

---

# Practical Work 1: Rule-based classifier

## Supervised and Experiential Learning

---

GUILLERMO CREUS BOTELLA

MSC IN ARTIFICIAL INTELLIGENCE



**UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH**

BARCELONA, APRIL 12, 2022

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Description . . . . .	1
1.2	Rule-based classifiers . . . . .	1
1.3	Metrics . . . . .	1
<b>2</b>	<b>Methods</b>	<b>2</b>
2.1	Input . . . . .	2
2.2	Training algorithm . . . . .	2
2.3	Prediction . . . . .	3
<b>3</b>	<b>Results</b>	<b>5</b>
3.1	Wines dataset . . . . .	5
3.2	Breast cancer Wisconsin dataset . . . . .	7
3.3	Seismic bumps dataset . . . . .	8
<b>4</b>	<b>Conclusion</b>	<b>10</b>
<b>5</b>	<b>Appendix</b>	<b>13</b>
5.1	Wines dataset . . . . .	13
5.2	Breast cancer Wisconsin dataset . . . . .	13
5.3	Seismic bumps dataset . . . . .	15

# 1 Introduction

## 1.1 Description

The purpose of this report is to gain insights on rule based classifiers. In particular, this report will be centered on the implementation of PRISM [1]. Based on the work of Jadzia Cendrowska [1], the PRISM algorithm will be developed and evaluated on three datasets from the UC Irvine Machine Learning Repository [2].

The motivation of this work is double sided. On one hand, academically, it is highly stimulating to implement from scratch an algorithm and have full control over it, where the opportunities to learn are endless. In addition, the mistakes on the way will be key at mastering this algorithm, contrary to other ready-to-use solutions.

On the other hand, by developing this algorithm one has the opportunity to test its performance on datasets and analyze where it fails, so as to improve it in the future.

As a side note, PRISM will be evaluated on three datasets of different size in order to determine how it scales. It is relevant to check if rule-based classifiers' performance is independent of the size and dataset used.

## 1.2 Rule-based classifiers

Whenever patterns are present in a dataset, a rule-based classifier aims at learning the rules that are followed by the instances. This classifier will be fed a training dataset, comprised of several instances which will have values in the attributes domain:  $\{\alpha_1, \dots, \alpha_{n_a}\}$ . The possible set of values for attribute  $i$  will be:  $\{z_{i,0}, \dots, z_{i,n_i}\}$ . The target value will be denoted as  $\delta_k$ , where  $k \in \{1, \dots, N_c\}$ . Having said that, the goal of the rule-based classifier will be to output several expressions like the following:

$$R := A_1 \wedge \dots \wedge A_s \rightarrow \delta_k \quad (1)$$

where,

$$A_m = (\alpha_i == z_{i,l}) \quad (2)$$

$A_m$  will be true when attribute  $i$  ( $\alpha_i$ ) matches value  $l$  ( $z_{i,l}$ ). Consequently, rule  $R$  will be fired when all  $A_m$  are true. That being said, it is relevant to introduce the coverage and precision of rules:

- $\text{Coverage}_R = \# \text{ instances that fire rule } R$
- $\text{Precision}_R = \frac{\# \text{ Correctly classified instances}}{\text{Coverage}_R}$

## 1.3 Metrics

Since rule-based classifiers can handle multiclass problems, i.e., ones where there are more than two target values, it is important to explain the metrics [3] that will be used. First of all, accuracy is defined as:

$$\text{Accuracy} = \frac{\sum_{k=1}^{N_c} TP_k}{\# \text{ of instances}}$$

Where  $TP_k$  is the number of correctly classified instances of class  $\delta_k$  and  $FP_k$  is the number of instances that were wrongly predicted as class  $\delta_k$ .

For every class  $\delta_k$ , one can define the

$$\begin{aligned} \text{Precision}_k &= \frac{TP_k}{TP_k + FP_k} \\ \text{Recall}_k &= \frac{TP_k}{\# \text{ of instances of class } \delta_k} \\ F1_k &= \frac{2 * \text{Precision}_k * \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \end{aligned}$$

When one reports these metrics, readily available in [7], it is important to announce whether they are **weighted** by the amount of instances belonging to each class, or if a regular **macro average** is being done, which does not take into account the amount of instances in each class. This is extremely important in unbalanced datasets, since an extremely low metric in the minimal class can drop the overall metric significantly.

## 2 Methods

As mentioned in the previous section, the rule-based classifier that will be explored in this work will be PRISM. In the following sections, the pseudo-code of the PRISM algorithm will be shown. In addition, the expected input will be displayed and it will be shown how the predictions are computed and some disambiguation techniques.

### 2.1 Input

In the first place, it is important to point out that PRISM [1] was developed as a rule-based classifier of datasets where the attributes were only categorical. Since one of the secondary goals of this work is to extend the functionality of the original algorithm, the datasets will be preprocessed beforehand so all the attributes are categorical.

The strategy followed is to discretize variables using *KBinsDiscretizer* of *sklearn* [7]. Its input will be a real or integer valued array, and it will output an array with  $n_{bins}$  different values. For example, if one selects  $n_{bins} = 4$ , the discretizer will sort data and assign the category based on its quartile. Therefore, first quartile instances will be assigned a “0”, and so on.

From here on, the number of bins will be set to 3. Although this number looks quite low, it is the sweet spot between interpretability and complexity. Setting it to 3 allows one to transform the input to three levels: “low”, “medium” and “high”, which will be encoded as “L”, “M” and “H”.

Additionally, in order to make the algorithm even more robust, the missing values will be imputed. In the case of missing values of categorical attributes, the mode will be imputed, and in the case of continuous attributes, the mean will be imputed.

### 2.2 Training algorithm

In algorithm 1, the pseudo-code of PRISM can be seen. To understand it better it is relevant to introduce  $p(x, y)(\delta_k | \alpha_z) \approx \frac{\sum_{j=1}^{N_{samples}} \mathbb{1}_{y_j = \delta_k}}{\sum_{j=1}^{N_{samples}} \mathbb{1}_{x_j[\alpha] = z \wedge y_j = \delta_k}}$ , which the probability that a sample’s target value is equal to  $\delta_k$  given that its attribute  $\alpha$  is equal to  $z$ . Here,  $x_j[\alpha]$  is the attribute  $\alpha$  of the  $j^{th}$  sample, and  $y_j$  is the label of the  $j^{th}$  sample.

Two aspects that must be highlighted when computing  $p(x, y)(\delta_k | \alpha_z)$  are the following:

- $p(x, y)(\delta_k | \alpha_z)$  can not be computed whenever there are no instances of label  $\delta_k$  and attribute  $\alpha$  equal to  $z$  in  $(x, y)$ .
- $p(x, y)(\delta_k | \alpha_z)$  is highly dependent on the dataset  $(x, y)$

As stated in [6], there are two strategies when growing rules: general-to-specific and specific-to-general. PRISM uses the first one, implying that every rule starts as empty. Going back to the notation presented in equations 1, 2, given a class  $\delta_k$ ,  $A_m$  will be induced as the combination  $(\alpha, z)$  that has the highest probability of having label  $\delta_k$  conditioned to attribute  $\alpha$  being equal to  $z$ . In the case that two combinations had the same probability, the one that induced a rule with higher coverage would be selected.

This process continues until one of two situations arise:

- Whenever the rule is fired only by instances of label  $\delta_k$ , it will be added to the set of rules that will be returned.

- Whenever the combinations  $\alpha_z$  are exhausted, the rule will be marked as complete. That can happen when the discretization transforms two instances with similar continuous values to the same categorical values, but their labels remain different. In this case, the precision of this rule will not be 100% because some instances will fire it while having labels different to  $\delta_k$

Finally, the algorithm ends when there are no instances left in any of the classes  $\delta_k$ . That way, the returned rules will cover all of the training instances.

## 2.3 Prediction

In algorithm 2 one can see the implementation followed to predict a new instance. The steps needed to classify an instance are the following:

1. Gather all rules that are fired by the instance.
2. In case that one or more rules are fired, choose the one that had the highest coverage in the training set. In the event of two rules having the same coverage, choose the first one.
3. If no rule is fired, predict the mode of the training labels.

The second step is an heuristic extracted from [1], which attempts to decide between several rules by selecting the one with highest coverage.

The third step is an idea developed during this work which might work well in balanced datasets. However, it is relevant to point out that it might not be best in imbalanced ones. That is caused by the mere fact that the underrepresented class will never be predicted when considering this step.

Therefore, if a rare event comes that is not represented by the rule set, its label will be predicted as the over-represented class (mode). This will work well in the majority of cases but if one is trying to predict a critical event (hazardous situations, cancer, ...), it will never consider the possibility that its label might come from the underrepresented class.

---

**Algorithm 1** PRISM Algorithm

---

**Input:**  $X_{\text{trn}}, y_{\text{trn}}$  – Training features and labels**Output:**  $\text{all}_R$  – Set of  $\{R, \delta_k\}$  that describe the training dataset

```
1: Initialize  $S_{\text{not used}}$  as the set of all attribute-value combinations
2:  $X_{\text{trn}}$  will denote the features of the training set
3:  $y_{\text{trn}}$  will denote the labels of the training set
4: for each  $\delta_k$  in all target values do
5:    $X_{\text{remaining}}, y_{\text{remaining}} \leftarrow X_{\text{trn}}, y_{\text{trn}}$ 
6:    $X_{\text{rule}}, y_{\text{rule}} \leftarrow X_{\text{trn}}, y_{\text{trn}}$ 
7:   repeat
8:      $X_{\text{rule}}, y_{\text{rule}} \leftarrow X_{\text{remaining}}, y_{\text{remaining}}$ 
9:      $S_{\text{attributes}} \leftarrow S_{\text{not used}}$ 
10:     $R \leftarrow \emptyset$ 
11:    repeat
12:      // Creation of the rule
13:       $p_{\text{max}} \leftarrow -1$ 
14:       $n_{\text{max}} \leftarrow -1$ 
15:       $\hat{\alpha}_{z_0} \leftarrow \text{nothing}$ 
16:
17:      for each  $\alpha_z$  (i.e. all {attribute, value} combination)  $\in S_{\text{attributes}}$  do
18:        // Selecting {attribute, value} with highest  $p$ 
19:         $R_{\text{now}} = R \cup \alpha_z$ 
20:        if  $p_{(X_{\text{rule}}, y_{\text{rule}})}(\delta_k | \alpha_z)$  can be computed then
21:          if  $p_{(X_{\text{rule}}, y_{\text{rule}})}(\delta_k | \alpha_z) > p_{\text{max}} \vee ( p_{(X_{\text{rule}}, y_{\text{rule}})}(\delta_k | \alpha_z) = p_{\text{max}} \wedge \text{Coverage}_{R_{\text{now}}} > n_{\text{max}} )$  then
22:             $p_{\text{max}} \leftarrow p_{(X_{\text{rule}}, y_{\text{rule}})}(\delta_k | \alpha_z)$ 
23:             $n_{\text{max}} \leftarrow \text{Coverage}_{R_{\text{now}}}$ 
24:             $\hat{\alpha}_{z_0} \leftarrow \alpha_z$ 
25:          end if
26:        end if
27:      end for
28:
29:      // Updating rule
30:       $R \leftarrow R \cup \hat{\alpha}_{z_0}$ 
31:      Remove all  $\alpha_z$  in  $S_{\text{attributes}}$  s.t.  $\alpha = \hat{\alpha}$ 
32:      Filter  $(X_{\text{rule}}, y_{\text{rule}})$  s.t. all instances fire rule  $R$ 
33:    until  $(S_{\text{attributes}} \neq \emptyset) \vee (\text{all instances in } y_{\text{rule}} \text{ belong to class } \delta_k)$ 
34:
35:    // Adding rule to the set of existing rules
36:    Compute the coverage of rule  $R$  –  $\text{Coverage}_R$ 
37:     $\text{all}_R \leftarrow \text{all}_R \cup \{R, \delta_k, \text{Coverage}_R\}$ 
38:    Filter  $(X_{\text{remaining}}, y_{\text{remaining}})$  so no instance fires rule  $R$ 
39:  until no instances with label  $\delta_k$  in  $X_{\text{remaining}}$ 
40:
41: end for
42:
43: return  $\text{all}_R$ 
```

---

---

**Algorithm 2** Prediction algorithm

---

**Input:**  $X_{\text{tst}}$  – Test features,  $y_{\text{trn}}$  – Train labels,  $\text{all}_R$  – set of rules

**Output:**  $\hat{y}$

```
1: Initialize  $\delta_{\text{mode}}$  as the mode of the training labels (computed from  $y_{\text{trn}}$ )
2:  $\hat{y} \leftarrow []$ 
3: for each instance  $x_j$  in  $X_{\text{tst}}$  do
4:    $\hat{\delta} \leftarrow \text{nothing}$ 
5:    $n_{\text{max}} \leftarrow -1$ 
6:   for each  $\{R, \delta_k, \text{Coverage}_R\} \in \text{all}_R$  do
7:     if  $(\text{Coverage}_R > n_{\text{max}}) \wedge (x_j \text{ fires } R)$  then
8:        $n_{\text{max}} \leftarrow \text{Coverage}_R$ 
9:        $\hat{\delta} \leftarrow \delta_k$ 
10:    end if
11:  end for
12:
13:  // Predict the mode if no rule is fired
14:  if  $\hat{\delta}$  is nothing then
15:     $\hat{\delta} \leftarrow \delta_{\text{mode}}$ 
16:  end if
17:  Append  $\hat{\delta}$  to  $\hat{y}$ 
18: end for
19:
20: return  $\text{all}_R$ 
```

---

### 3 Results

In this subsection, the results of applying the PRISM algorithm and assessing its performance on three datasets will be shown. These datasets are the following:

- Wines: small dataset with 178 instances.
- Breast cancer Wisconsin: medium dataset with 699 instances.
- Seismic bumps: large dataset with 2584 instances.

Analyzing the performance on three different datasets will be important to assess the algorithm’s scalability and behaviour in imbalanced datasets.

In order to feed it to PRISM, discretization of numerical attributes was needed. When deciding the number of bins, explainability was heavily weighted, so all numerical attributes have been transformed into three categories: L (Low), M (Medium) and H (High). In addition, whenever instances had missing values they were imputed as the mean or mode, whatever deemed relevant.

The training procedure consisted of dividing the dataset into two parts, a training and a test set. The first one gathered 80% of the data and will be used to infer the dataset’s rules, while the latter will only be used for testing purposes.

#### 3.1 Wines dataset

##### Description

The Wines dataset contains the results of a chemical analysis of wines from three different cultivars. Each instance has 13 numerical attributes derived from the analysis: Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, Proline.

As previously mentioned, the dataset's instances can be classified into three classes:  $\delta_1$ ,  $\delta_2$  and  $\delta_3$ . Its distribution is the following:  $\delta_1$  – 59 (33.1%),  $\delta_2$  – 71 (39.9%) and  $\delta_3$  – 48 (27.0%), which indicates that it is a well behaved dataset, with an even class distribution.

## Rules

In this subsection, the 10 rules with the highest coverage are shown, while the complete set of rules is displayed in the Appendix 5.1.

- R1: **Proline** = M  $\wedge$  **Total phenols** = M  $\rightarrow \delta_1$  (n = 20, p = 100.0%)
- R2: **Alcohol** = L  $\wedge$  **Hue** = M  $\rightarrow \delta_2$  (n = 17, p = 100.0%)
- R3: **Proline** = H  $\rightarrow \delta_1$  (n = 14, p = 100.0%)
- R4: **Color intensity** = H  $\rightarrow \delta_3$  (n = 10, p = 100.0%)
- R5: **OD280/OD315 of diluted wines** = L  $\wedge$  **Color intensity** = M  $\wedge$  **Hue** = L  $\rightarrow \delta_3$  (n = 10, p = 100.0%)
- R6: **Flavanoids** = H  $\wedge$  **Proline** = M  $\wedge$  **Ash** = M  $\rightarrow \delta_1$  (n = 9, p = 100.0%)
- R7: **OD280/OD315 of diluted wines** = L  $\wedge$  **Malic acid** = H  $\rightarrow \delta_3$  (n = 9, p = 100.0%)
- R8: **Ash** = L  $\rightarrow \delta_2$  (n = 9, p = 100.0%)
- R9: **Hue** = H  $\rightarrow \delta_2$  (n = 8, p = 100.0%)
- R10: **OD280/OD315 of diluted wines** = L  $\wedge$  **Ash** = H  $\rightarrow \delta_3$  (n = 5, p = 100.0%)

## Metrics

This subsection will show the Accuracy and the weighted/macro Precision, Recall and F1-Score in the train and test sets.

Table 1: Macro average metrics in the Wines dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	100.0	100.0	100.0	100.0
Test	91.7	93.3	93.0	93.1

Table 2: Weighted average metrics in the Wines dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	100.0	100.0	100.0	100.0
Test	91.7	91.8	91.7	91.6

## Analysis

Regarding the rules obtained, it is important to point out that all of them have 100% precision. That means that all of the training instances that fire them are correctly classified. This particular fact is a feature from PRISM algorithm since it has two stopping criteria: 100% precision (all instances correctly classified) or lack of attributes to add to the antecedents. In addition, it is interesting to see that the maximum number of conditions in a single rule is equal to 3 and the total number of rules is 23.

All of these findings help one conclude that training data can be divided into 23 homogeneous groups, i.e., with the same label (same cultivar). If data was uniformly generated, the performance on the test set should be very high. Indeed, as one can see in tables 1 and 2 all of the metrics are above 91%.



The first thing that should be commented is the drop in performance in the training and test set. This is completely normal since this algorithm relies on overfitting to generalize in another dataset. Also, no bias seems to happen in the algorithm since the difference in performance between the macro and weighted average is very slim. In other words, the macro average, which does not take into account the number of instances in each class performs similar than the weighted one.

All in all, PRISM has solved this dataset and it has developed a set of rules that can generalize on unseen data. It is also interpretable since one can easily see where the combinations with highest weight (coverage) and precision in the model are.

### 3.2 Breast cancer Wisconsin dataset

#### Description

Data was collected by Dr. William H. Wolberg at the University of Wisconsin Hospitals, Madison [5], culminating in a breast cancer database consisting of several attributes: Sample code number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses. The first one, Sample code number, has been dropped since it does not have predicting power. The others are integer valued (1 – 10) and have been discretized into three categories: L, M, H. It should be noted that the column Bare Nuclei was the only one that had missing values and its mode was imputed.

The classification aims to predict instances into two groups:  $\delta_2$  and  $\delta_4$ . The first class,  $\delta_2$ , represents benign cases, while  $\delta_4$  represents malignant ones. The distribution is the following:  $\delta_2$  – 458 (65.5%) and  $\delta_4$  – 241 (34.5%). Since the objective is to predict malignant cases, this problem has been addressed as binary classification.

#### Rules

In this subsection, the 10 rules with the highest coverage are shown, while the complete set of rules is displayed in the Appendix 5.2.

- R1: **Clump Thickness = L  $\wedge$  Bare Nuclei = L  $\wedge$  Marginal Adhesion = L  $\rightarrow \delta_2$**  (n = 223, p = 100.0%)
- R2: **Clump Thickness = M  $\wedge$  Uniformity of Cell Shape = L  $\wedge$  Marginal Adhesion = L  $\wedge$  Uniformity of Cell Size = L  $\wedge$  Bland Chromatin = L  $\wedge$  Single Epithelial Cell Size = L  $\wedge$  Normal Nucleoli = L  $\wedge$  Mitoses = L  $\wedge$  Bare Nuclei = L  $\rightarrow \delta_4$**  (n = 107, p = 0.9%)
- R3: **Uniformity of Cell Shape = L  $\wedge$  Bare Nuclei = L  $\wedge$  Bland Chromatin = L  $\wedge$  Single Epithelial Cell Size = L  $\wedge$  Marginal Adhesion = L  $\wedge$  Mitoses = L  $\wedge$  Clump Thickness = M  $\wedge$  Uniformity of Cell Size = L  $\wedge$  Normal Nucleoli = L  $\rightarrow \delta_2$**  (n = 107, p = 99.1%)
- R4: **Uniformity of Cell Shape = H  $\wedge$  Bland Chromatin = H  $\rightarrow \delta_4$**  (n = 59, p = 100.0%)
- R5: **Marginal Adhesion = H  $\wedge$  Bare Nuclei = H  $\rightarrow \delta_4$**  (n = 25, p = 100.0%)
- R6: **Uniformity of Cell Size = H  $\wedge$  Clump Thickness = H  $\rightarrow \delta_4$**  (n = 19, p = 100.0%)
- R7: **Mitoses = M  $\rightarrow \delta_4$**  (n = 17, p = 100.0%)
- R8: **Bare Nuclei = H  $\wedge$  Bland Chromatin = M  $\rightarrow \delta_4$**  (n = 16, p = 100.0%)
- R9: **Clump Thickness = H  $\wedge$  Bare Nuclei = M  $\rightarrow \delta_4$**  (n = 9, p = 100.0%)
- R10: **Bare Nuclei = H  $\wedge$  Clump Thickness = H  $\wedge$  Marginal Adhesion = L  $\rightarrow \delta_4$**  (n = 9, p = 100.0%)

#### Metrics

Since this is a binary classification problem, the reported metrics are different. In table 3 one can see the accuracy, precision, recall and F1-score of predicting a malignant instance,  $\delta_4$ . After that, in table 4 the weighted average metrics can be seen.

Table 3: Binary classification metrics in the Breast cancer Wisconsin dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	80.7	63.8	100.0	77.9
Test	70.0	55.6	88.2	68.2

Table 4: Weighted average metrics in the Breast cancer Wisconsin dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	80.7	87.7	80.7	81.2
Test	70.0	77.3	70.0	70.4

## Analysis

The rules in this dataset are interesting. Except R2 and R3, all of them have 100% precision. The former, R2 and R3, are complementary due to the discretization, since they both have the same antecedents, but yield different predictions. The algorithm has tried all the parameter combinations, but it seems like these instances share the same features but have different class values. Consequently, R2 covers 107 instances with a 0.9% precision (of predicting cancer) while R3 covers the same instances but with a 99.1% precision (of benign tumor detection).

Interestingly, the recall shown in table 3 is 100.0% while in the weighted case (table 4) it is 80.7%. This means that the algorithm recalls 100% of the cancer tumors with a precision that is lower than 100%. This contrasts one of the features of PRISM, which states that the rules will create a classifier with 100% precision and recall.

The previous statement stems from the fact that discretization can make instances that were previously distinguishable, virtually similar. In addition, the classifier will not know what to do, since the input will be equal. Therefore, two rules will divide instances with the same features but with different target values. When predicting one of these instances, the predicted class will be chosen at random, since the coverage by definition is the same. Obviously, this is a problem since it increases the instability of the classifier.

## 3.3 Seismic bumps dataset

### Description

Mining activity can bring situation which are called mining hazards. A special case is called a high energy seismic hazard which frequently occurs in many underground mines. As mentioned by the authors of [4], in the data set, each row contains a summary statement about seismic activity in the rock mass within one shift (8 hours). If decision attribute has the value 1, then in the next shift any seismic bump with an energy higher than  $10^4$  J was registered. Instances have 18 attributes, of which 4 are categorical:

- seismic:  $\{a, b, c, d\}$
- seismoacoustic:  $\{a, b, c, d\}$
- shift:  $\{W, N\}$
- ghazard:  $\{a, b, c, d\}$

The rest of the attributes are real valued: genergy, gpuls, gdenergy, gdpuls, nbumps, nbumps2, nbumps3, nbumps4, nbumps5, nbumps6, nbumps7, nbumps89, energy and maxenergy.

This dataset is part of a binary classification problem where a positive prediction means that there is earthquake hazard. Consequently, good predictions of the positive class are very important and of practical usage. The codification of the classes is the following: no earthquake hazard ( $\delta_0$ ) and earthquake hazard ( $\delta_1$ ). Taking a look at the class distribution, one can see that data is extremely imbalanced, with  $\delta_1$  only representing 6.6% of the instances.

## Rules

In this subsection, the 10 rules with the highest coverage are shown, while the complete set of rules is displayed in the Appendix 5.3.

- R1: **shift** = N  $\wedge$  **seismoacoustic** = a  $\wedge$  **nbumps** = L  $\wedge$  **nbumps4** = L  $\wedge$  **gdpuls** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **gpuls** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **seismic** = a  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_0$  (n = 406, p = 98.0%)
- R2: **seismoacoustic** = a  $\wedge$  **seismic** = a  $\wedge$  **shift** = N  $\wedge$  **gdpuls** = L  $\wedge$  **nbumps** = L  $\wedge$  **nbumps4** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **gpuls** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_1$  (n = 406, p = 2.0%)
- R3: **nbumps** = L  $\wedge$  **gpuls** = L  $\wedge$  **seismoacoustic** = a  $\wedge$  **seismic** = a  $\wedge$  **gdpuls** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **nbumps4** = L  $\wedge$  **shift** = W  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_0$  (n = 312, p = 94.9%)
- R4: **shift** = W  $\wedge$  **seismoacoustic** = a  $\wedge$  **seismic** = a  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps4** = L  $\wedge$  **genergy** = L  $\wedge$  **gpuls** = L  $\wedge$  **gdpuls** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **nbumps** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\rightarrow \delta_1$  (n = 312, p = 5.1%)
- R5: **nbumps** = L  $\wedge$  **gpuls** = L  $\wedge$  **seismoacoustic** = a  $\wedge$  **nbumps4** = L  $\wedge$  **nbumps3** = L  $\wedge$  **gdpuls** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **shift** = W  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **seismic** = b  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_0$  (n = 225, p = 93.3%)
- R6: **shift** = W  $\wedge$  **ghazard** = a  $\wedge$  **nbumps4** = L  $\wedge$  **seismic** = b  $\wedge$  **seismoacoustic** = a  $\wedge$  **nbumps** = L  $\wedge$  **gpuls** = L  $\wedge$  **energy** = L  $\wedge$  **gdpuls** = L  $\wedge$  **nbumps7** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_1$  (n = 225, p = 6.7%)
- R7: **seismoacoustic** = b  $\wedge$  **nbumps** = L  $\wedge$  **gpuls** = L  $\wedge$  **shift** = W  $\wedge$  **ghazard** = a  $\wedge$  **genergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps4** = L  $\wedge$  **gdpuls** = L  $\wedge$  **nbumps7** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **seismic** = a  $\wedge$  **nbumps89** = L  $\rightarrow \delta_0$  (n = 177, p = 93.8%)
- R8: **shift** = W  $\wedge$  **seismoacoustic** = b  $\wedge$  **ghazard** = a  $\wedge$  **nbumps** = L  $\wedge$  **seismic** = a  $\wedge$  **nbumps4** = L  $\wedge$  **gpuls** = L  $\wedge$  **energy** = L  $\wedge$  **gdpuls** = L  $\wedge$  **nbumps7** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_1$  (n = 177, p = 6.2%)
- R9: **shift** = N  $\wedge$  **seismic** = a  $\wedge$  **seismoacoustic** = b  $\wedge$  **nbumps4** = L  $\wedge$  **gdpuls** = L  $\wedge$  **ghazard** = a  $\wedge$  **nbumps7** = L  $\wedge$  **nbumps** = L  $\wedge$  **gpuls** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps3** = L  $\wedge$  **nbumps2** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_0$  (n = 155, p = 98.7%)
- R10: **seismic** = a  $\wedge$  **shift** = N  $\wedge$  **ghazard** = a  $\wedge$  **gdpuls** = L  $\wedge$  **nbumps** = L  $\wedge$  **nbumps4** = L  $\wedge$  **nbumps3** = L  $\wedge$  **energy** = L  $\wedge$  **nbumps7** = L  $\wedge$  **gpuls** = L  $\wedge$  **gdenenergy** = L  $\wedge$  **nbumps2** = L  $\wedge$  **nbumps6** = L  $\wedge$  **nbumps5** = L  $\wedge$  **maxenergy** = L  $\wedge$  **seismoacoustic** = b  $\wedge$  **nbumps89** = L  $\wedge$  **genergy** = L  $\rightarrow \delta_1$  (n = 155, p = 1.3%)

## Metrics

This subsection will show the Accuracy and the weighted/macro Precision, Recall and F1-Score in the train and test sets.

Table 5: Macro average metrics in the Seismic bumps dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	94.3	100.0	13.9	24.4
Test	92.6	14.3	3.0	5.0

Table 6: Weighted average metrics in the Seismic bumps dataset

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Train	94.3	94.6	94.3	92.2
Test	92.6	88.7	92.6	90.4

## Analysis

The results from this dataset are very interesting to analyze. It contains over 2500 instances, so it is the test the algorithm needs to prove it can generalize in bigger and more complex datasets.

First of all, the 10 rules with the highest coverage are complementary. This concept was explained in the previous section (Breast cancer Wisconsin), which makes the classifier unstable, since the instances with the same features are predicted the class value at random. In this case, judging from table 5, one can see that the recall at 13.9% is low while the precision at 100.0% is very high. This behaviour is interesting since in the dataset Breast cancer Wisconsin the opposite happened, recall was high, while the precision was low.

The results that follow in the test set are expected. As one can see in table 5, only 3% of the seismic bumps were recalled and the precision was at 14.3%. The performance observed can be attributed to the fact that the dataset is imbalanced and PRISM can not compensate this situation. Therefore, the algorithm ends up predicting the majority class (no hazard of seismic bump).

This results do not mean that the algorithm is not “working” per se. In fact, if one observes table 6, the weighted metrics are similar in the train and test sets and are high. It just means that the algorithm is mostly predicting the majority class which accounts for the 94-95% of data.

Overall, PRISM has not generalized as it was expected. Part of the suboptimal performance comes from the dataset being imbalanced and the other part comes from the algorithm not being powerful enough to distinguish instances from the minimal class.

## 4 Conclusion

To conclude this report, it is important to highlight that all of the objectives were met. Firstly, a classifier was created so that it could read CSV files, induce the set of rules, and display them and their coverage and precision. Additionally, the classifier was improved from the base form and applied to three different databases: Wines (small), Breast cancer Wisconsin (medium) and Seismic bumps (large). Subsequently, the metrics in the train and test sets were displayed, and they were analyzed.

One of the main strengths observed is the interpretability since one can see the rules with the highest coverage and the attributes assigned to them. Therefore, when one is building an expert system it will be very useful. In addition, in small and medium sized datasets, the accuracy is quite high, so it can even be comparable with other methods where interpretability is not even considered.

In contrast, when the dataset grew in size, the performance deteriorated. It must be said that the Seismic bumps (large) dataset, was heavily imbalanced. In the future, some resampling techniques or data generation could be used. Distribution of the minimal class could be estimated and used to generate more data and upsample that class.

All in all, PRISM is a solid classifier which should be used in datasets with categorical attributes. If not, one is at risk of discretizing data with a low number of bins, transforming some instances to the same set of features, but

with a different class value, and giving not ideal predictions. However, given the right dataset it brings a lot of advantages to the table.

## References

- [1] Jadzia Cendrowska. Prism: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27(4):349–370, 1987.
- [2] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [3] Mohammad Hossin and Md Nasir Sulaiman. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1, 2015.
- [4] Andrzej Janusz, Marek Grzegorowski, Marcin Michalak, Łukasz Wróbel, Marek Sikora, and Dominik Slezak. Predicting seismic events in coal mines based on underground sensor measurements. *Engineering Applications of Artificial Intelligence*, 64:83–94, 2017.
- [5] Olvi L Mangasarian and William H Wolberg. Cancer diagnosis via linear programming. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 1990.
- [6] B. S. Panda. Rule based classification. <https://web.iitd.ac.in/~bspanda/rb.pdf>.
- [7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

## 5 Appendix

### 5.1 Wines dataset

- R1: **Proline** = M  $\wedge$  **Total phenols** = M  $\rightarrow \delta_1$  (n = 20, p = 100.0%)
- R2: **Alcohol** = L  $\wedge$  **Hue** = M  $\rightarrow \delta_2$  (n = 17, p = 100.0%)
- R3: **Proline** = H  $\rightarrow \delta_1$  (n = 14, p = 100.0%)
- R4: **Color intensity** = H  $\rightarrow \delta_3$  (n = 10, p = 100.0%)
- R5: **OD280/OD315 of diluted wines** = L  $\wedge$  **Color intensity** = M  $\wedge$  **Hue** = L  $\rightarrow \delta_3$  (n = 10, p = 100.0%)
- R6: **Flavanoids** = H  $\wedge$  **Proline** = M  $\wedge$  **Ash** = M  $\rightarrow \delta_1$  (n = 9, p = 100.0%)
- R7: **OD280/OD315 of diluted wines** = L  $\wedge$  **Malic acid** = H  $\rightarrow \delta_3$  (n = 9, p = 100.0%)
- R8: **Ash** = L  $\rightarrow \delta_2$  (n = 9, p = 100.0%)
- R9: **Hue** = H  $\rightarrow \delta_2$  (n = 8, p = 100.0%)
- R10: **OD280/OD315 of diluted wines** = L  $\wedge$  **Ash** = H  $\rightarrow \delta_3$  (n = 5, p = 100.0%)
- R11: **Color intensity** = L  $\wedge$  **Magnesium** = L  $\wedge$  **Malic acid** = L  $\rightarrow \delta_2$  (n = 5, p = 100.0%)
- R12: **Flavanoids** = M  $\wedge$  **Proline** = L  $\wedge$  **Color intensity** = L  $\rightarrow \delta_2$  (n = 5, p = 100.0%)
- R13: **Flavanoids** = L  $\wedge$  **Hue** = L  $\wedge$  **Alcalinity of ash** = M  $\rightarrow \delta_3$  (n = 4, p = 100.0%)
- R14: **Alcohol** = L  $\wedge$  **Flavanoids** = M  $\rightarrow \delta_2$  (n = 3, p = 100.0%)
- R15: **Nonflavanoid phenols** = H  $\wedge$  **Alcalinity of ash** = L  $\rightarrow \delta_2$  (n = 3, p = 100.0%)
- R16: **Flavanoids** = H  $\wedge$  **Alcohol** = H  $\wedge$  **Nonflavanoid phenols** = M  $\rightarrow \delta_1$  (n = 2, p = 100.0%)
- R17: **Flavanoids** = M  $\wedge$  **Nonflavanoid phenols** = H  $\rightarrow \delta_2$  (n = 2, p = 100.0%)
- R18: **Flavanoids** = M  $\wedge$  **Malic acid** = H  $\rightarrow \delta_2$  (n = 2, p = 100.0%)
- R19: **OD280/OD315 of diluted wines** = L  $\wedge$  **Magnesium** = H  $\rightarrow \delta_3$  (n = 1, p = 100.0%)
- R20: **OD280/OD315 of diluted wines** = L  $\wedge$  **Alcohol** = H  $\rightarrow \delta_3$  (n = 1, p = 100.0%)
- R21: **OD280/OD315 of diluted wines** = L  $\wedge$  **Alcalinity of ash** = H  $\rightarrow \delta_3$  (n = 1, p = 100.0%)
- R22: **Alcohol** = L  $\wedge$  **Flavanoids** = H  $\rightarrow \delta_2$  (n = 1, p = 100.0%)
- R23: **Magnesium** = L  $\wedge$  **Alcalinity of ash** = L  $\wedge$  **Hue** = L  $\rightarrow \delta_2$  (n = 1, p = 100.0%)

### 5.2 Breast cancer Wisconsin dataset

- R1: **Clump Thickness** = L  $\wedge$  **Bare Nuclei** = L  $\wedge$  **Marginal Adhesion** = L  $\rightarrow \delta_2$  (n = 223, p = 100.0%)
- R2: **Clump Thickness** = M  $\wedge$  **Uniformity of Cell Shape** = L  $\wedge$  **Marginal Adhesion** = L  $\wedge$  **Uniformity of Cell Size** = L  $\wedge$  **Bland Chromatin** = L  $\wedge$  **Single Epithelial Cell Size** = L  $\wedge$  **Normal Nucleoli** = L  $\wedge$  **Mitoses** = L  $\wedge$  **Bare Nuclei** = L  $\rightarrow \delta_4$  (n = 107, p = 0.9%)
- R3: **Uniformity of Cell Shape** = L  $\wedge$  **Bare Nuclei** = L  $\wedge$  **Bland Chromatin** = L  $\wedge$  **Single Epithelial Cell Size** = L  $\wedge$  **Marginal Adhesion** = L  $\wedge$  **Mitoses** = L  $\wedge$  **Clump Thickness** = M  $\wedge$  **Uniformity of Cell Size** = L  $\wedge$  **Normal Nucleoli** = L  $\rightarrow \delta_2$  (n = 107, p = 99.1%)
- R4: **Uniformity of Cell Shape** = H  $\wedge$  **Bland Chromatin** = H  $\rightarrow \delta_4$  (n = 59, p = 100.0%)
- R5: **Marginal Adhesion** = H  $\wedge$  **Bare Nuclei** = H  $\rightarrow \delta_4$  (n = 25, p = 100.0%)
- R6: **Uniformity of Cell Size** = H  $\wedge$  **Clump Thickness** = H  $\rightarrow \delta_4$  (n = 19, p = 100.0%)

- R7: **Mitoses** =  $M \rightarrow \delta_4$  ( $n = 17, p = 100.0\%$ )
- R8: **Bare Nuclei** =  $H \wedge \text{Bland Chromatin} = M \rightarrow \delta_4$  ( $n = 16, p = 100.0\%$ )
- R9: **Clump Thickness** =  $H \wedge \text{Bare Nuclei} = M \rightarrow \delta_4$  ( $n = 9, p = 100.0\%$ )
- R10: **Bare Nuclei** =  $H \wedge \text{Clump Thickness} = H \wedge \text{Marginal Adhesion} = L \rightarrow \delta_4$  ( $n = 9, p = 100.0\%$ )
- R11: **Uniformity of Cell Shape** =  $L \wedge \text{Bland Chromatin} = L \wedge \text{Clump Thickness} = L \rightarrow \delta_2$  ( $n = 8, p = 100.0\%$ )
- R12: **Bare Nuclei** =  $L \wedge \text{Clump Thickness} = M \wedge \text{Bland Chromatin} = L \rightarrow \delta_2$  ( $n = 8, p = 100.0\%$ )
- R13: **Bland Chromatin** =  $H \wedge \text{Normal Nucleoli} = M \rightarrow \delta_4$  ( $n = 4, p = 100.0\%$ )
- R14: **Bare Nuclei** =  $H \wedge \text{Bland Chromatin} = H \wedge \text{Normal Nucleoli} = L \rightarrow \delta_4$  ( $n = 4, p = 100.0\%$ )
- R15: **Clump Thickness** =  $H \wedge \text{Normal Nucleoli} = M \rightarrow \delta_4$  ( $n = 3, p = 100.0\%$ )
- R16: **Bland Chromatin** =  $H \wedge \text{Single Epithelial Cell Size} = H \rightarrow \delta_4$  ( $n = 3, p = 100.0\%$ )
- R17: **Bare Nuclei** =  $H \wedge \text{Single Epithelial Cell Size} = L \wedge \text{Uniformity of Cell Size} = M \rightarrow \delta_4$  ( $n = 3, p = 100.0\%$ )
- R18: **Uniformity of Cell Shape** =  $L \wedge \text{Bare Nuclei} = L \wedge \text{Marginal Adhesion} = M \rightarrow \delta_2$  ( $n = 3, p = 100.0\%$ )
- R19: **Uniformity of Cell Shape** =  $L \wedge \text{Bare Nuclei} = L \wedge \text{Bland Chromatin} = L \wedge \text{Clump Thickness} = H \rightarrow \delta_2$  ( $n = 3, p = 100.0\%$ )
- R20: **Normal Nucleoli** =  $H \wedge \text{Uniformity of Cell Size} = L \rightarrow \delta_4$  ( $n = 2, p = 100.0\%$ )
- R21: **Mitoses** =  $H \wedge \text{Clump Thickness} = H \rightarrow \delta_4$  ( $n = 2, p = 100.0\%$ )
- R22: **Uniformity of Cell Size** =  $H \wedge \text{Single Epithelial Cell Size} = H \rightarrow \delta_4$  ( $n = 2, p = 100.0\%$ )
- R23: **Clump Thickness** =  $H \wedge \text{Bland Chromatin} = H \wedge \text{Marginal Adhesion} = L \rightarrow \delta_4$  ( $n = 2, p = 100.0\%$ )
- R24: **Marginal Adhesion** =  $H \wedge \text{Normal Nucleoli} = H \wedge \text{Uniformity of Cell Size} = M \wedge \text{Mitoses} = L \wedge \text{Bland Chromatin} = H \wedge \text{Clump Thickness} = M \wedge \text{Bare Nuclei} = L \wedge \text{Single Epithelial Cell Size} = M \wedge \text{Uniformity of Cell Shape} = M \rightarrow \delta_4$  ( $n = 2, p = 50.0\%$ )
- R25: **Bare Nuclei** =  $H \wedge \text{Uniformity of Cell Shape} = L \wedge \text{Clump Thickness} = M \wedge \text{Single Epithelial Cell Size} = L \wedge \text{Mitoses} = L \wedge \text{Normal Nucleoli} = M \wedge \text{Uniformity of Cell Size} = L \wedge \text{Marginal Adhesion} = M \wedge \text{Bland Chromatin} = L \rightarrow \delta_4$  ( $n = 2, p = 50.0\%$ )
- R26: **Uniformity of Cell Shape** =  $L \wedge \text{Bland Chromatin} = L \wedge \text{Uniformity of Cell Size} = M \rightarrow \delta_2$  ( $n = 2, p = 100.0\%$ )
- R27: **Uniformity of Cell Shape** =  $L \wedge \text{Clump Thickness} = L \wedge \text{Single Epithelial Cell Size} = L \rightarrow \delta_2$  ( $n = 2, p = 100.0\%$ )
- R28: **Bland Chromatin** =  $L \wedge \text{Clump Thickness} = M \wedge \text{Uniformity of Cell Size} = H \rightarrow \delta_2$  ( $n = 2, p = 100.0\%$ )
- R29: **Bare Nuclei** =  $L \wedge \text{Clump Thickness} = M \wedge \text{Normal Nucleoli} = L \wedge \text{Uniformity of Cell Shape} = L \rightarrow \delta_2$  ( $n = 2, p = 100.0\%$ )
- R30: **Bare Nuclei** =  $L \wedge \text{Uniformity of Cell Size} = M \wedge \text{Normal Nucleoli} = H \wedge \text{Mitoses} = L \wedge \text{Bland Chromatin} = H \wedge \text{Clump Thickness} = M \wedge \text{Marginal Adhesion} = H \wedge \text{Single Epithelial Cell Size} = M \wedge \text{Uniformity of Cell Shape} = M \rightarrow \delta_2$  ( $n = 2, p = 50.0\%$ )
- R31: **Marginal Adhesion** =  $M \wedge \text{Uniformity of Cell Shape} = L \wedge \text{Bland Chromatin} = L \wedge \text{Single Epithelial Cell Size} = L \wedge \text{Bare Nuclei} = H \wedge \text{Mitoses} = L \wedge \text{Clump Thickness} = M \wedge \text{Normal Nucleoli} = M \wedge \text{Uniformity of Cell Size} = L \rightarrow \delta_2$  ( $n = 2, p = 50.0\%$ )



- R32: **Bare Nuclei** =  $H \wedge \text{Mitoses} = H \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R33: **Marginal Adhesion** =  $H \wedge \text{Single Epithelial Cell Size} = L \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R34: **Normal Nucleoli** =  $M \wedge \text{Bare Nuclei} = M \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R35: **Bare Nuclei** =  $H \wedge \text{Uniformity of Cell Shape} = L \wedge \text{Single Epithelial Cell Size} = M \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R36: **Bland Chromatin** =  $M \wedge \text{Single Epithelial Cell Size} = M \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R37: **Normal Nucleoli** =  $M \wedge \text{Bland Chromatin} = M \wedge \text{Uniformity of Cell Shape} = L \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R38: **Bare Nuclei** =  $M \wedge \text{Clump Thickness} = M \wedge \text{Bland Chromatin} = M \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R39: **Bare Nuclei** =  $M \wedge \text{Clump Thickness} = M \wedge \text{Marginal Adhesion} = L \wedge \text{Uniformity of Cell Size} = L \rightarrow \delta_4$  ( $n = 1, p = 100.0\%$ )
- R40: **Uniformity of Cell Shape** =  $L \wedge \text{Bland Chromatin} = L \wedge \text{Marginal Adhesion} = H \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R41: **Uniformity of Cell Shape** =  $L \wedge \text{Bare Nuclei} = L \wedge \text{Bland Chromatin} = L \wedge \text{Normal Nucleoli} = M \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R42: **Uniformity of Cell Shape** =  $L \wedge \text{Bare Nuclei} = L \wedge \text{Bland Chromatin} = L \wedge \text{Single Epithelial Cell Size} = M \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R43: **Bare Nuclei** =  $L \wedge \text{Clump Thickness} = M \wedge \text{Uniformity of Cell Size} = M \wedge \text{Single Epithelial Cell Size} = H \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R44: **Uniformity of Cell Shape** =  $L \wedge \text{Marginal Adhesion} = M \wedge \text{Bare Nuclei} = M \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R45: **Bland Chromatin** =  $L \wedge \text{Clump Thickness} = M \wedge \text{Marginal Adhesion} = M \wedge \text{Single Epithelial Cell Size} = H \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R46: **Bare Nuclei** =  $L \wedge \text{Uniformity of Cell Size} = M \wedge \text{Marginal Adhesion} = L \wedge \text{Uniformity of Cell Shape} = M \wedge \text{Bland Chromatin} = M \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )
- R47: **Marginal Adhesion** =  $M \wedge \text{Uniformity of Cell Size} = M \wedge \text{Normal Nucleoli} = H \wedge \text{Uniformity of Cell Shape} = M \wedge \text{Clump Thickness} = H \rightarrow \delta_2$  ( $n = 1, p = 100.0\%$ )

### 5.3 Seismic bumps dataset

- R1: **shift** =  $N \wedge \text{seismoacoustic} = a \wedge \text{nbumps} = L \wedge \text{nbumps4} = L \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{gpuls} = L \wedge \text{gdenery} = L \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{seismic} = a \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_0$  ( $n = 406, p = 98.0\%$ )
- R2: **seismoacoustic** =  $a \wedge \text{seismic} = a \wedge \text{shift} = N \wedge \text{gdpuls} = L \wedge \text{nbumps} = L \wedge \text{nbumps4} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{gpuls} = L \wedge \text{gdenery} = L \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_1$  ( $n = 406, p = 2.0\%$ )
- R3: **nbumps** =  $L \wedge \text{gpuls} = L \wedge \text{seismoacoustic} = a \wedge \text{seismic} = a \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps4} = L \wedge \text{shift} = W \wedge \text{gdenery} = L \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_0$  ( $n = 312, p = 94.9\%$ )
- R4: **shift** =  $W \wedge \text{seismoacoustic} = a \wedge \text{seismic} = a \wedge \text{nbumps3} = L \wedge \text{nbumps4} = L \wedge \text{genergy} = L \wedge \text{gpuls} = L \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps} = L \wedge \text{gdenery} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \rightarrow \delta_1$  ( $n = 312, p = 5.1\%$ )











- [illegible]

- R95:  $\text{nbumps} = M \wedge \text{gpuls} = L \wedge \text{shift} = W \wedge \text{seismoacoustic} = b \wedge \text{nbumps3} = M \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps4} = L \wedge \text{gdenenergy} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{seismic} = b \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_0$  (n = 3, p = 66.7%)
- R96:  $\text{gpuls} = M \wedge \text{genergy} = H \wedge \text{nbumps4} = L \wedge \text{nbumps3} = L \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps} = L \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{seismoacoustic} = a \wedge \text{seismic} = a \wedge \text{nbumps89} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R97:  $\text{nbumps3} = M \wedge \text{nbumps4} = M \wedge \text{nbumps} = H \wedge \text{gpuls} = L \wedge \text{seismoacoustic} = b \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps2} = M \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{seismic} = b \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R98:  $\text{nbumps3} = M \wedge \text{seismic} = b \wedge \text{nbumps} = M \wedge \text{nbumps4} = L \wedge \text{genergy} = L \wedge \text{gpuls} = M \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{seismoacoustic} = a \wedge \text{nbumps89} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R99:  $\text{nbumps3} = M \wedge \text{gpuls} = L \wedge \text{nbumps} = M \wedge \text{seismic} = b \wedge \text{seismoacoustic} = b \wedge \text{nbumps2} = L \wedge \text{gdpuls} = L \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{nbumps4} = L \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \wedge \text{genergy} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R100:  $\text{gpuls} = M \wedge \text{nbumps} = L \wedge \text{nbumps3} = L \wedge \text{nbumps4} = L \wedge \text{seismic} = b \wedge \text{gdpuls} = L \wedge \text{seismoacoustic} = b \wedge \text{genergy} = M \wedge \text{ghazard} = a \wedge \text{nbumps7} = L \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R101:  $\text{shift} = W \wedge \text{ghazard} = a \wedge \text{seismoacoustic} = b \wedge \text{genergy} = M \wedge \text{gpuls} = L \wedge \text{seismic} = a \wedge \text{gdpuls} = L \wedge \text{nbumps7} = L \wedge \text{nbumps} = L \wedge \text{nbumps4} = L \wedge \text{gdenenergy} = L \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \rightarrow \delta_1$  (n = 3, p = 33.3%)
- R102:  $\text{shift} = N \wedge \text{seismoacoustic} = c \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R103:  $\text{shift} = N \wedge \text{nbumps3} = L \wedge \text{nbumps} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R104:  $\text{nbumps} = L \wedge \text{nbumps4} = H \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R105:  $\text{seismoacoustic} = b \wedge \text{nbumps} = L \wedge \text{genergy} = H \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R106:  $\text{ghazard} = b \wedge \text{nbumps3} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R107:  $\text{ghazard} = b \wedge \text{seismic} = b \wedge \text{nbumps2} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R108:  $\text{ghazard} = b \wedge \text{seismic} = b \wedge \text{gpuls} = L \wedge \text{nbumps4} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R109:  $\text{genergy} = M \wedge \text{seismoacoustic} = c \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R110:  $\text{seismic} = a \wedge \text{genergy} = L \wedge \text{seismoacoustic} = b \wedge \text{nbumps3} = H \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R111:  $\text{nbumps3} = M \wedge \text{nbumps} = L \wedge \text{seismic} = b \wedge \text{gpuls} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R112:  $\text{nbumps} = L \wedge \text{nbumps3} = M \wedge \text{seismoacoustic} = b \wedge \text{gpuls} = M \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R113:  $\text{seismic} = b \wedge \text{nbumps3} = H \wedge \text{nbumps4} = L \wedge \text{seismoacoustic} = b \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R114:  $\text{seismoacoustic} = a \wedge \text{genergy} = M \wedge \text{nbumps} = L \rightarrow \delta_0$  (n = 2, p = 100.0%)
- R115:  $\text{ghazard} = b \wedge \text{seismic} = a \wedge \text{nbumps7} = L \wedge \text{gpuls} = L \wedge \text{shift} = W \wedge \text{gdenenergy} = L \wedge \text{nbumps3} = L \wedge \text{energy} = L \wedge \text{nbumps6} = L \wedge \text{nbumps5} = L \wedge \text{maxenergy} = L \wedge \text{nbumps89} = L \wedge \text{genergy} = L \wedge \text{gdpuls} = L \wedge \text{nbumps} = M \wedge \text{nbumps2} = M \wedge \text{nbumps4} = L \wedge \text{seismoacoustic} = b \rightarrow \delta_0$  (n = 2, p = 50.0%)





- R131:  $\text{shift} = N \wedge \text{seismoacoustic} = a \wedge \text{nbumps} = L \wedge \text{seismic} = a \wedge \text{nbumps4} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R132:  $\text{nbumps} = L \wedge \text{gpuls} = L \wedge \text{shift} = W \wedge \text{ghazard} = a \wedge \text{gdpuls} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R133:  $\text{nbumps} = L \wedge \text{gpuls} = L \wedge \text{shift} = W \wedge \text{ghazard} = a \wedge \text{nbumps3} = L \wedge \text{seismoacoustic} = b \wedge \text{seismic} = b \wedge \text{genergy} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R134:  $\text{ghazard} = b \wedge \text{genergy} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R135:  $\text{genergy} = M \wedge \text{nbumps4} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R136:  $\text{seismic} = a \wedge \text{nbumps} = H \wedge \text{nbumps4} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R137:  $\text{seismic} = a \wedge \text{nbumps3} = M \wedge \text{nbumps4} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R138:  $\text{seismic} = a \wedge \text{nbumps3} = M \wedge \text{genergy} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R139:  $\text{nbumps4} = M \wedge \text{gdpuls} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R140:  $\text{nbumps4} = M \wedge \text{nbumps} = L \wedge \text{genergy} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R141:  $\text{seismic} = a \wedge \text{seismoacoustic} = a \wedge \text{gpuls} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R142:  $\text{gpuls} = L \wedge \text{nbumps2} = M \wedge \text{nbumps4} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R143:  $\text{seismic} = a \wedge \text{nbumps} = M \wedge \text{nbumps3} = M \wedge \text{seismoacoustic} = b \wedge \text{nbumps2} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R144:  $\text{nbumps4} = M \wedge \text{gpuls} = L \wedge \text{nbumps} = M \wedge \text{nbumps2} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R145:  $\text{seismic} = a \wedge \text{nbumps3} = M \wedge \text{gpuls} = L \wedge \text{nbumps} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R146:  $\text{nbumps4} = M \wedge \text{nbumps3} = M \wedge \text{nbumps} = M \wedge \text{genergy} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R147:  $\text{nbumps4} = M \wedge \text{nbumps} = L \wedge \text{seismic} = b \wedge \text{seismoacoustic} = b \wedge \text{shift} = N \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R148:  $\text{seismic} = a \wedge \text{nbumps2} = M \wedge \text{nbumps3} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R149:  $\text{seismic} = a \wedge \text{nbumps2} = M \wedge \text{gpuls} = L \wedge \text{nbumps} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R150:  $\text{nbumps} = M \wedge \text{nbumps4} = M \wedge \text{genergy} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R151:  $\text{seismic} = a \wedge \text{nbumps3} = L \wedge \text{seismoacoustic} = a \wedge \text{gpuls} = M \wedge \text{nbumps} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R152:  $\text{nbumps} = M \wedge \text{nbumps2} = M \wedge \text{seismoacoustic} = a \wedge \text{gpuls} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R153:  $\text{nbumps} = M \wedge \text{seismic} = a \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{ghazard} = b \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R154:  $\text{seismic} = b \wedge \text{nbumps3} = H \wedge \text{gpuls} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R155:  $\text{nbumps4} = M \wedge \text{nbumps3} = M \wedge \text{nbumps} = M \wedge \text{seismoacoustic} = b \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R156:  $\text{nbumps4} = M \wedge \text{nbumps3} = M \wedge \text{genergy} = L \wedge \text{gpuls} = M \wedge \text{nbumps} = H \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R157:  $\text{gdpuls} = M \wedge \text{gpuls} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R158:  $\text{nbumps3} = L \wedge \text{gpuls} = M \wedge \text{seismoacoustic} = b \wedge \text{ghazard} = a \wedge \text{nbumps4} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)
- R159:  $\text{seismic} = b \wedge \text{gpuls} = M \wedge \text{seismoacoustic} = b \wedge \text{ghazard} = a \wedge \text{nbumps3} = L \wedge \text{nbumps} = M \rightarrow \delta_0$  (n = 1, p = 100.0%)

- R160:  $\text{nbumps} = H \wedge \text{genergy} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R161:  $\text{nbumps3} = H \wedge \text{nbumps4} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R162:  $\text{gpuls} = M \wedge \text{nbumps3} = H \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R163:  $\text{gpuls} = M \wedge \text{nbumps2} = M \wedge \text{genergy} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R164:  $\text{gpuls} = M \wedge \text{genergy} = H \wedge \text{nbumps} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R165:  $\text{gpuls} = M \wedge \text{nbumps} = H \wedge \text{seismic} = a \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R166:  $\text{gpuls} = M \wedge \text{ghazard} = b \wedge \text{nbumps} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R167:  $\text{gpuls} = M \wedge \text{seismic} = b \wedge \text{nbumps4} = L \wedge \text{seismoacoustic} = a \wedge \text{nbumps3} = L \wedge \text{nbumps2} = L \wedge \text{nbumps} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R168:  $\text{nbumps} = M \wedge \text{seismoacoustic} = c \wedge \text{gpuls} = L \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R169:  $\text{nbumps} = M \wedge \text{shift} = N \wedge \text{seismic} = b \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R170:  $\text{nbumps} = M \wedge \text{shift} = N \wedge \text{nbumps3} = M \wedge \text{seismoacoustic} = a \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R171:  $\text{nbumps} = M \wedge \text{seismic} = b \wedge \text{gpuls} = H \wedge \text{nbumps4} = M \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R172:  $\text{nbumps} = H \wedge \text{nbumps2} = L \wedge \text{seismoacoustic} = a \wedge \text{seismic} = a \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R173:  $\text{nbumps} = H \wedge \text{nbumps4} = H \wedge \text{seismoacoustic} = b \wedge \text{nbumps2} = L \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R174:  $\text{genergy} = M \wedge \text{gpuls} = H \wedge \text{nbumps} = L \wedge \text{seismoacoustic} = b \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R175:  $\text{gpuls} = M \wedge \text{nbumps} = L \wedge \text{genergy} = L \wedge \text{nbumps3} = M \wedge \text{seismoacoustic} = a \wedge \text{seismic} = a \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R176:  $\text{nbumps5} = H \wedge \text{energy} = L \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R177:  $\text{gdpuls} = M \wedge \text{seismoacoustic} = b \wedge \text{gpuls} = L \wedge \text{ghazard} = a \wedge \text{seismic} = b \rightarrow \delta_1$  (n = 1, p = 100.0%)
- R178:  $\text{nbumps4} = M \wedge \text{shift} = N \wedge \text{seismic} = a \wedge \text{seismoacoustic} = b \rightarrow \delta_1$  (n = 1, p = 100.0%)