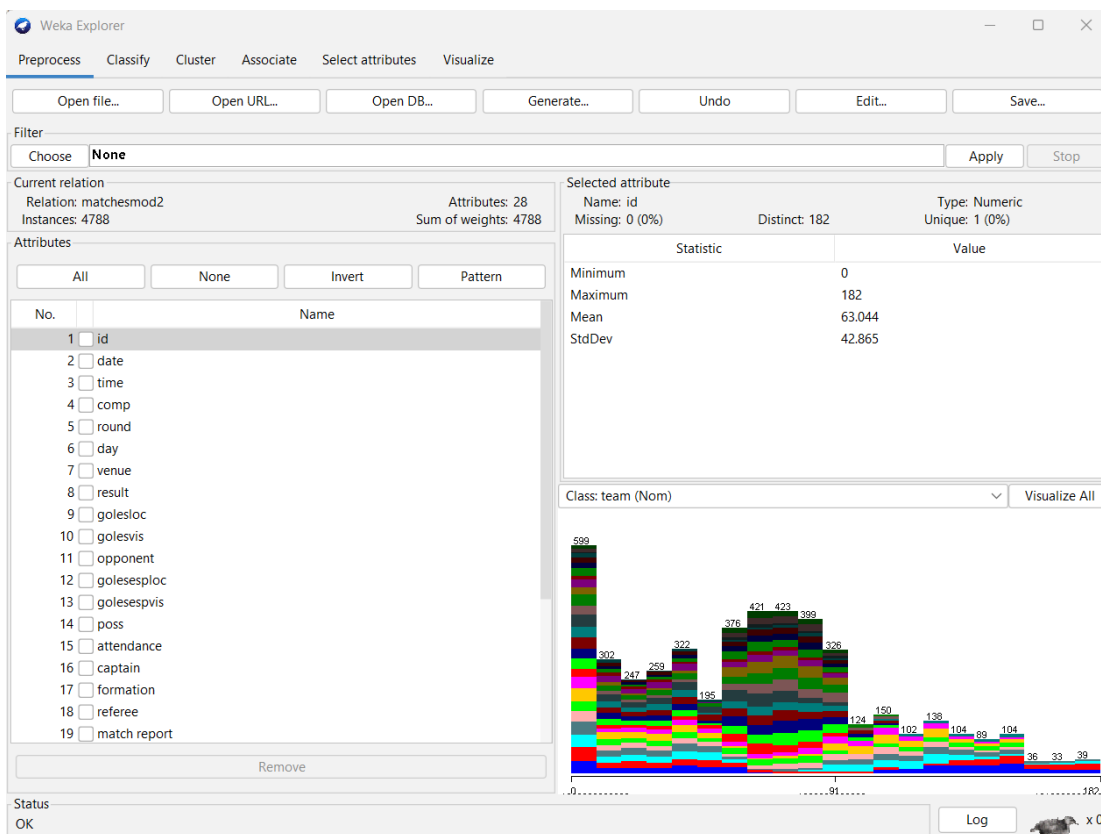
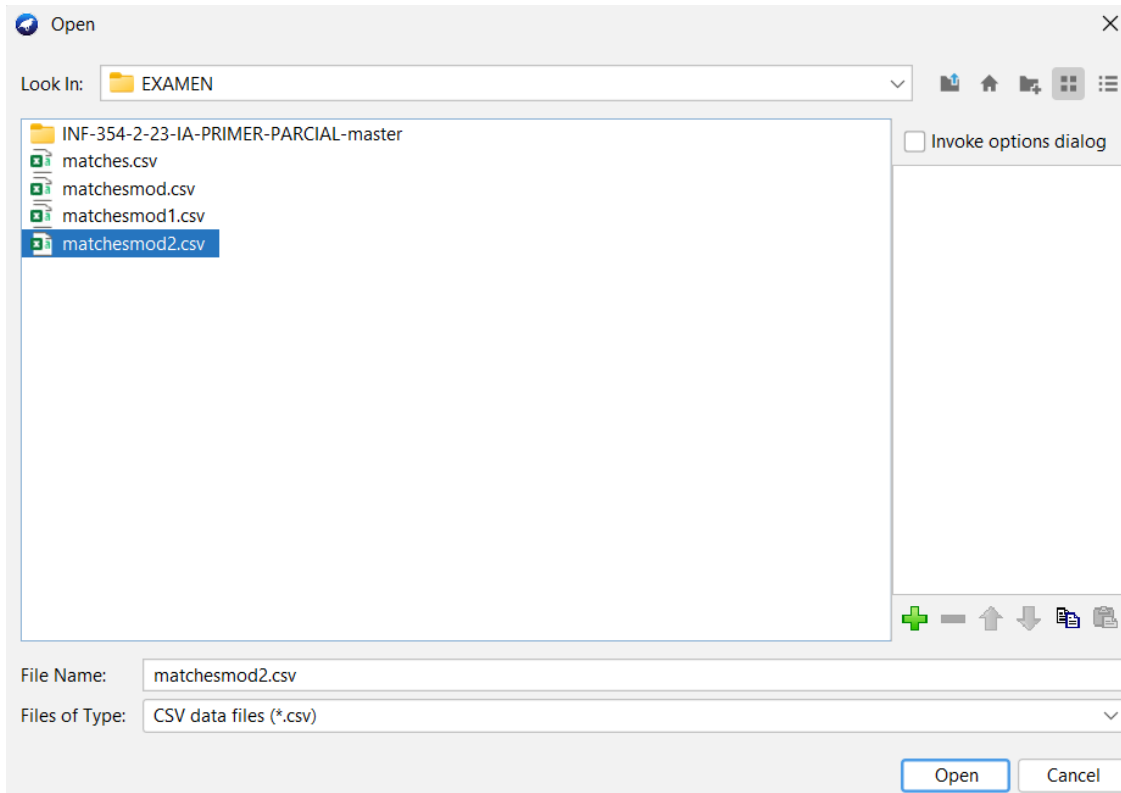


3. Para el preprocesamiento debe usted migrar su dataset de csv u otro formato a arff, una vez realizado ello. Realice un etiqueta onehotencoder, labelencoder, discretización y normalización.



OneHotEncoder :

- Ve a la pestaña “Preprocess”.
- Selecciona el atributo categórico y elige el filtro unsupervised.attribute.NominalToBinary

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter
Choose **NumericToBinary** -R first-last Apply Stop

Current relation
Relation: matchesmod2
Instances: 4788
Attributes: 28
Sum of weights: 4788

Attributes
All None Invert Pattern

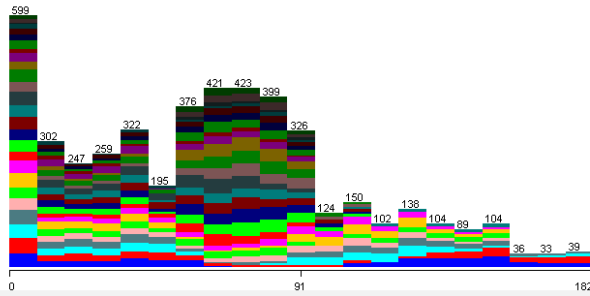
| No. | Name |
|---------------------------------------|--------------|
| 1 <input checked="" type="checkbox"/> | id |
| 2 <input type="checkbox"/> | date |
| 3 <input type="checkbox"/> | time |
| 4 <input type="checkbox"/> | comp |
| 5 <input type="checkbox"/> | round |
| 6 <input type="checkbox"/> | day |
| 7 <input type="checkbox"/> | venue |
| 8 <input type="checkbox"/> | result |
| 9 <input type="checkbox"/> | golesloc |
| 10 <input type="checkbox"/> | golesvis |
| 11 <input type="checkbox"/> | opponent |
| 12 <input type="checkbox"/> | golesesploc |
| 13 <input type="checkbox"/> | golesespvis |
| 14 <input type="checkbox"/> | poss |
| 15 <input type="checkbox"/> | attendance |
| 16 <input type="checkbox"/> | captain |
| 17 <input type="checkbox"/> | formation |
| 18 <input type="checkbox"/> | referee |
| 19 <input type="checkbox"/> | match report |

Remove

Selected attribute
Name: id
Missing: 0 (0%)
Distinct: 182
Type: Numeric
Unique: 1 (0%)

| Statistic | Value |
|-----------|--------|
| Minimum | 0 |
| Maximum | 182 |
| Mean | 63.044 |
| StdDev | 42.865 |

Class: team (Nom) Visualize All



Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter
Choose **NumericToBinary** -R first-last Apply Stop

Current relation
Relation: matchesmod2-weka.filters.unsupervised.attribute.Num...
Instances: 4788
Attributes: 28
Sum of weights: 4788

Attributes
All None Invert Pattern


| No. | Name |
|-----------------------------|--------------------------|
| 10 <input type="checkbox"/> | golesvis_binarized |
| 11 <input type="checkbox"/> | opponent |
| 12 <input type="checkbox"/> | golesesploc_binarized |
| 13 <input type="checkbox"/> | golesespvis_binarized |
| 14 <input type="checkbox"/> | poss_binarized |
| 15 <input type="checkbox"/> | attendance_binarized |
| 16 <input type="checkbox"/> | captain |
| 17 <input type="checkbox"/> | formation |
| 18 <input type="checkbox"/> | referee |
| 19 <input type="checkbox"/> | match report |
| 20 <input type="checkbox"/> | notes |
| 21 <input type="checkbox"/> | tirosloc_binarized |
| 22 <input type="checkbox"/> | tirosalarco_binarized |
| 23 <input type="checkbox"/> | distpromptir_binarized |
| 24 <input type="checkbox"/> | tiroslibloc_binarized |
| 25 <input type="checkbox"/> | tirospenloc_binarized |
| 26 <input type="checkbox"/> | tirospenintloc_binarized |
| 27 <input type="checkbox"/> | season_binarized |
| 28 <input type="checkbox"/> | team |

Remove

Selected attribute
Name: id_binarized
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | 0 | 106 | 106 |
| 2 | 1 | 4682 | 4682 |

Class: team (Nom) Visualize All



Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayesMultinomialText** -P 0 -M 3.0 -norm 1.0 -lnorm 2.0 -stopwords-handler weka.core.stopwords.Null -tokenizer "weka.core.tokenizers.WordTokenizer -delimiters \"\

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) team

Result list (right-click for options)

01:10:37 - bayes.NaiveBayesMultinomialText

01:10:44 - bayes.NaiveBayesMultinomialText

Classifier output

```

=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayesMultinomialText -P 0 -M 3.0 -norm 1.0 -lnorm 2.0 -stop
Relation:    matchesmod2-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last
Instances:   4788
Attributes:  28
             id_binarized
             date
             time
             comp
             round
             day
             venue
             result
             golesloc_binarized
             golesvis_binarized
             opponent
             golesesploc_binarized
             golesespvis_binarized
             poss_binarized
             attendance_binarized
             captain
             formation
             referee
             match report
             notes
             tirosloc_binarized
             tirosalarco_binarized
             distpromptir_binarized
             tiroslibloc_binarized
             tirospenloc_binarized
             tirospenintloc_binarized
             season_binarized

```

weka.gui.GenericObjectEditor

weka.classifiers.bayes.NaiveBayesMultinomialText

About

Multinomial naive bayes for text data.

LNorm

batchSize

debug

doNotCheckCapabilities

lowercaseTokens

minWordFrequency

norm

normalizeDocLength

numDecimalPlaces

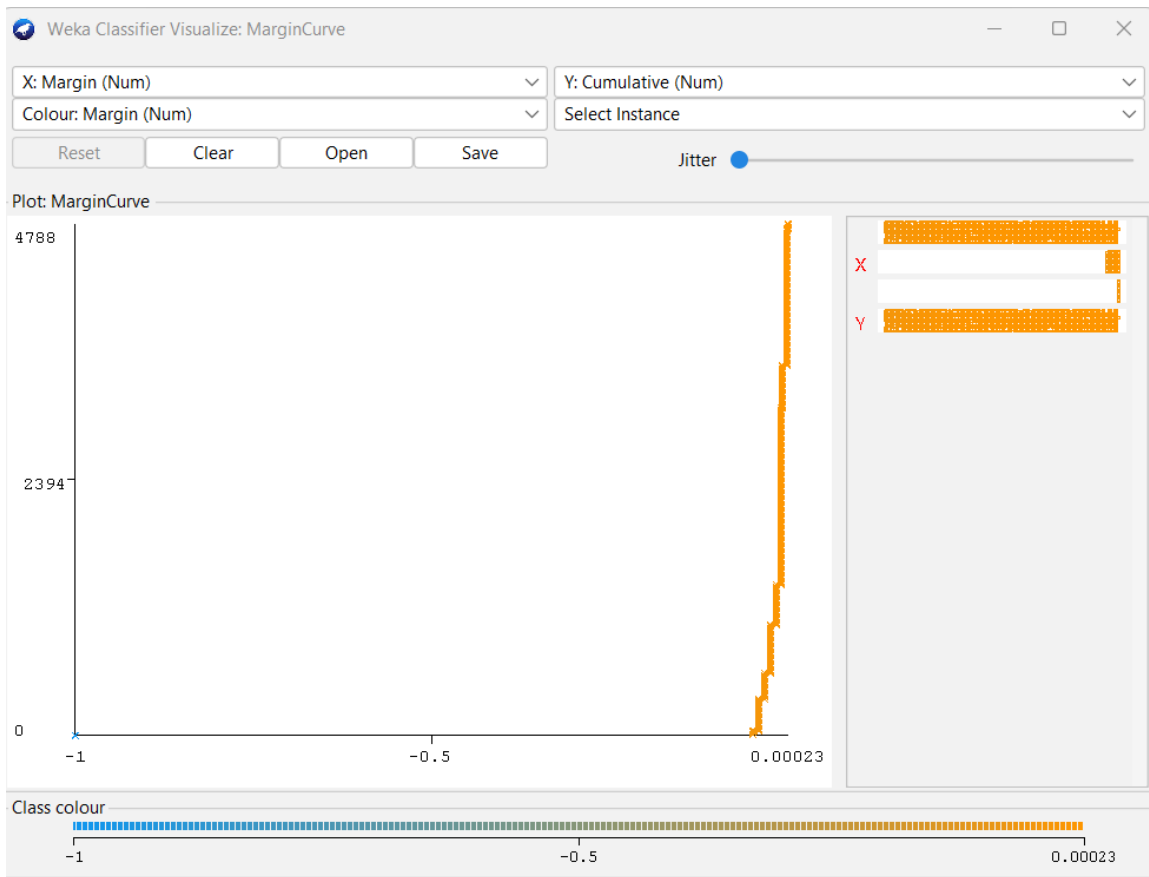
periodicPruning

stemmer **NullStemmer**

stopwordsHandler **Null**

tokenizer **WordTokenizer** -delimiters "\"\r\n\t,\";\'\\"{}?!"

useWordFrequencies



LabelEncoder:

- Aplica `unsupervised.attribute.StringToNominal` para convertir cadenas en números.

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter Choose **StringToNominal -R last** Apply Stop

Current relation
 Relation: matchesmod2-weka.filters.unsupervised.attribute.Num... Attributes: 28
 Instances: 4788 Sum of weights: 4788

Attributes

All None Invert Pattern

| No. | Name |
|-----|--|
| 1 | <input checked="" type="checkbox"/> id_binarized |
| 2 | <input type="checkbox"/> date |
| 3 | <input type="checkbox"/> time |
| 4 | <input type="checkbox"/> comp |
| 5 | <input type="checkbox"/> round |
| 6 | <input type="checkbox"/> day |
| 7 | <input type="checkbox"/> venue |
| 8 | <input type="checkbox"/> result |
| 9 | <input type="checkbox"/> golesloc_binarized |
| 10 | <input type="checkbox"/> golesvis_binarized |
| 11 | <input type="checkbox"/> opponent |
| 12 | <input type="checkbox"/> golesesploc_binarized |
| 13 | <input type="checkbox"/> golesespvis_binarized |
| 14 | <input type="checkbox"/> poss_binarized |
| 15 | <input type="checkbox"/> attendance_binarized |
| 16 | <input type="checkbox"/> captain |
| 17 | <input type="checkbox"/> formation |
| 18 | <input type="checkbox"/> referee |
| 19 | <input type="checkbox"/> match report |

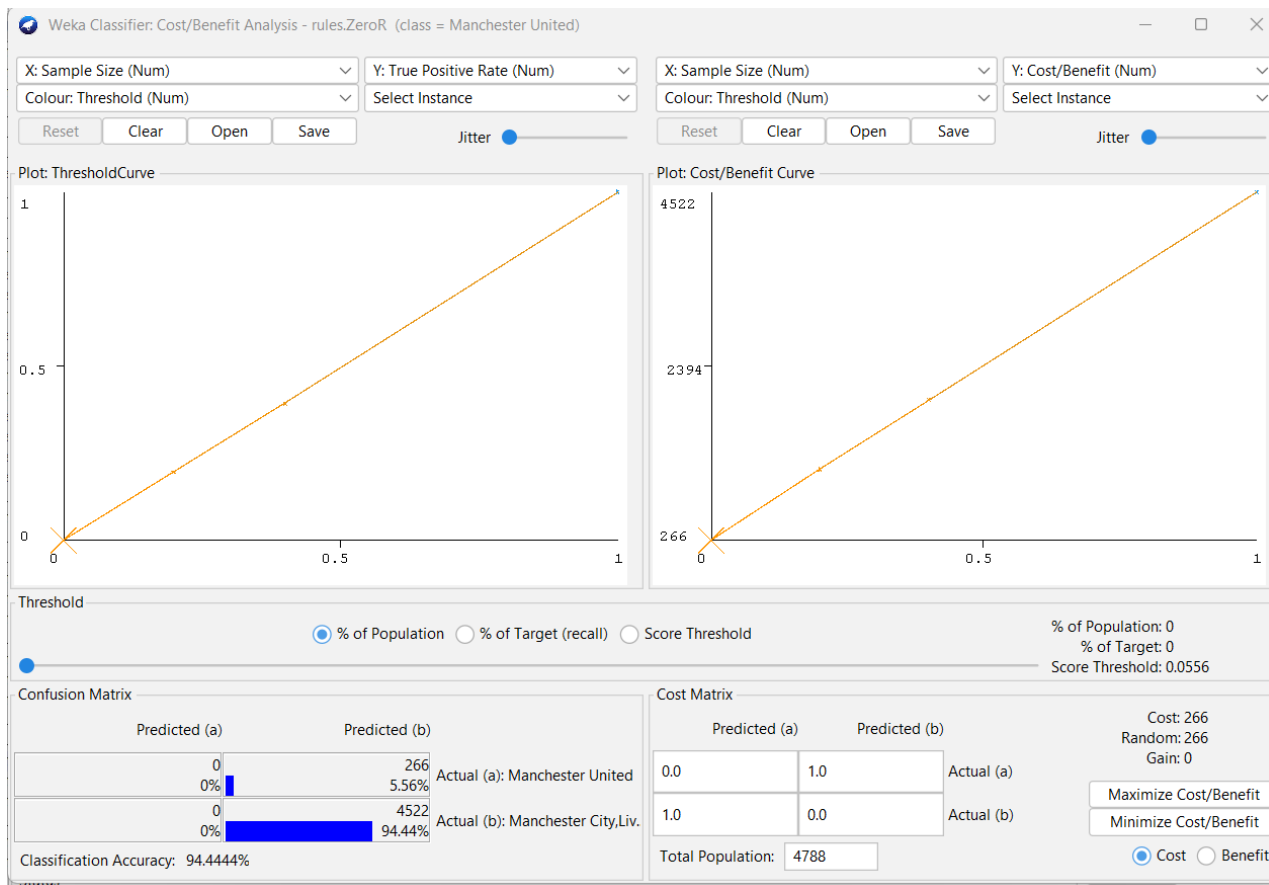
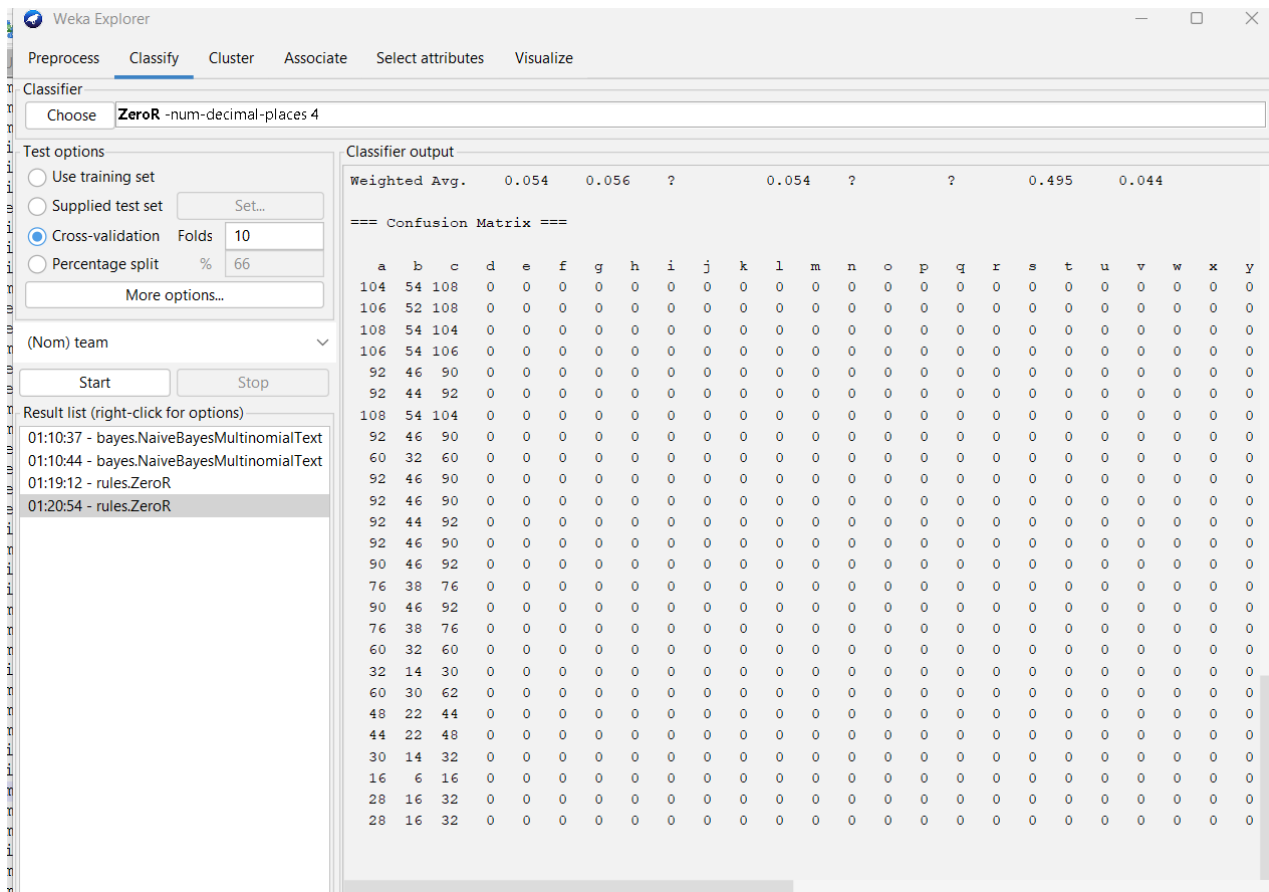
Remove

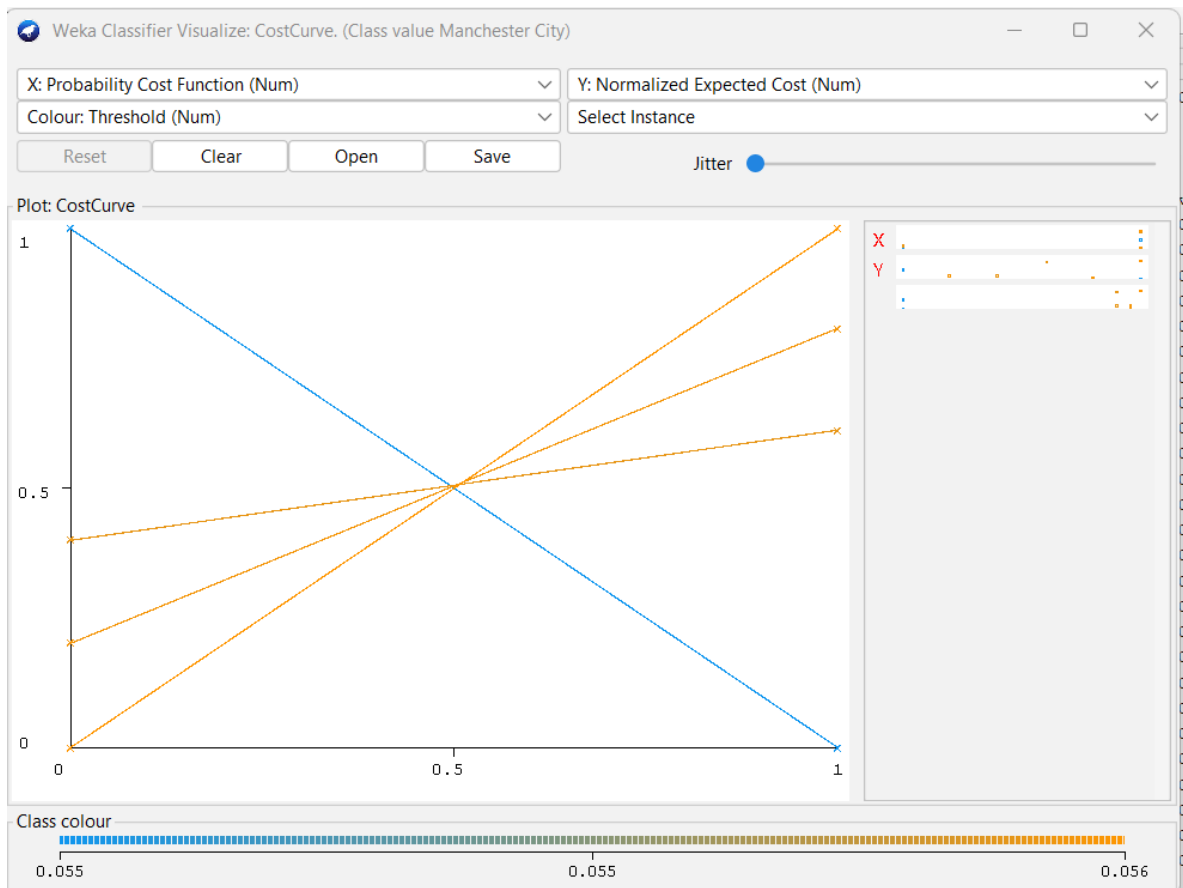
Selected attribute
 Name: id_binarized
 Missing: 0 (0%) Distinct: 2 Type: Nominal
 Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | 0 | 106 | 106 |
| 2 | 1 | 4682 | 4682 |

Class: team (Nom) Visualize All

Status OK Log x 0





Discretización:

- Utiliza `unsupervised.attribute.Discretize`.

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter Choose **Discretize** -B 10 -M -1.0 -R first-last -precision 6 Apply Stop

Current relation
Relation: matchesmod2-weka.filters.unsupervised.attribute.Num... Attributes: 28
Instances: 4788 Sum of weights: 4788

Selected attribute
Name: id_binarized Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | 0 | 106 | 106 |
| 2 | 1 | 4682 | 4682 |

Class: team (Nom) Visualize All

4682

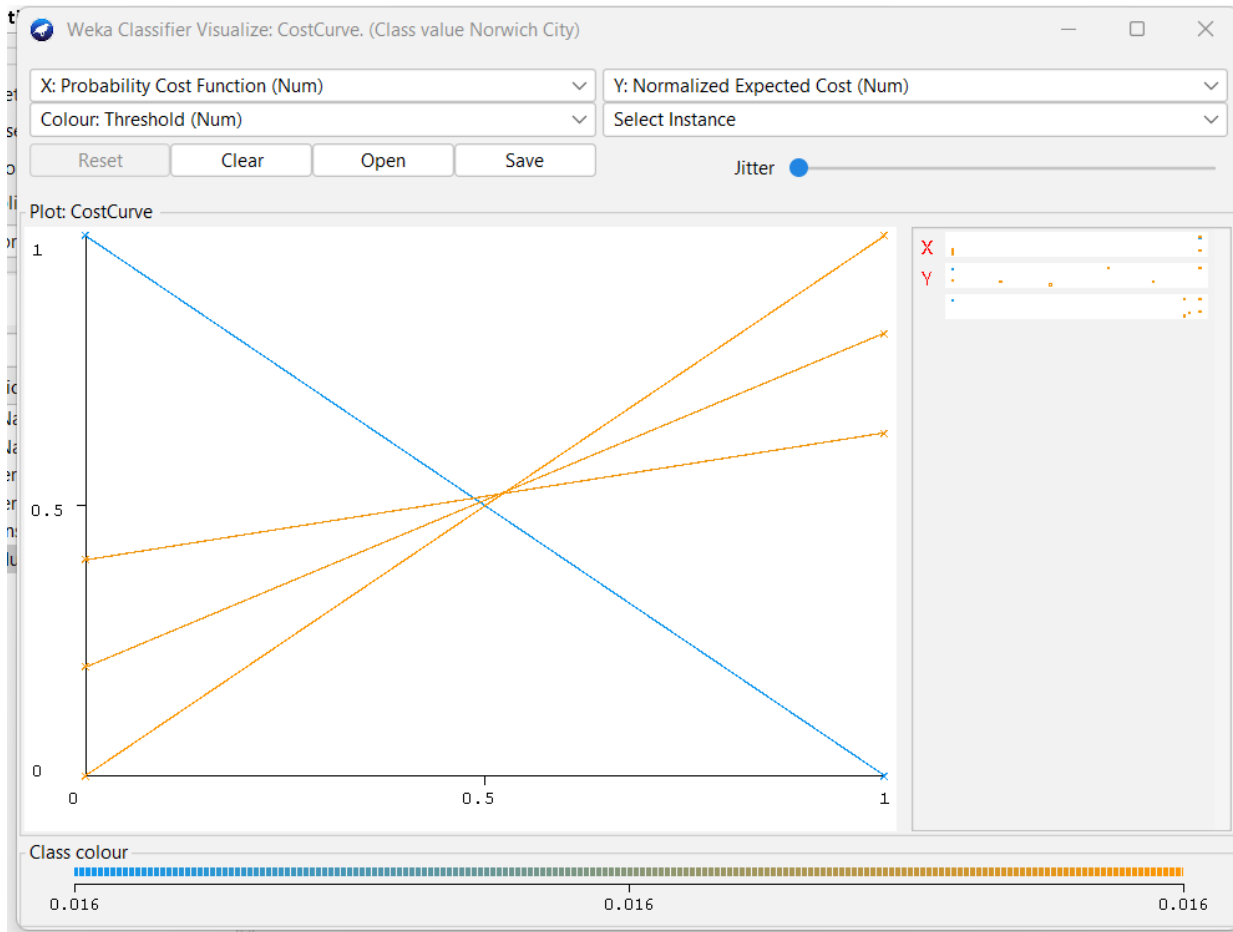
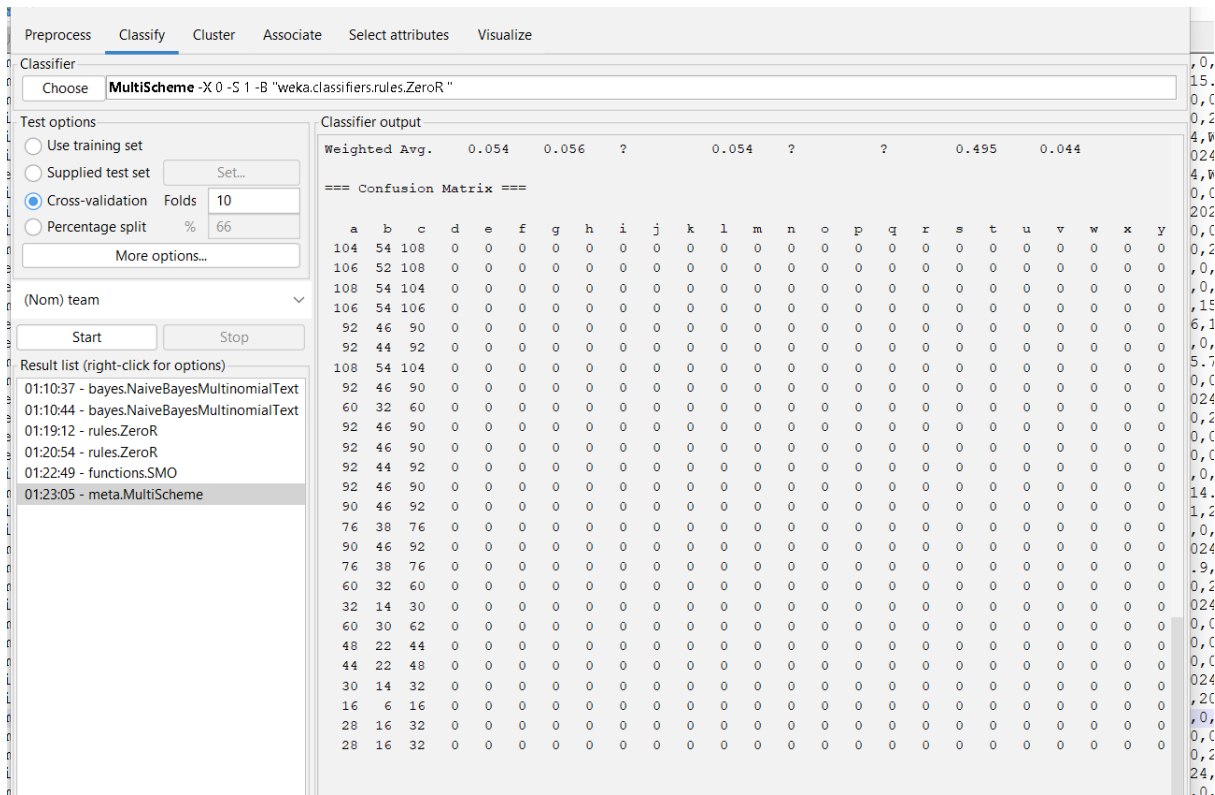
106

Attributes

All None Invert Pattern

| No. | Name |
|-----|--|
| 1 | <input checked="" type="checkbox"/> id_binarized |
| 2 | <input type="checkbox"/> date |
| 3 | <input type="checkbox"/> time |
| 4 | <input type="checkbox"/> comp |
| 5 | <input type="checkbox"/> round |
| 6 | <input type="checkbox"/> day |
| 7 | <input type="checkbox"/> venue |
| 8 | <input type="checkbox"/> result |
| 9 | <input type="checkbox"/> golesloc_binarized |
| 10 | <input type="checkbox"/> golesvis_binarized |
| 11 | <input type="checkbox"/> opponent |
| 12 | <input type="checkbox"/> golesesploc_binarized |
| 13 | <input type="checkbox"/> golesesvis_binarized |
| 14 | <input type="checkbox"/> poss_binarized |
| 15 | <input type="checkbox"/> attendance_binarized |
| 16 | <input type="checkbox"/> captain |
| 17 | <input type="checkbox"/> formation |
| 18 | <input type="checkbox"/> referee |
| 19 | <input type="checkbox"/> match report |

Remove



Normalización:

- Usa el filtro `unsupervised.attribute.Normalize`.

X: team (Nom)

Y: predicted team (Nom)

Colour: team (Nom)

Select Instance

Reset

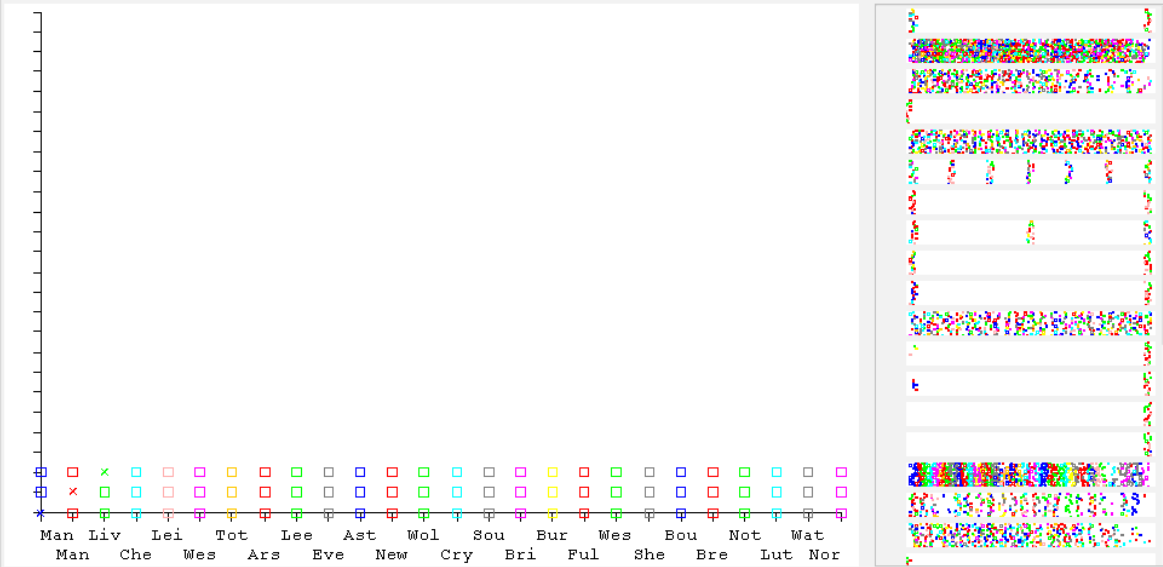
Clear

Open

Save

Jitter

Plot: matchesmod2-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last-weka.filters.unsupervised.attribute.StringToNominal-Rlast-w...



Class colour

Mancheste Mancheste Liverpool Chelsea Leicester West Ham Tottenham Arsenal Leeds Uni Everton Aston Vill Newcastle
Wolverham Crystal P Southampto Brighton Burnley Fulham West Brom Sheffield Bournemouth Brentford Nottinoha Luton Town