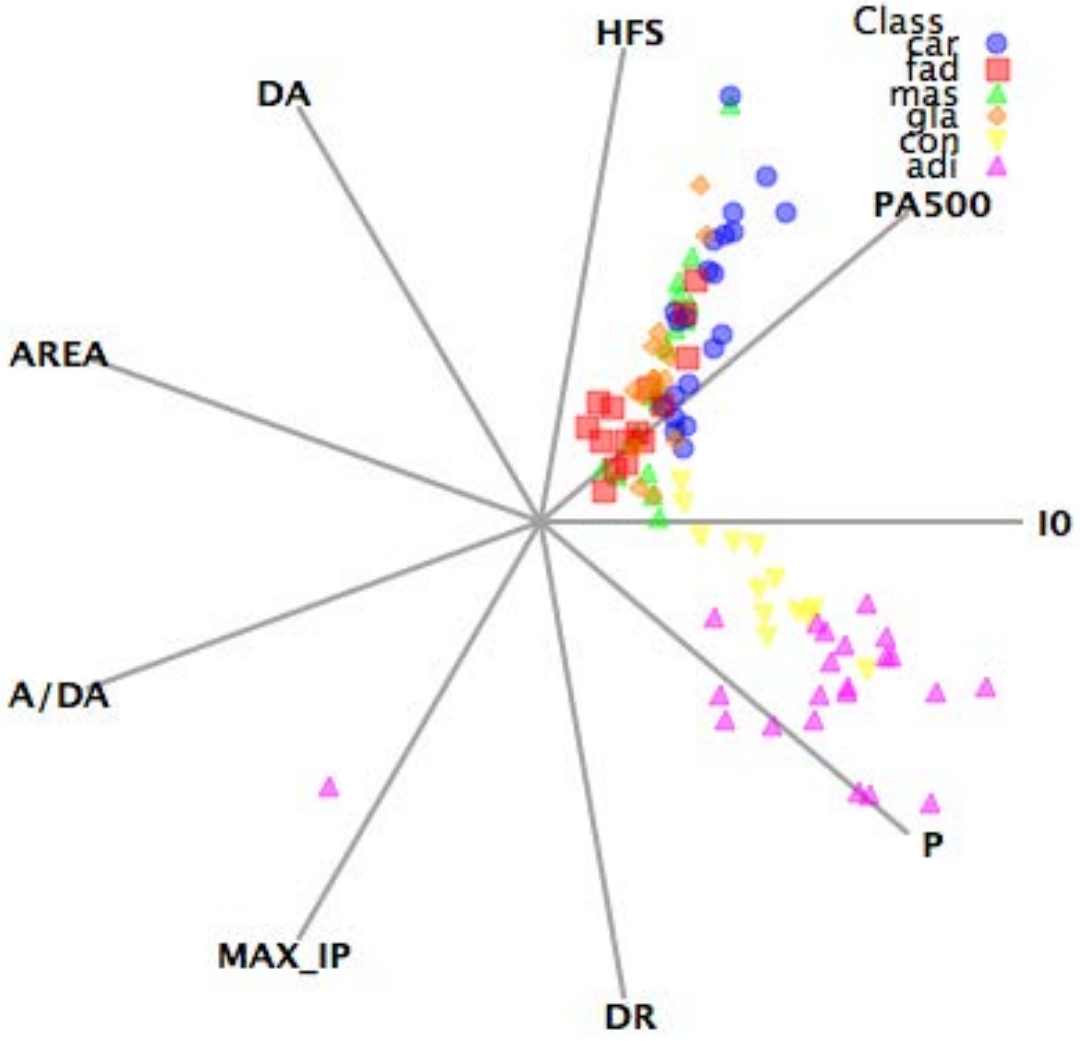

Figure 34 Linear projection of Breast Tissue

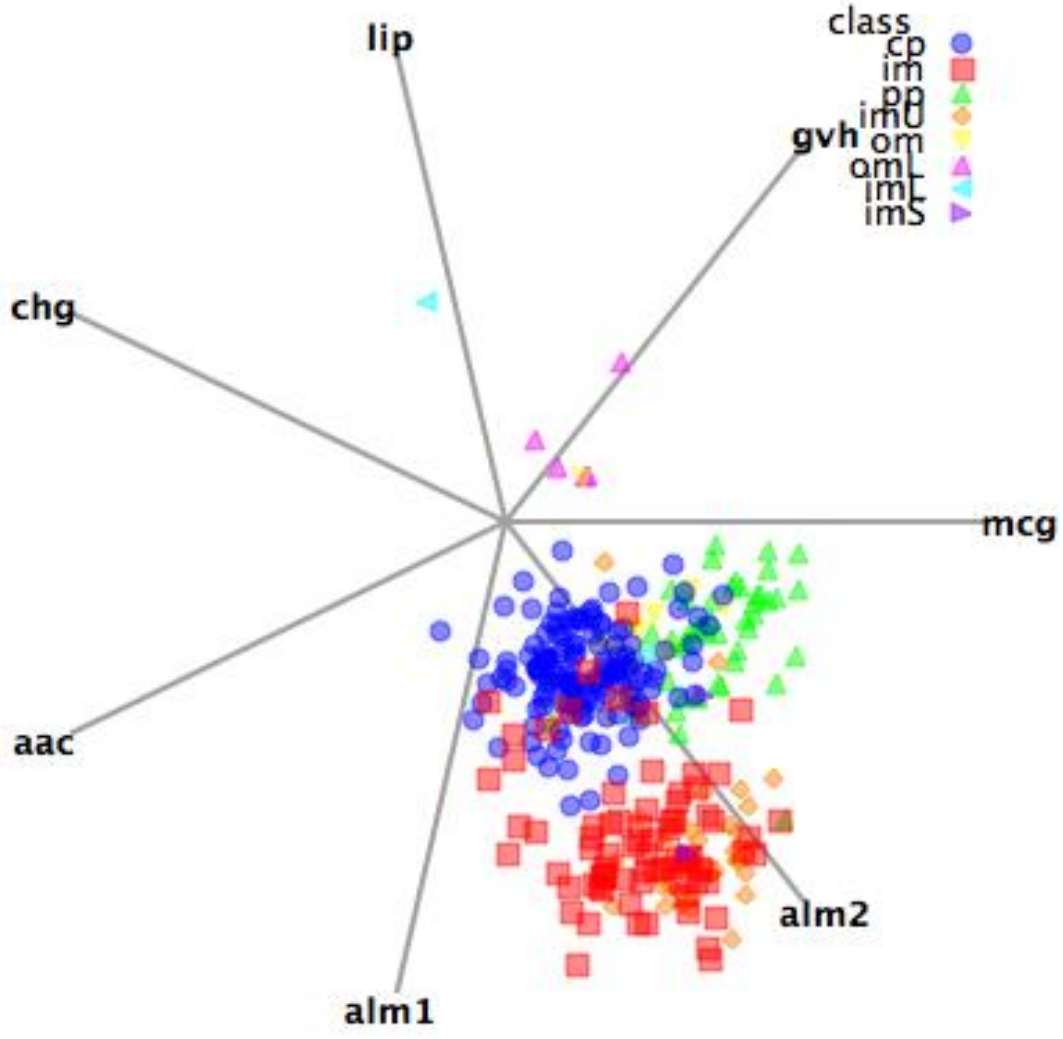

Figure 35 Linear projection of Ecoli

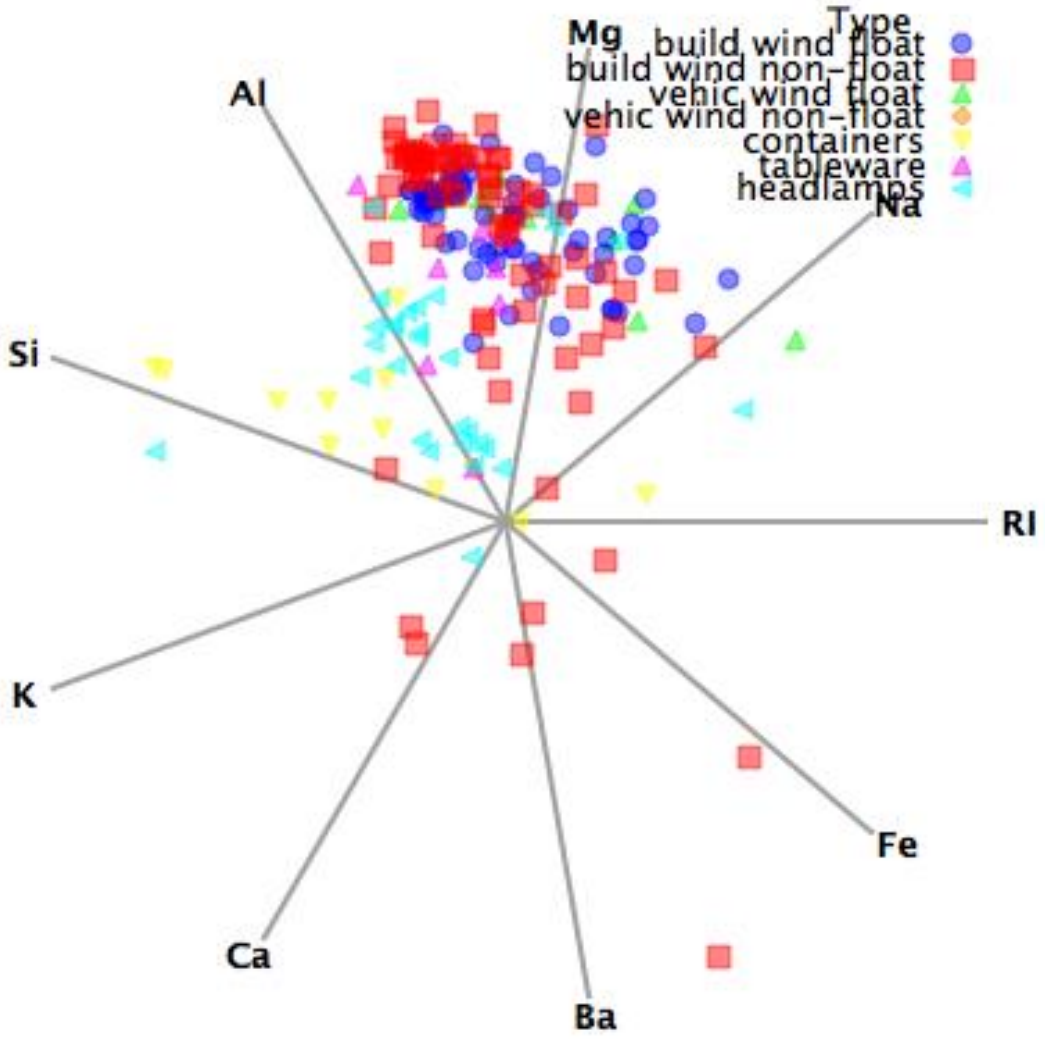

Figure 36 Linear projection of Glass Identification

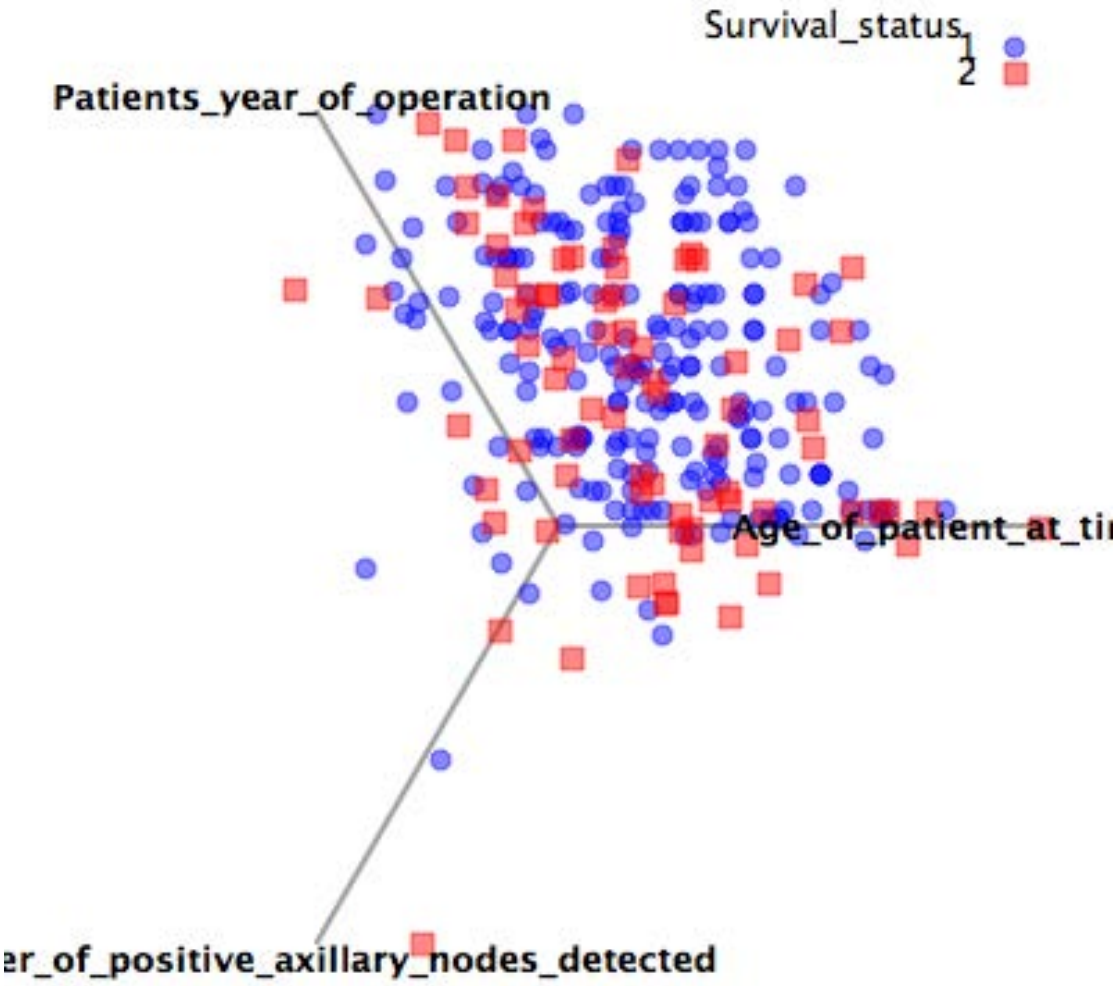

Figure 37 Linear projection of Haberman's Survival

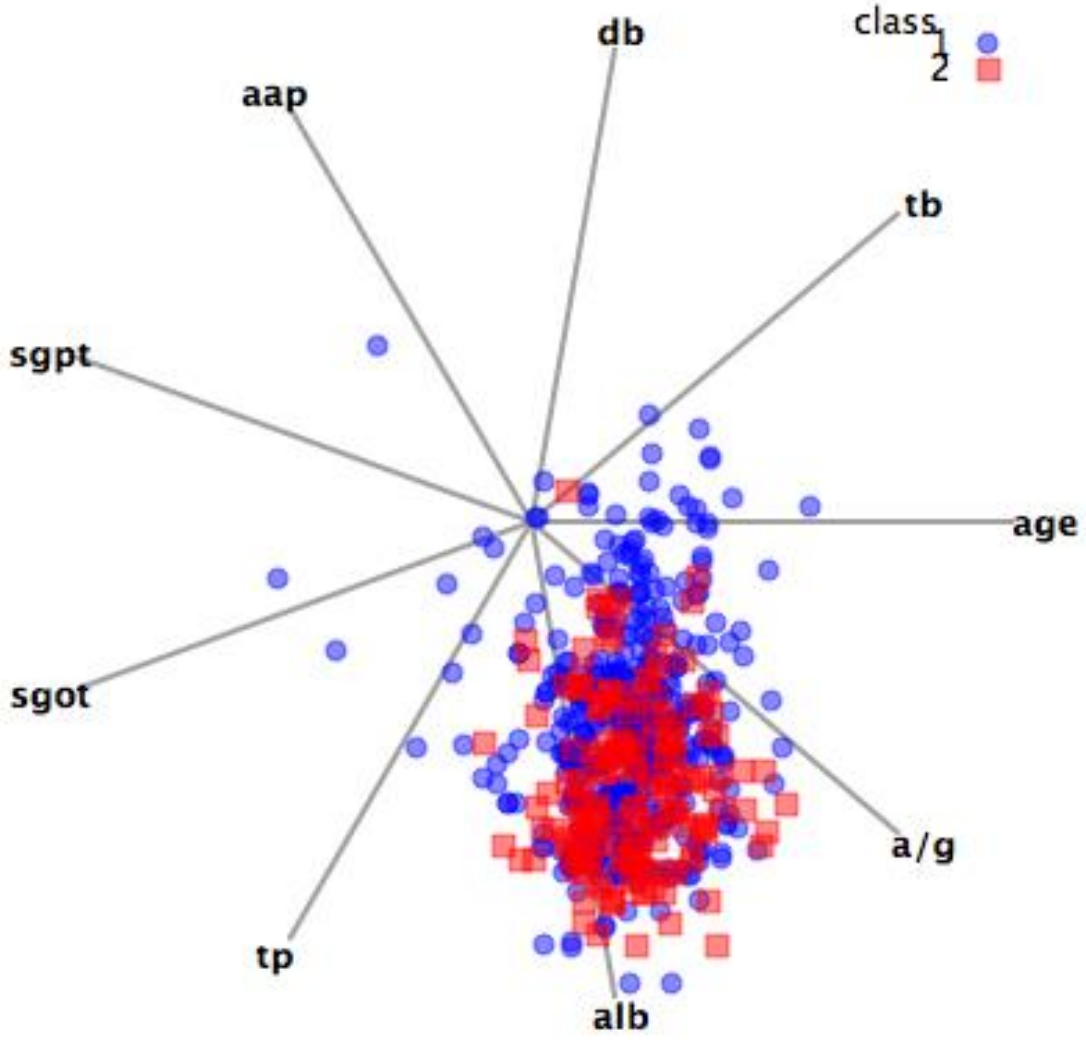

Figure 38 Linear projection of ILPD (Indian Liver Patient Dataset)

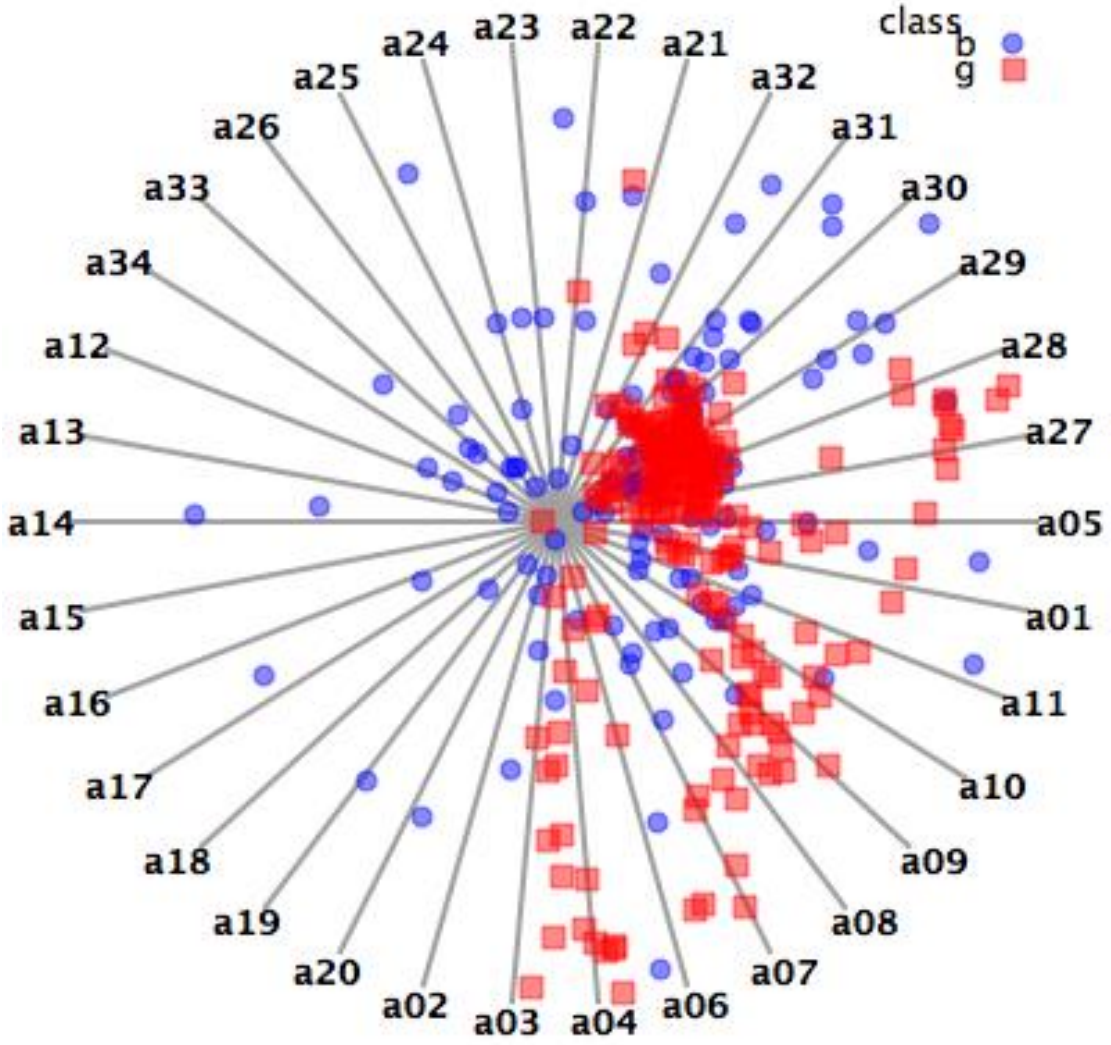

Figure 39 Linear projection of Ionosphere

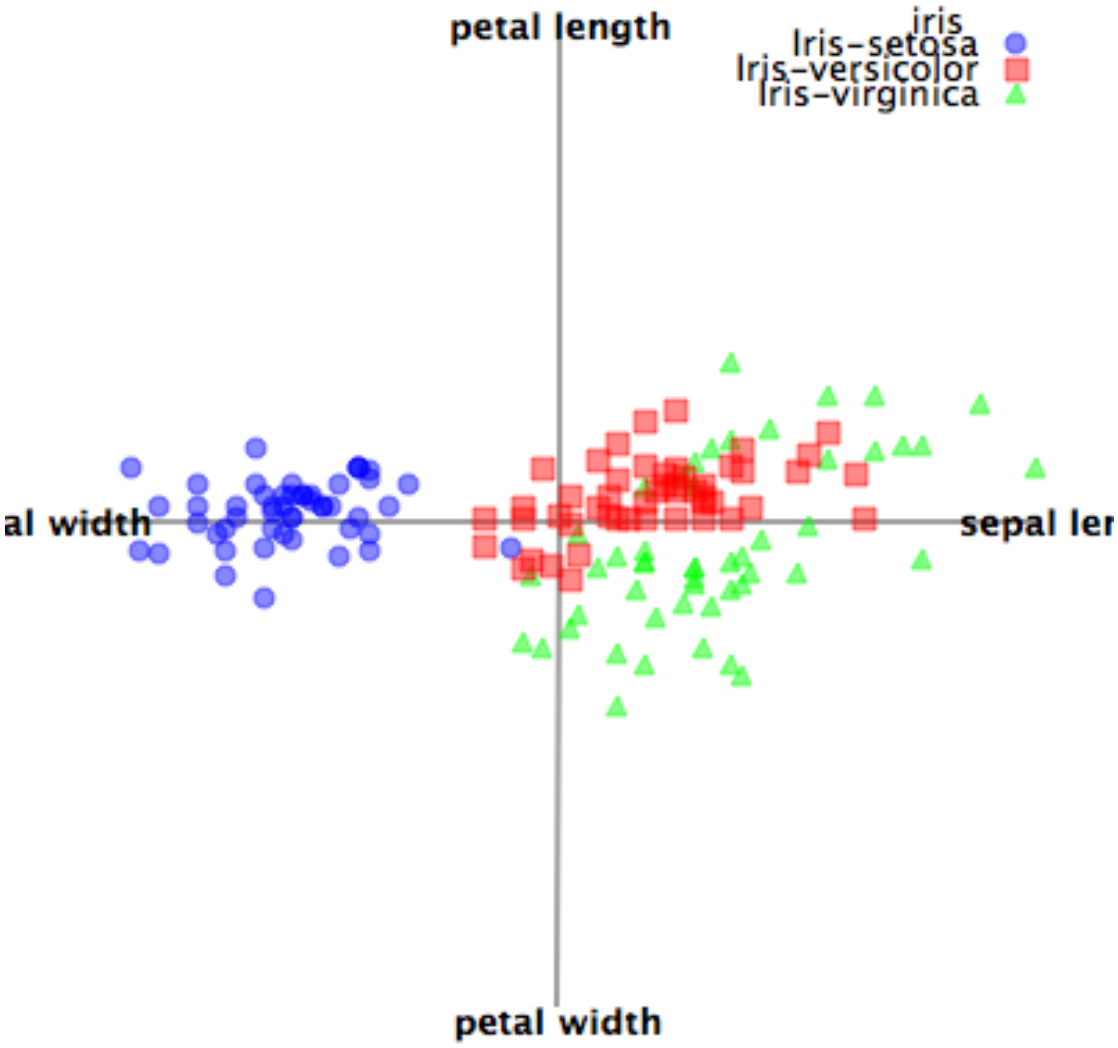

Figure 40 Linear projection of Iris

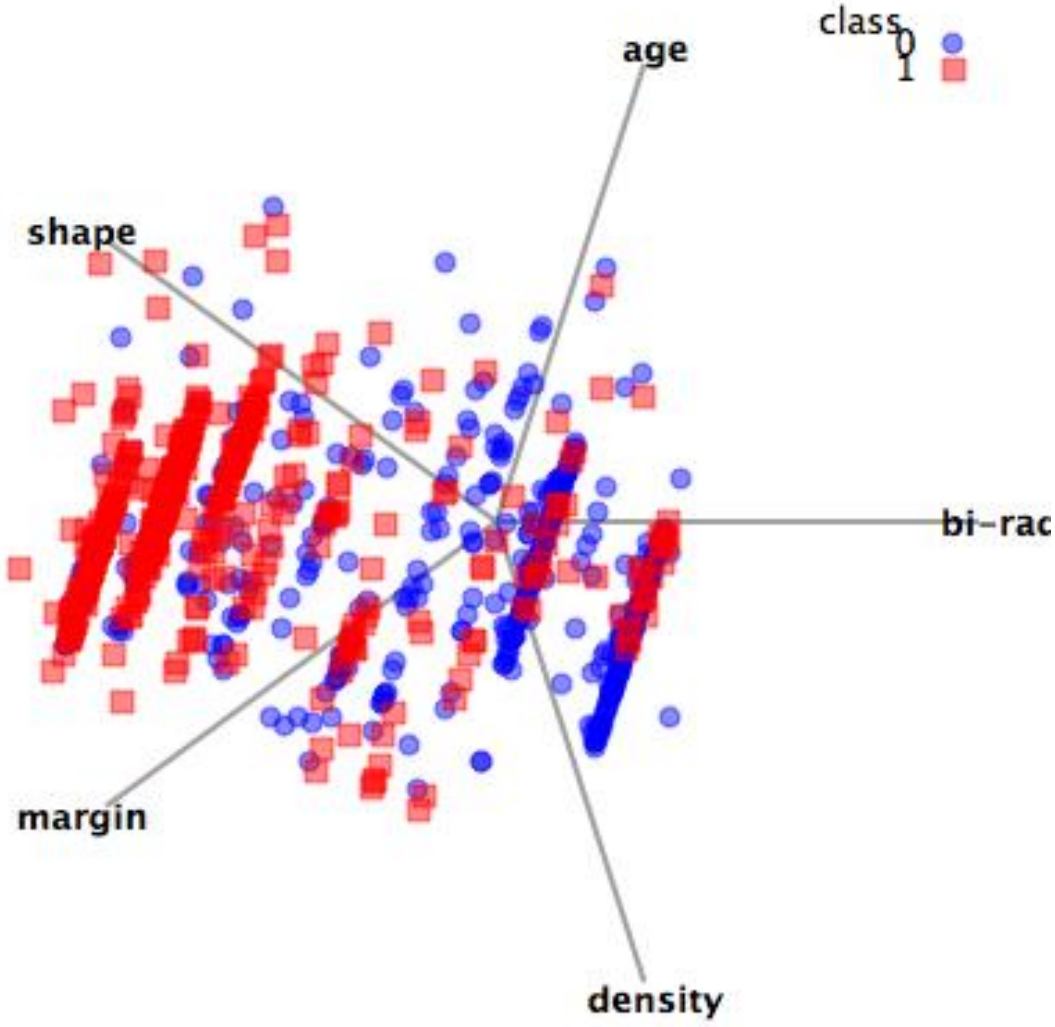

Figure 41 Linear projection of Mammographic Mass

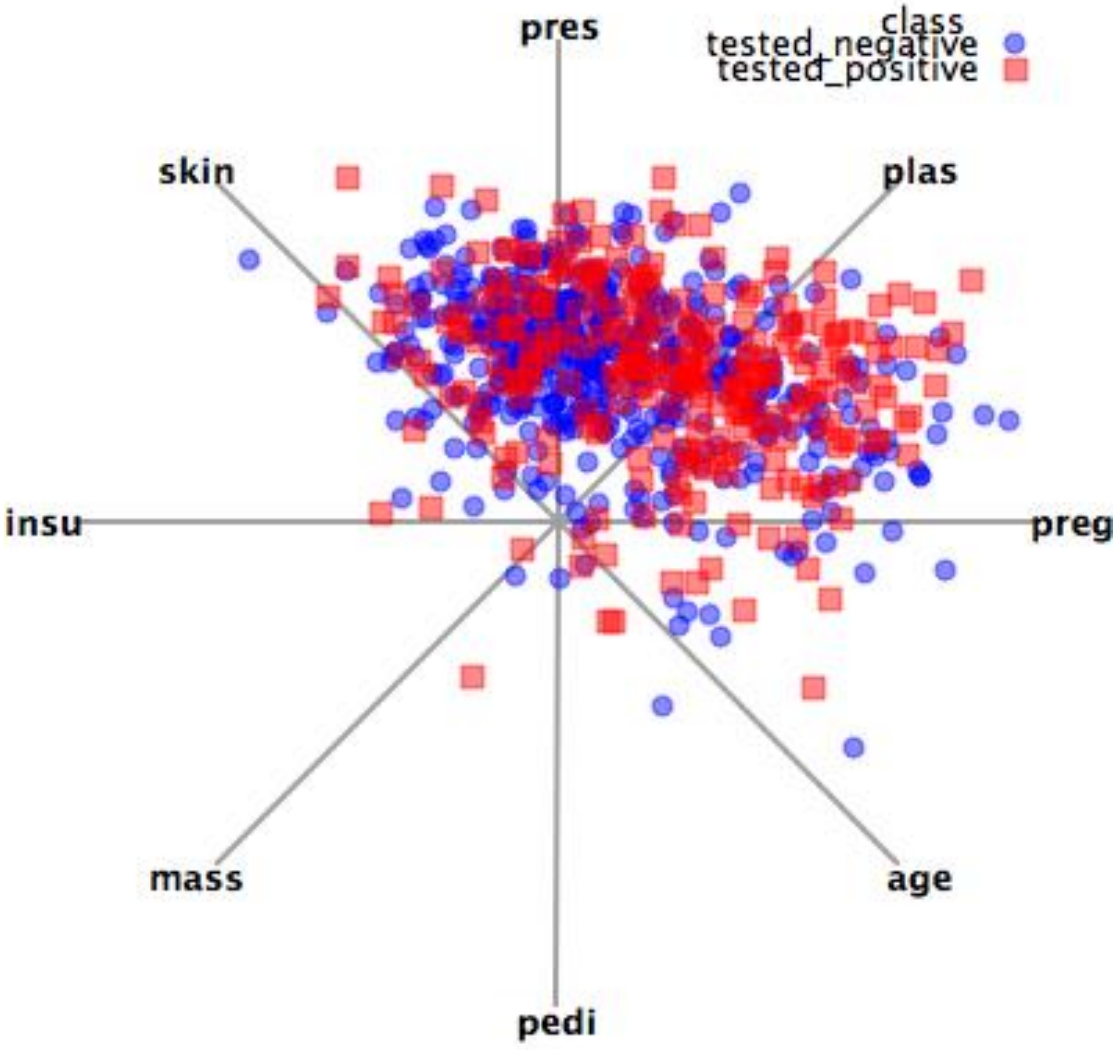

Figure 42 Linear projection of Pima Indians Diabetes

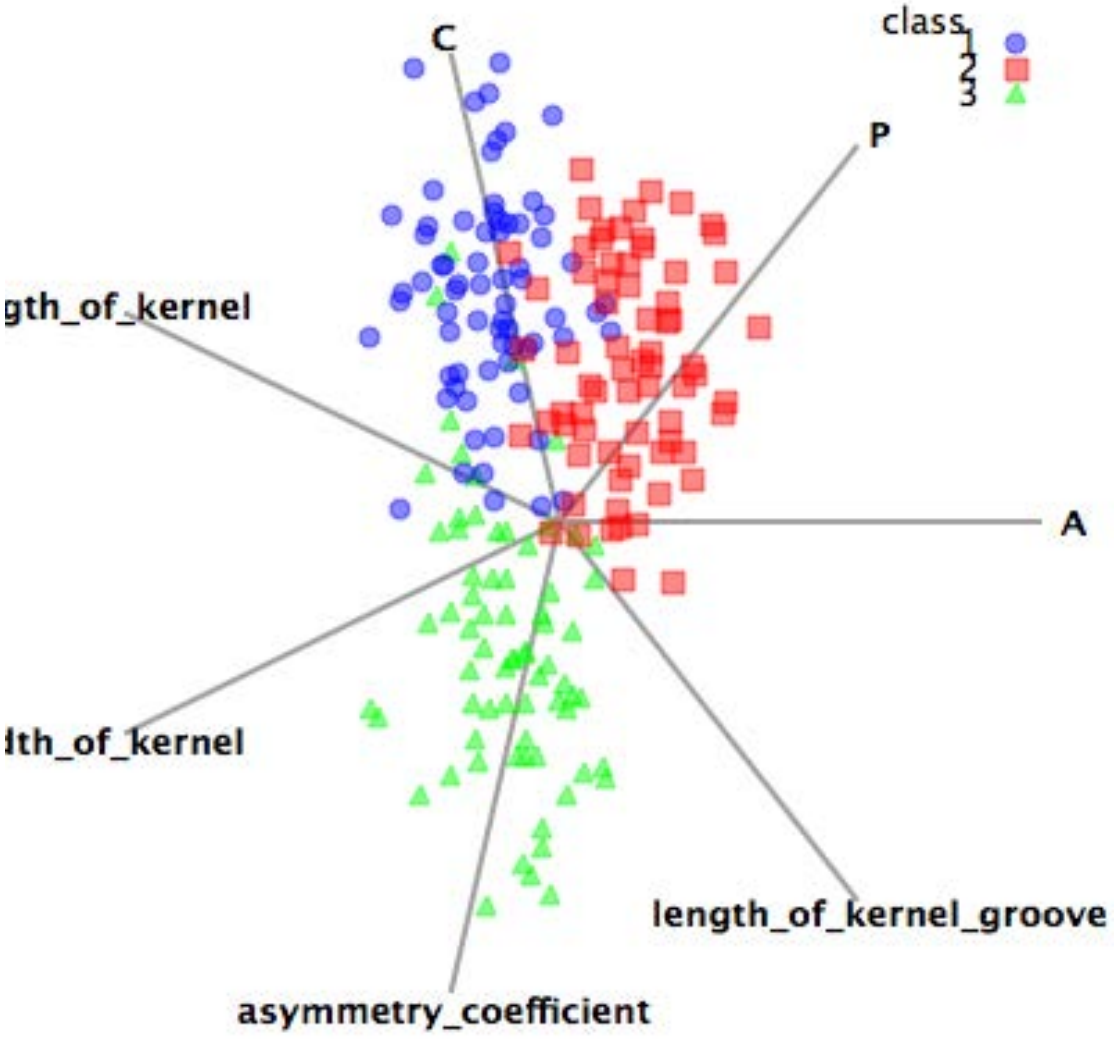

Figure 43 Linear projection of Seeds

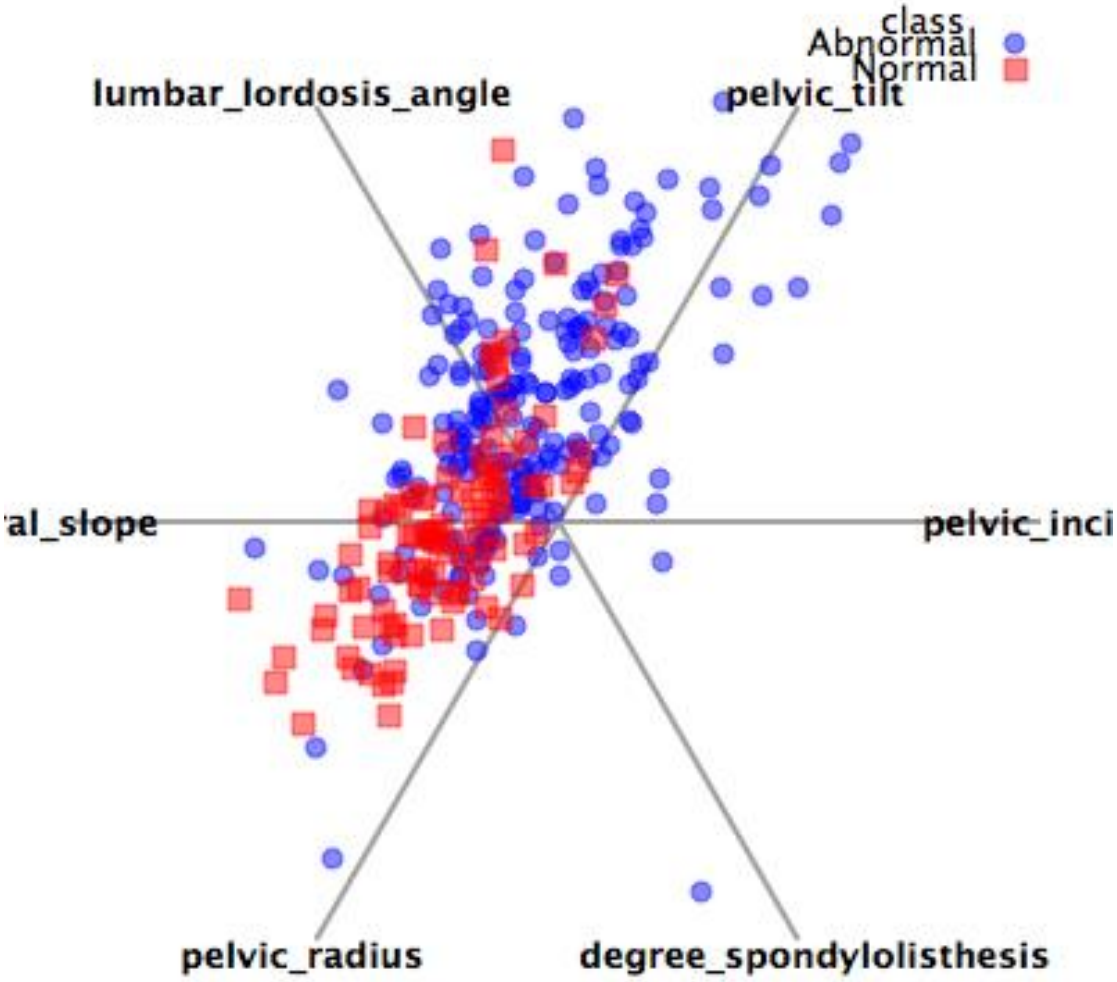

Figure 44 Linear projection of Vertebral Column

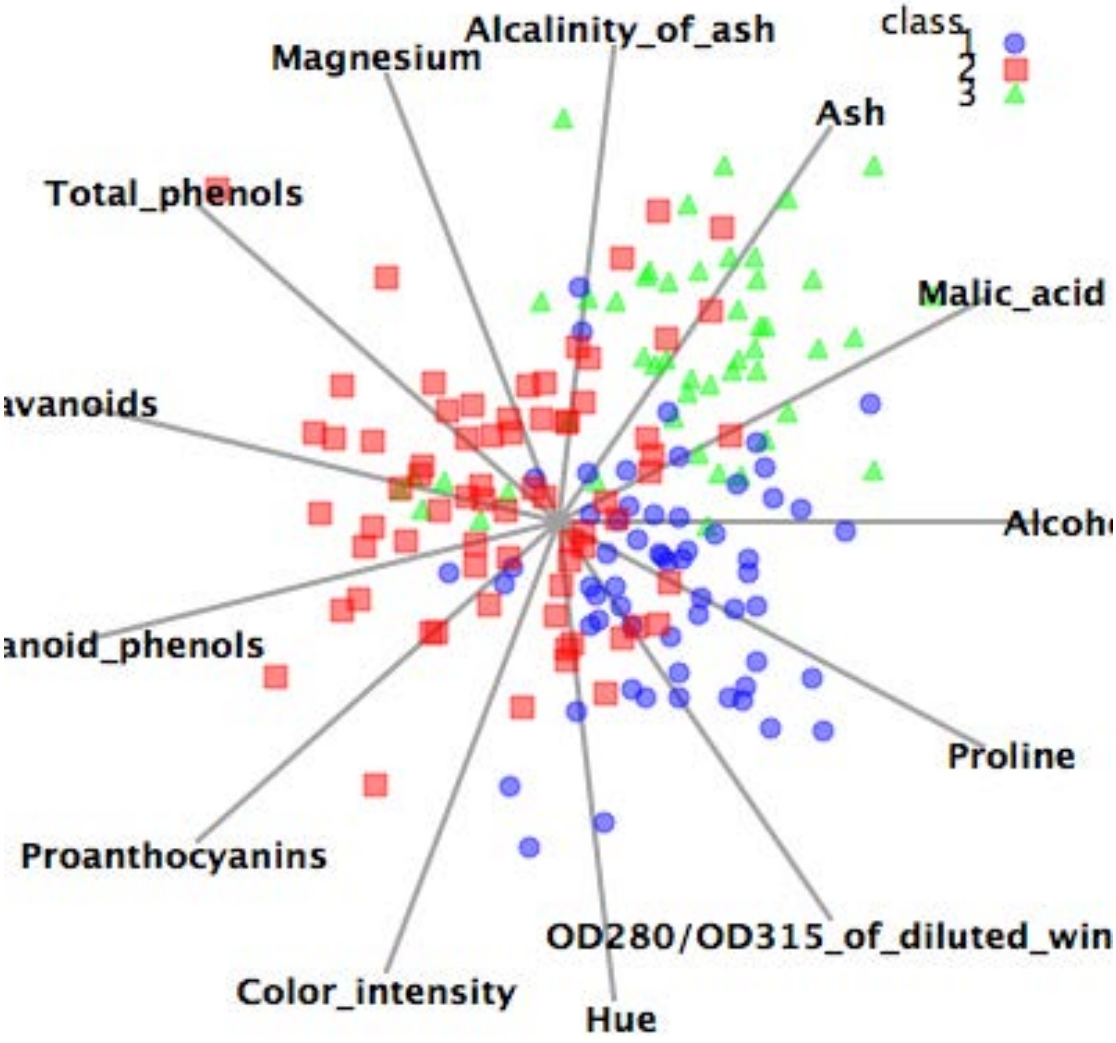

Figure 45 Linear projection of Wine

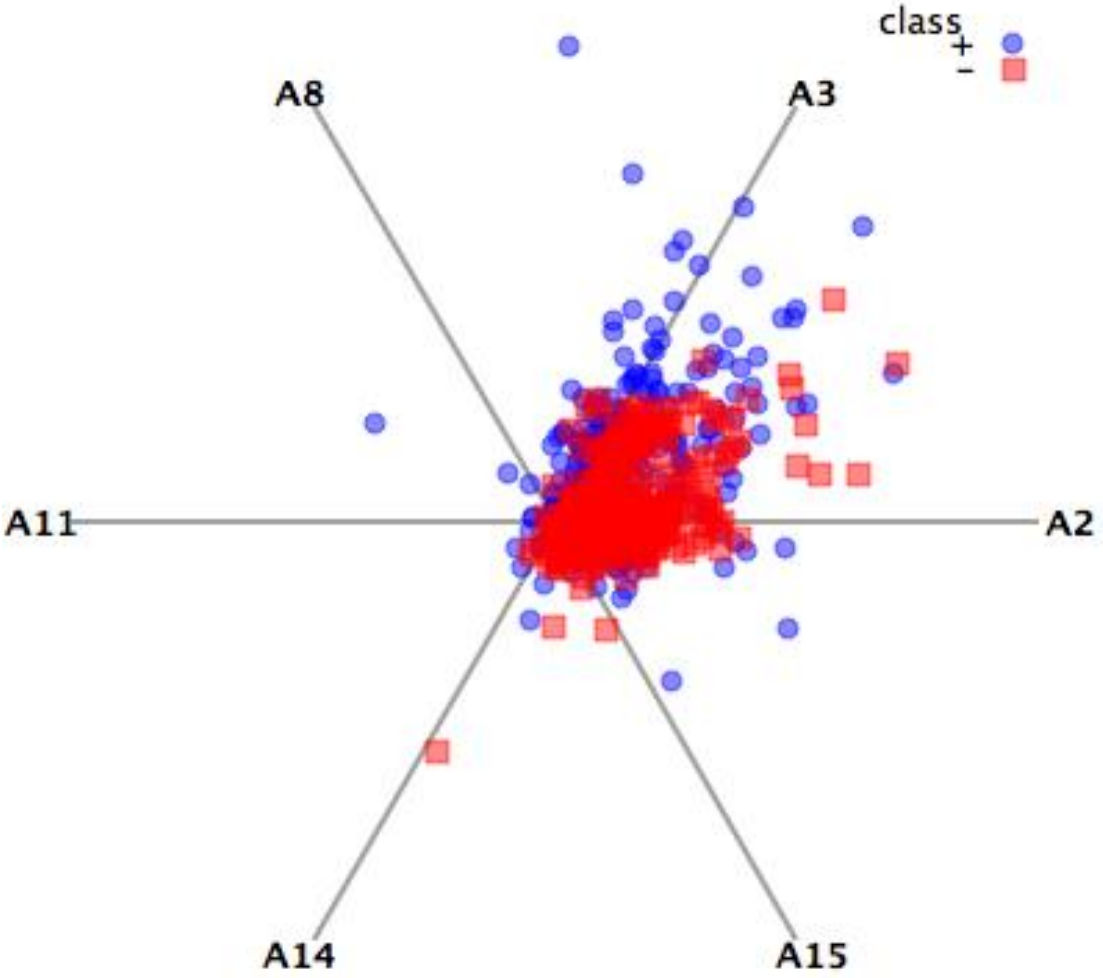

Figure 46 Linear projection of Credit Approval

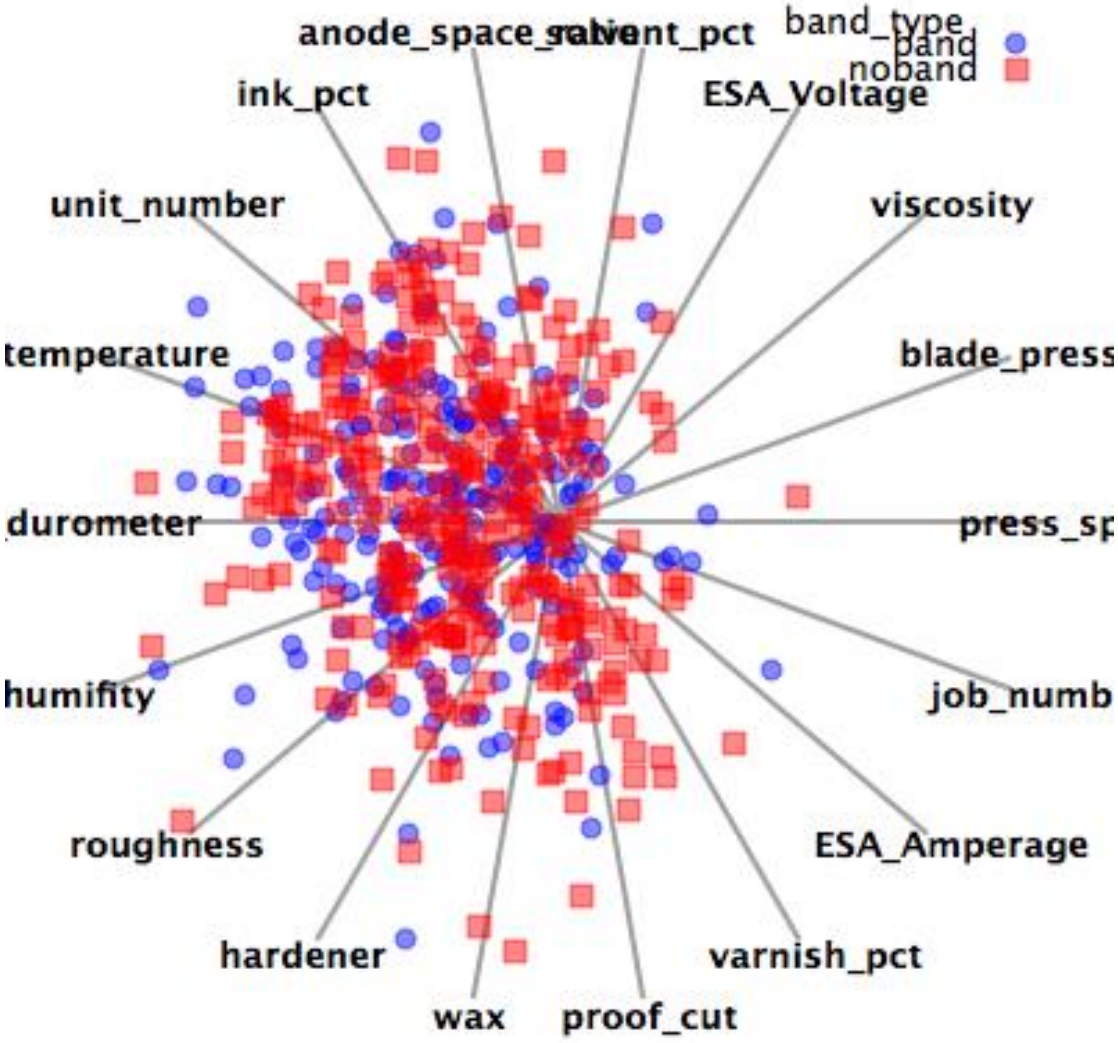

Figure 47 Linear projection of Cylinder Bands

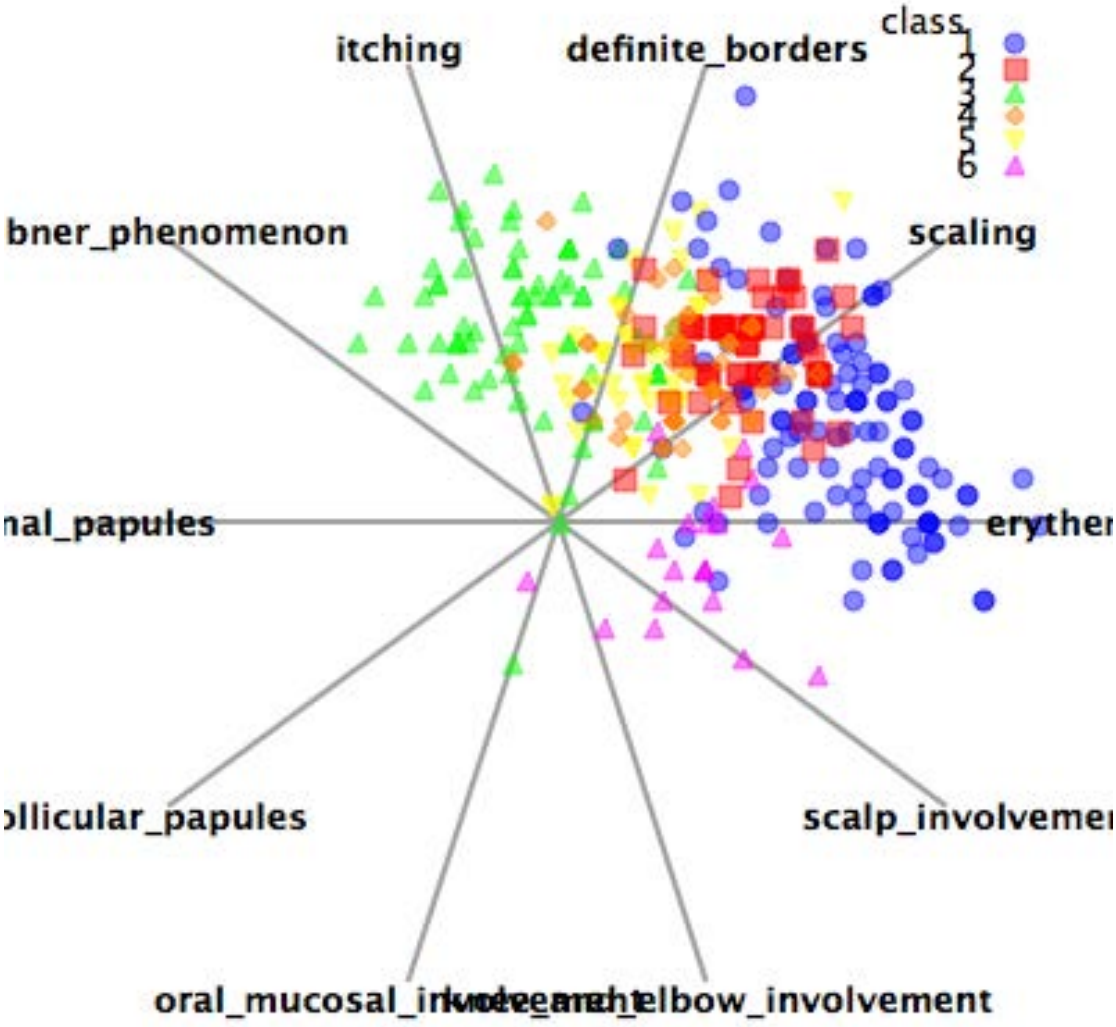

Figure 48 Linear projection of Dermatology

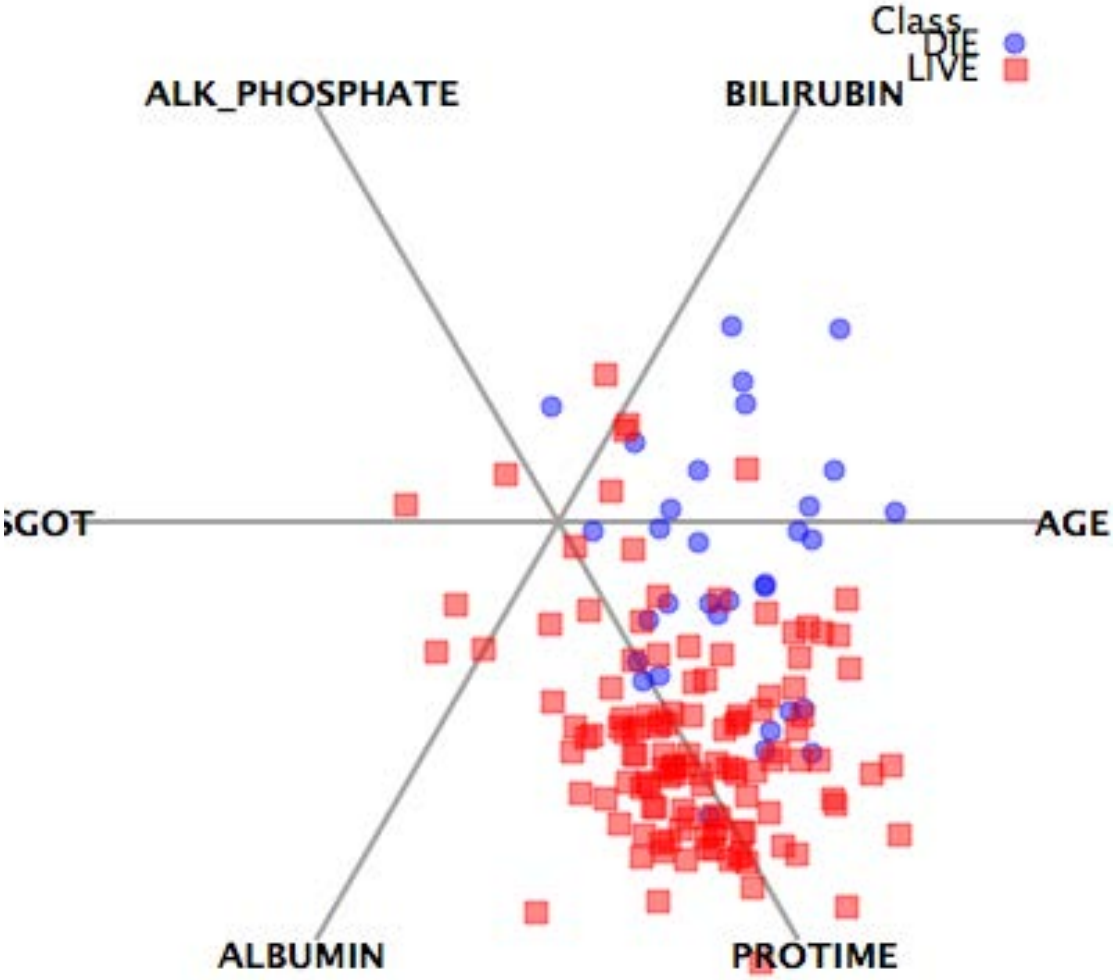

Figure 49 Linear projection of Hepatitis

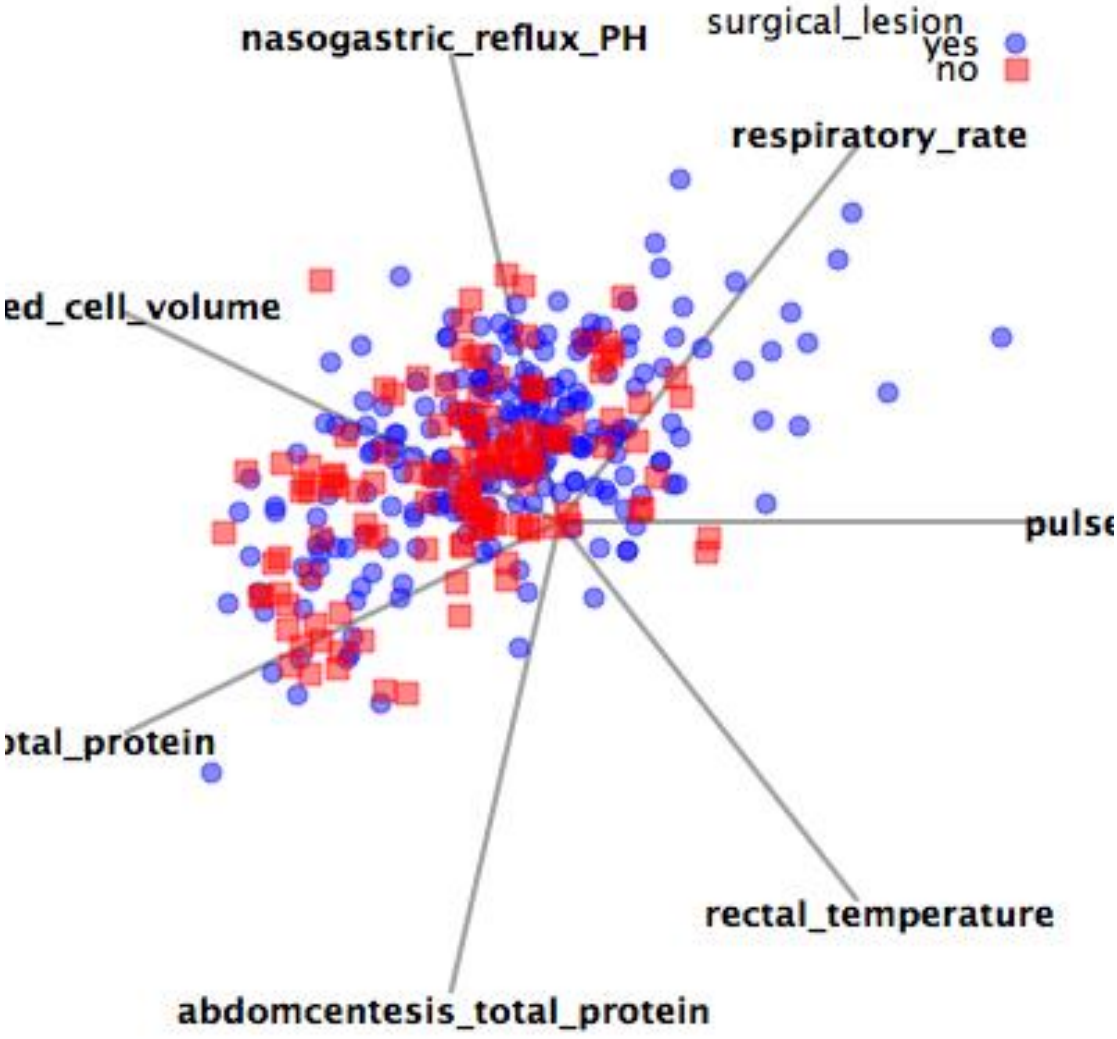

Figure 50 Linear projection of Horse Colic

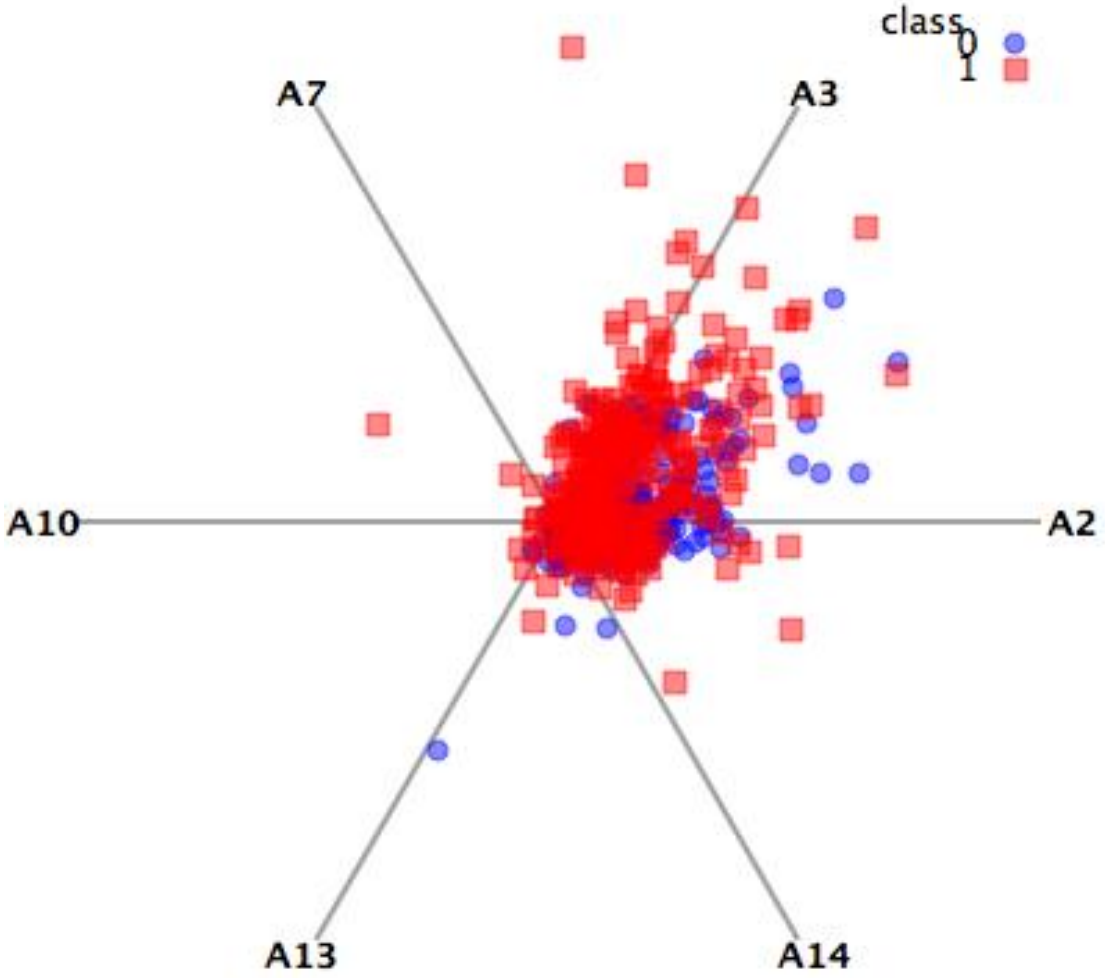

Figure 51 Linear projection of Statlog (Australian Credit Approval)

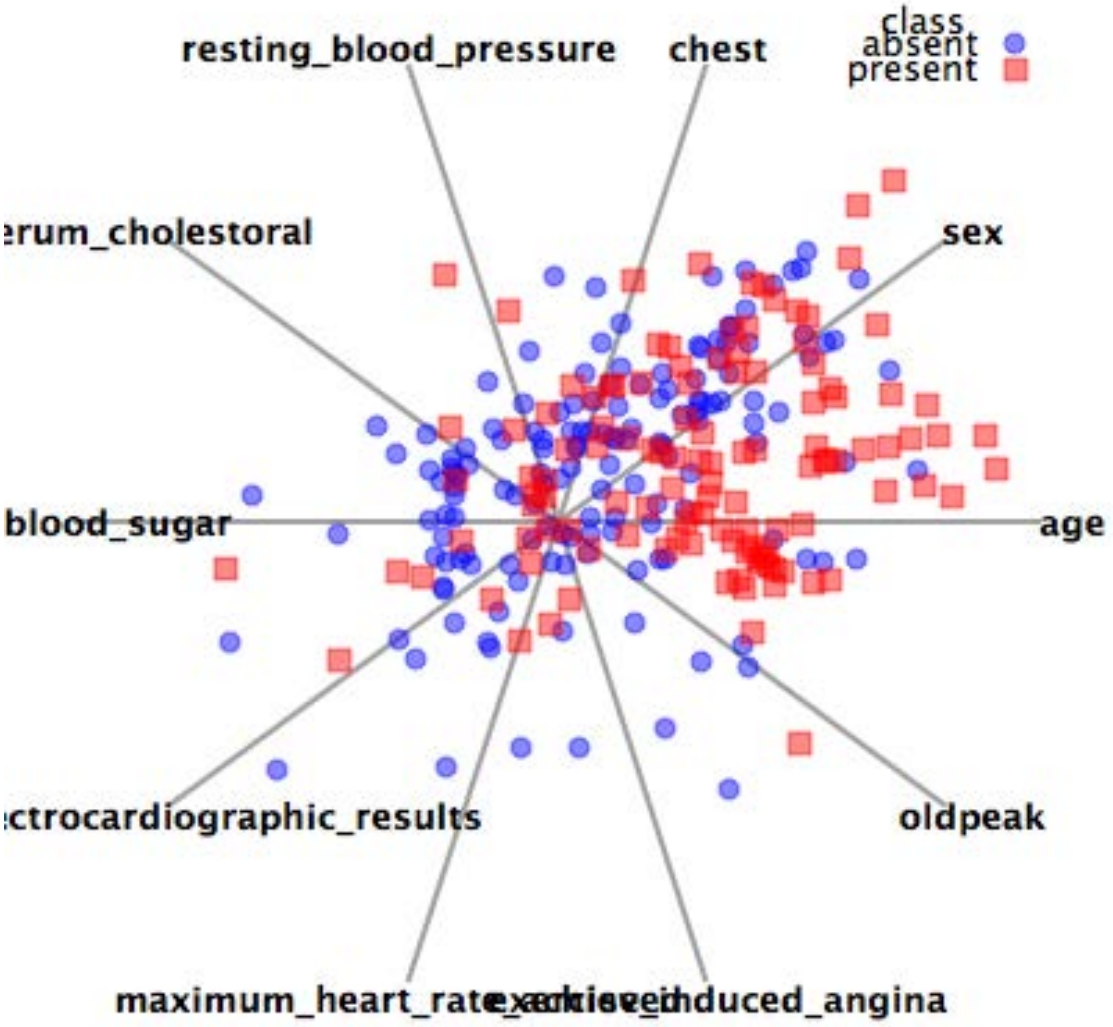

Figure 52 Linear projection of Statlog (Heart)

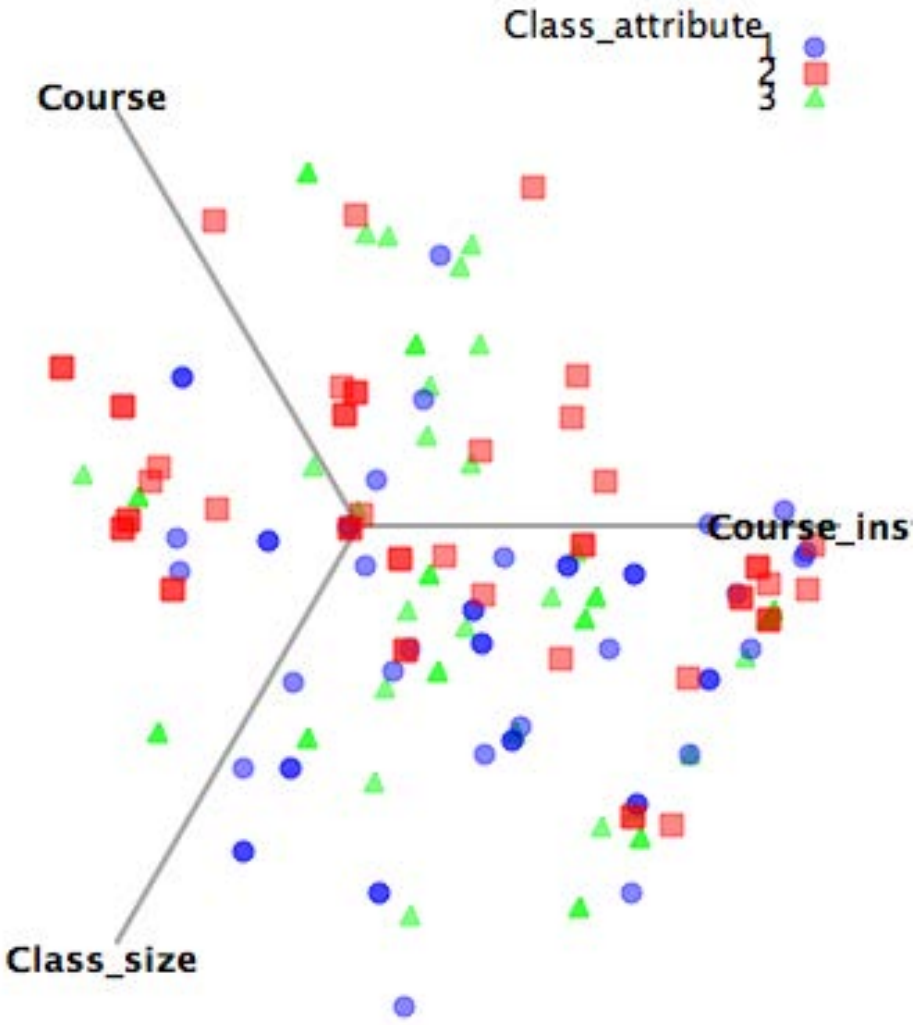

Figure 53 Linear projection of Teaching Assistant Evaluation

VITA

Name	Charoenchai Sirisomboonrat
Date of Birth	07 July 1987
Place of Birth	Rayong, Thailand
Education	B.Sc. (Computer Science), Chulalongkorn University
Publication	Sirisomboonrat, Charoenchai, and Krung Sinapiromsaran. "Breast cancer diagnosis using multi-attributed lens recursive partitioning algorithm." In *ICT and Knowledge Engineering (ICT & Knowledge Engineering), 2012 10th International Conference on*, pp. 40-45. IEEE, 2012.