

CSE 634 - DATA MINING

PROJECT REPORT

Aditi Ghotikar - 111353349

Komal Gyanani - 111491719

Nirvik Ghosh - 111461888

Sweta Kumari - 111497926

Swetha Tatavarthy - 111326521

Classification of Bakery Data using WEKA

INTRODUCTION

The goal of this project is to use a classification tool - WEKA - to build two types of classifiers namely, a descriptive and a non-descriptive classifier and compare their results.

DATA MINING PROCESS STEPS

Step 1: Data Preparation -

Preparing the dataset: The .xls format of the dataset was converted to .csv, for WEKA to recognize the data. Also, blank sheets in the .xls dataset were deleted.

We corrected a few misspellings found in the naming of the classes - for example, Pyrate was misspelled as Pyrrite in a row. Since we need 6 classes, R. Carbonatees and R. Carbonatees impures were clubbed to a single class C1.

Now, there are 6 classes in the given dataset:

C1 - R. Carbonatees and R. Carbonatees impures

C2 : Pyrate

C3 : Charcopyrite

C4 : Galene

C5 : Spahlerite

C6 : Sediments terrigenes

Cleaning the data: We observed that the class attribute *Type de Roche* occurred twice in the given dataset (at the second and again at the last column). WEKA throws an error warning of duplicated attributes. Hence, we deleted the last column. The second row of the dataset, which was blank, was also deleted as WEKA considers even blank rows as attributes.

Additionally, one value for the Li attribute read '<0.3'. Since we cannot accept special characters, we filled in a value of 0.29 for this attribute.

Filling missing values: We observed that some attributes had <85% of their values missing, while others had >85%.

Two attributes - Mo and Co - had >85% of their values missing. Hence, we ignored these attributes during the classification. For others which had about <85% of their values

missing, we filled in the missing values by replacing them with the mean of the values over that specific attribute values using “ReplaceMissingValues” filter in Weka.

Motivation for the above steps: During the KDD process, we need noiseless and consistent data, which was the motivation to perform the above data preparation steps.

Result of this step: The clean data was loaded into WEKA and was ready was data preprocessing.

Step 2: Data Preprocessing -

Descriptive classifier: To build a descriptive classifier - a decision tree - we used two techniques of data discretization. Discretization methods reduce the number of values for continuous attributes. This is done by dividing the attribute range into intervals. This makes it easier for the learning algorithm (decision tree, in this case) to spot patterns in the data thereby making the result more accurate and faster.

Non-Descriptive classifier: For a Neural Network non-descriptive classifier, we first normalized the data using the WEKA ‘Normalize’ preprocess filter.

Step 3: Building the classifier:

For a descriptive classifier, we have used 2 types of data discretization methods:

Discretization Method 1: Equal Width Binning: It is a discretization algorithm which discretizes values by creating bins of equal width intervals.

Discretization Method 2: Equal Frequency Binning: It is a discretization algorithm which discretizes values by creating bins where an equal number of values are assigned to each bin.

Additionally, while trying out various methods in WEKA, we found a Supervised Discretization method and observed that it fares better than the above two unsupervised discretization methods. We have also documented the results for the same. It is an algorithm which discretizes values based on classes. This ensures that it implicitly uses information gain on parameters to classify them into bins.

We have used cross validation with 10 folds method in Weka to train the model on training and validation data.

To implement the discretization, we used meta classifier to ensure that the discretization is only applied on the test data and not the data being used for validation in the 10 folds cross validation being used to train the model.

Here are the results for **decision tree descriptive classifier** discretization methods:

a. Equal width binning discretization method

| |
|--|
| Classifier Model J48 pruned tree |
| <pre> K2O_1 = '(-inf-0.1]' Cu_1 = '(-inf-0.1]' CaO+MgO_3 = '(-inf-0.3]': Galene (2.0/1.0) CaO+MgO_3 = '(0.3-inf)' Cd_1 = '(-inf-0.1]': R. Carbonatees and R. Carbonatees impures (83.0/6.0) Cd_1 = '(0.1-inf)': Spahlerite (3.0/1.0) Cu_1 = '(0.1-inf)': Charcopyrite (2.0) K2O_1 = '(0.1-inf)': Sediments terrigenes (8.0) </pre> |
| Number of Leaves : 5 Size of the tree : 9 |

Discriminant rules:

Rule 1: If the K2O and Cu attributes are both less than 0.1 and CaO+MgO is less than 0.3, then the type of rock is Galene.

Rule 2: If the K2O and Cu attributes are both less than 0.1, CaO+MgO is more than 0.3 and Cd is less than 0.1, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 3: If the K2O and Cu attributes are both less than 0.1, CaO+MgO is more than 0.3 and Cd is more than 0.1, then the type of rock is Spahlerite.

Rule 4: If K2O is less than 0.1 and Cu is more than 0.1, then the type of rock is Charcopyrite.

Rule 5: If K2O is more than 0.1, then the type of rock is Sediments terrigenes.

Classifier output

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

Summary ===

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 83 | 84.6939 % |
| Incorrectly Classified Instances | 15 | 15.3061 % |
| Kappa statistic | 0.4926 | |
| Mean absolute error | 0.0665 | |
| Root mean squared error | 0.2158 | |
| Relative absolute error | 49.7643 % | |
| Root relative squared error | 86.3974 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|------------------------------------|---------|---------|-----------|--------|-----------|--------|----------|----------|---|
| 17:50:39 - meta.FilteredClassifier | 0.987 | 0.476 | 0.884 | 0.987 | 0.933 | 0.639 | 0.758 | 0.896 | R. Carbonatees and R. Carbonatees impures |
| 17:51:05 - meta.FilteredClassifier | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | -0.021 | 0.547 | 0.057 | Pyrate |
| 17:52:06 - meta.FilteredClassifier | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.500 | 0.020 | Charcopyrite |
| 17:55:26 - meta.FilteredClassifier | 0.333 | 0.011 | 0.500 | 0.333 | 0.400 | 0.393 | 0.704 | 0.399 | Galene |
| 17:55:52 - meta.FilteredClassifier | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | -0.018 | 0.207 | 0.025 | Spahlerite |
| 17:56:12 - meta.FilteredClassifier | 0.667 | 0.022 | 0.750 | 0.667 | 0.706 | 0.679 | 0.799 | 0.539 | Sediments terrigenes |
| Weighted Avg. | 0.847 | 0.377 | ? | 0.847 | ? | ? | 0.729 | 0.769 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | ← classified as |
|----|---|---|---|---|---|---|---|
| 76 | 0 | 0 | 0 | 1 | 1 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 1 | b = Pyrate |
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | c = Charcopyrite |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | d = Galene |
| 2 | 0 | 0 | 1 | 0 | 0 | 1 | e = Spahlerite |
| 3 | 0 | 0 | 0 | 0 | 6 | 1 | f = Sediments terrigenes |

b. Equal frequency discretization method:

| |
|---|
| Classifier Model J48 pruned tree |
| Pb_6 = '(-inf-0.053537]' Cs_6 = '(-inf-0.189189]' Fe2O3*_9 = '(-inf-0.06657]' S_9 = '(-inf-0.041379]' Cd_4 = '(-inf-0.094552]': R. Carbonatees and R. Carbonatees impures (80.0/3.0) Cd_4 = '(0.094552-inf)': Spahlerite (2.0) S_9 = '(0.041379-inf)': Pyrate (2.0) Fe2O3*_9 = '(0.06657-inf)': Charcopyrite (3.0/1.0) Cs_6 = '(0.189189-inf)': Sediments terrigenes (8.0) Pb_6 = '(0.053537-inf)': Galene (3.0) |
| Number of Leaves : 6 Size of the tree : 11 |

Discriminant rules:

Rule 1: If Pb is less than 0.0535, Cs is less than 0.189 and Fe2O3 is less than 0.066, S is less than 0.0414, Cd is less than 0.095, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 2: If Pb is less than 0.0535, Cs is less than 0.189 and Fe2O3 is less than 0.066, S is less than 0.0414, Cd is more than 0.095, then the type of rock is Spahlerite.

Rule 3: If Pb is less than 0.0535, Cs is less than 0.189 and Fe2O3 is less than 0.066, S is more than 0.0414, then the type of rock is Pyrate.

Rule 4: If Pb is less than 0.0535, Cs is less than 0.189 and Fe2O3 is more than 0.066, then the type of rock is Charcopyrite.

Rule 5: If Pb is less than 0.0535, Cs is more than 0.189, then the type of rock is Sediments terrigenes.

Rule 6: If Pb is less than 0.0535, Cs is more than 0.189, then the type of rock is Galene.

PreprocessClassifyClusterAssociateSelect attributesVisualize

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds10

Percentage split

%66

More options...

(Nom) Type de roche

StartStop

Result list (right-click for options)

17:50:39 - meta.FilteredClassifier

17:51:05 - meta.FilteredClassifier

17:52:06 - meta.FilteredClassifier

17:55:26 - meta.FilteredClassifier

17:55:52 - meta.FilteredClassifier

17:56:12 - meta.FilteredClassifier

Classifier output

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

8687.7551 %

Incorrectly Classified Instances

1212.2449 %

Kappa statistic

0.6313

Mean absolute error

0.0476

Root mean squared error

0.1977

Relative absolute error

35.5977 %

Root relative squared error

79.158 %

Total Number of Instances

98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|--------|----------|----------|---|
| | 0.974 | 0.333 | 0.915 | 0.974 | 0.943 | 0.711 | 0.753 | 0.861 | R. Carbonatees and R. Carbonatees impures |
| | 0.500 | 0.021 | 0.500 | 0.500 | 0.500 | 0.479 | 0.598 | 0.354 | Pyrate |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.703 | 0.510 | Charcopyrite |
| | 0.667 | 0.011 | 0.667 | 0.667 | 0.667 | 0.656 | 0.828 | 0.455 | Galene |
| | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | -0.018 | 0.242 | 0.031 | Spahlerite |
| | 0.778 | 0.011 | 0.875 | 0.778 | 0.824 | 0.809 | 0.823 | 0.701 | Sediments terrigenes |
| Weighted Avg. | 0.878 | 0.264 | ? | 0.878 | ? | ? | 0.738 | 0.781 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 75 | 0 | 0 | 0 | 1 | 1 | | a = R. Carbonatees and R. Carbonatees impures |
| 2 | 2 | 0 | 0 | 0 | 0 | | b = Pyrate |
| 0 | 2 | 0 | 0 | 0 | 0 | | c = Charcopyrite |
| 1 | 0 | 0 | 2 | 0 | 0 | | d = Galene |
| 2 | 0 | 0 | 1 | 0 | 0 | | e = Spahlerite |
| 2 | 0 | 0 | 0 | 0 | 7 | | f = Sediments terrigenes |

Status

OK

Log

x 0

c. Supervised Discretization - We also performed a supervised discretization in addition to the above 2 data discretization methods and found that we obtain a better accuracy (87.8%) here:

| |
|---|
| Classifier Model J48 pruned tree |
| <pre> K2O = '(-inf-0.084975]' Pb = '(-inf-0.038576]' Sc = '(-inf-0.107798]' Fe2O3* = '(-inf-0.010288]': R. Carbonatees and R. Carbonatees impures (74.0/1.0) Fe2O3* = '(0.010288-inf)' S = '(-inf-0.00611]': R. Carbonatees and R. Carbonatees impures (4.0/1.0) S = '(0.00611-inf)': Pyrate (4.0) Sc = '(0.107798-inf)': Charcopyrite (3.0/1.0) Pb = '(0.038576-0.053537]': Spahlerite (1.0) Pb = '(0.053537-inf)': Galene (3.0) K2O = '(0.084975-inf)': Sediments terrigenes (9.0) </pre> |
| Number of Leaves : 7 Size of the tree : 12 |

Discriminant rules:

Rule 1: If K2O is less than 0.085, Pb is less than 0.0386 and Sc is less than 0.1078, Fe2O3 is less than 0.01029, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 2: If K2O is less than 0.085, Pb is less than 0.0386 and Sc is less than 0.1078, Fe2O3 is more than 0.01029, S is less than 0.00611, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 3: If K2O is less than 0.085, Pb is less than 0.0386 and Sc is less than 0.1078, Fe2O3 is more than 0.01029, S is more than 0.00611, then the type of rock is Pyrate.

Rule 4: If K2O is less than 0.085, Pb is less than 0.0386 and Sc is more than 0.1078, then the type of rock is Charcopyrite.

Rule 5: If K2O is less than 0.085, Pb is between than 0.038576 and 0.053537, then the type of rock is Spahlerite.

Rule 6: If K2O is less than 0.085, Pb is more than 0.053537, then the type of rock is Galene.

Rule 7: If K2O is more than 0.085, then the type of rock is Sediments terrigenes.

PreprocessClassifyClusterAssociateSelect attributesVisualize

Weka Explorer

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds10

Percentage split

%66

More options...

(Nom) Type de roche

StartStop

Result list (right-click for options)

17:50:39 - meta.FilteredClassifier

17:51:05 - meta.FilteredClassifier

17:52:06 - meta.FilteredClassifier

17:55:26 - meta.FilteredClassifier

17:55:52 - meta.FilteredClassifier

17:56:12 - meta.FilteredClassifier

Classifier output

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances8687.7551 %

Incorrectly Classified Instances1212.2449 %

Kappa statistic0.6627

Mean absolute error0.0494

Root mean squared error0.1965

Relative absolute error37.001 %

Root relative squared error78.6448 %

Total Number of Instances98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|--------|----------|----------|---|
| | 0.961 | 0.190 | 0.949 | 0.961 | 0.955 | 0.785 | 0.889 | 0.931 | R. Carbonatees and R. Carbonatees impures |
| | 0.750 | 0.032 | 0.500 | 0.750 | 0.600 | 0.593 | 0.957 | 0.392 | Pyrate |
| | 0.000 | 0.021 | 0.000 | 0.000 | 0.000 | -0.021 | 0.948 | 0.286 | Charcopyrite |
| | 0.667 | 0.011 | 0.667 | 0.667 | 0.667 | 0.656 | 0.828 | 0.455 | Galene |
| | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | -0.018 | 0.147 | 0.031 | Spahlerite |
| | 0.778 | 0.011 | 0.875 | 0.778 | 0.824 | 0.809 | 0.882 | 0.678 | Sediments terrigenes |
| Weighted Avg. | 0.878 | 0.153 | 0.867 | 0.878 | 0.871 | 0.734 | 0.868 | 0.830 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 74 | 0 | 2 | 0 | 0 | 1 | 1 | a = R. Carbonatees and R. Carbonatees impures |
| 1 | 3 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 2 | 0 | 0 | 0 | 0 | 1 | c = Charcopyrite |
| 0 | 0 | 0 | 2 | 1 | 0 | 0 | d = Galene |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | e = Spahlerite |
| 1 | 1 | 0 | 0 | 0 | 7 | 1 | f = Sediments terrigenes |

Status

OK

Log

x 0

For a **non-descriptive classifier**, we have used a MultiLayer Perceptron Neural Network. We varied the hyperparameters for the network and observed the following accuracies:

Case 1:

| |
|--|
| Learning rate-0.3 |
| Momentum - 0.2 |
| Epoch-500 |
| Hidden layers- a (average of input and output) |
| Percentage split- 80% |
| Accuracy - 95% |

Weka Explorer

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

Classifier

Choose: **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

☐ Use training set
☐ Supplied test set
☐ Cross-validation Folds: 10
☒ Percentage split %: 80

More options...

(Nom) Type de roche

Start Stop

Result list (right-click for options)

16:56:13 - functions.MultilayerPerceptron

Classifier output

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

| | | | |
|----------------------------------|-----------|----|---|
| Correctly Classified Instances | 19 | 95 | % |
| Incorrectly Classified Instances | 1 | 5 | % |
| Kappa statistic | 0.7778 | | |
| Mean absolute error | 0.0192 | | |
| Root mean squared error | 0.1285 | | |
| Relative absolute error | 15.5622 % | | |
| Root relative squared error | 58.842 % | | |
| Total Number of Instances | 20 | | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|---|-------|
| 1.000 | 0.333 | 0.944 | 1.000 | 0.971 | 0.793 | 0.084 | 0.954 | R. Carbonatees and R. Carbonatees impures | |
| 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.368 | 0.077 | Pyrate | |
| ? | 0.000 | ? | ? | ? | ? | ? | ? | Charcopyrite | |
| ? | 0.000 | ? | ? | ? | ? | ? | ? | Galene | |
| ? | 0.000 | ? | ? | ? | ? | ? | ? | Spahlerite | |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | Sediments terrigenes | |
| Weighted Avg. | 0.950 | 0.283 | ? | 0.950 | ? | ? | 0.082 | 0.915 | |

=== Confusion Matrix ===

| a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|
| 17 | 0 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 1 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 0 | 0 | 0 | 0 | 0 | c = Charcopyrite |
| 0 | 0 | 0 | 0 | 0 | 0 | d = Galene |
| 0 | 0 | 0 | 0 | 0 | 0 | e = Spahlerite |
| 0 | 0 | 0 | 0 | 2 | 1 | f = Sediments terrigenes |

Status

OK Log x 0

Case 2: In this case, the parameters were same as above, although we used a cross-validation algorithm of 10 folds. Accuracy - 88.755%

Classifier

Choose: **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -Y 0 -S 0 -E 20 -H a

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: **10**
- ☐ Percentage split % 66

(Nom) Type de roche

Start Stop

Result list (right-click for options)

16:57:57 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 7.97 seconds

=== Stratified cross-validation ===

=== Summary ===

| | | |
|----------------------------------|-----------|----------|
| Correctly Classified Instances | 87 | 88.755 % |
| Incorrectly Classified Instances | 11 | 11.245 % |
| Kappa statistic | 0.6369 | |
| Mean absolute error | 0.0423 | |
| Root mean squared error | 0.1879 | |
| Relative absolute error | 31.6739 % | |
| Root relative squared error | 75.2082 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 0.987 | 0.429 | 0.894 | 0.987 | 0.938 | 0.676 | 0.858 | 0.938 | R. Carbonatees and R. Carbonatees impures |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.388 | 0.048 | Pyrate |
| | 1.000 | 0.010 | 0.667 | 1.000 | 0.800 | 0.812 | 0.990 | 0.583 | Charcopyrite |
| | 0.333 | 0.000 | 1.000 | 0.333 | 0.500 | 0.571 | 0.716 | 0.679 | Galene |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.870 | 0.168 | Spahlerite |
| | 0.889 | 0.011 | 0.889 | 0.889 | 0.889 | 0.878 | 0.993 | 0.956 | Sediments terrigenes |
| Weighted Avg. | 0.888 | 0.338 | ? | 0.888 | ? | ? | 0.850 | 0.865 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 76 | 0 | 1 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | c = Charcopyrite |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | d = Galene |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | e = Spahlerite |
| 1 | 0 | 0 | 0 | 0 | 8 | 1 | f = Sediments terrigenes |

Status: OK Log x 0

Case 3: Here, we maintained the same parameters as above but used 5 hidden layers. Accuracy - 87.7551%

Classifier

Choose: **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -Y 0 -S 0 -E 20 -H a

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: **10**
- ☐ Percentage split % 66

(Nom) Type de roche

Start Stop

Result list (right-click for options)

16:57:57 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 7.97 seconds

=== Stratified cross-validation ===

=== Summary ===

| | | |
|----------------------------------|-----------|----------|
| Correctly Classified Instances | 87 | 87.755 % |
| Incorrectly Classified Instances | 11 | 12.245 % |
| Kappa statistic | 0.6369 | |
| Mean absolute error | 0.0423 | |
| Root mean squared error | 0.1879 | |
| Relative absolute error | 31.6739 % | |
| Root relative squared error | 75.2082 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 0.987 | 0.429 | 0.894 | 0.987 | 0.938 | 0.676 | 0.858 | 0.938 | R. Carbonatees and R. Carbonatees impures |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.388 | 0.048 | Pyrate |
| | 1.000 | 0.010 | 0.667 | 1.000 | 0.800 | 0.812 | 0.990 | 0.583 | Charcopyrite |
| | 0.333 | 0.000 | 1.000 | 0.333 | 0.500 | 0.571 | 0.716 | 0.679 | Galene |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.870 | 0.168 | Spahlerite |
| | 0.889 | 0.011 | 0.889 | 0.889 | 0.889 | 0.878 | 0.993 | 0.956 | Sediments terrigenes |
| Weighted Avg. | 0.888 | 0.338 | ? | 0.888 | ? | ? | 0.850 | 0.865 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 76 | 0 | 1 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | c = Charcopyrite |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | d = Galene |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | e = Spahlerite |
| 1 | 0 | 0 | 0 | 0 | 8 | 1 | f = Sediments terrigenes |

Status: OK Log x 0

Case 4: The parameters were same as above but the number of epochs were changed to 300 and 5,10 hidden layers. Accuracy - 85.7143%

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -Y 0 -S 0 -E 20 -H a

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Type de roche

Start Stop

Result list (right-click for options)

16.57.57 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 7.97 seconds

=== Stratified cross-validation ===

=== Summary ===

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 87 | 88.7755 % |
| Incorrectly Classified Instances | 11 | 11.2245 % |
| Kappa statistic | 0.6369 | |
| Mean absolute error | 0.0423 | |
| Root mean squared error | 0.1879 | |
| Relative absolute error | 31.6739 % | |
| Root relative squared error | 75.2082 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 0.987 | 0.429 | 0.894 | 0.987 | 0.938 | 0.676 | 0.858 | 0.938 | R. Carbonatees and R. Carbonatees impures |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.388 | 0.048 | Pyrate |
| | 1.000 | 0.010 | 0.667 | 1.000 | 0.800 | 0.812 | 0.990 | 0.583 | Charcopyrite |
| | 0.333 | 0.000 | 1.000 | 0.333 | 0.500 | 0.571 | 0.716 | 0.679 | Galene |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.870 | 0.168 | Spahlerite |
| | 0.889 | 0.011 | 0.889 | 0.889 | 0.889 | 0.878 | 0.993 | 0.956 | Sediments terrigenes |
| Weighted Avg. | 0.888 | 0.338 | ? | 0.888 | ? | ? | 0.850 | 0.865 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 76 | 0 | 1 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | c = Charcopyrite |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 | d = Galene |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | e = Spahlerite |
| 1 | 0 | 0 | 0 | 0 | 8 | 8 | f = Sediments terrigenes |

Status

OK Log x 0

Experiment 2: Contrast classification: Here, class C1 is contrasted with all other classes(notC1).

Results for descriptive classifier discretization:

a. Equal width binning discretization

| |
|---|
| Classifier Model J48 pruned tree |
| CaO+MgO_5 = '(-inf-0.5]': Not C1 (10.0) CaO+MgO_5 = '(0.5-inf)' As_2 = '(-inf-0.2]' Rb_1 = '(-inf-0.1]': C1 (82.0/6.0) Rb_1 = '(0.1-inf)': Not C1 (3.0/1.0) As_2 = '(0.2-inf)': Not C1 (3.0) |
| Number of Leaves : 4 Size of the tree : 7 |

Discriminant rules:

Rule 1: If CaO+MgO is less than 0.5, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 2: If CaO+MgO is more than 0.5, As is less than 0.2 and Rb is less than 0.1, then the type of rock belongs to C1, i.e R. Carbonatees and R. Carbonatees impures.

Rule 3: If CaO+MgO is more than 0.5, As is less than 0.2 and Rb is more than 0.1, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 4: If CaO+MgO is more than 0.5, As is more than 0.2, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **FilteredClassifier** -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

More options...

(Nom) Type de roche

Start Stop

Result list (right-click for options)

- 18:19:46 - meta.FilteredClassifier
- 18:19:57 - meta.FilteredClassifier
- 18:20:29 - meta.FilteredClassifier

Classifier output

```
| - Rb_1 = '(-inf-0.1]': C1 (82.0/6.0)
|   Rb_1 = '(0.1-inf)': Not C1 (3.0/1.0)
|   As_2 = '(0.2-inf)': Not C1 (3.0)

Number of Leaves : 4
Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      88      89.7959 %
Incorrectly Classified Instances    10      10.2041 %
Kappa statistic                    0.6618
Mean absolute error                 0.1452
Root mean squared error             0.3113
Relative absolute error             42.635 %
Root relative squared error         75.8217 %
Total Number of Instances          98

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0.974    0.381    0.904    0.974    0.938    0.676    0.773    0.891    C1
      0.619    0.026    0.867    0.619    0.722    0.676    0.773    0.647    Not C1
Weighted Avg.   0.898    0.385    0.896    0.898    0.891    0.676    0.773    0.839

=== Confusion Matrix ===

a b <-- classified as
75 2 | a = C1
 8 13 | b = Not C1
```

Status

OK

Log

b. Equal frequency discretization

| |
|--|
| Classifier Model J48 pruned tree |
| CaO+MgO_1 = '(-inf-0.478646]': Not C1 (10.0) CaO+MgO_1 = '(0.478646-inf)' S_9 = '(-inf-0.041379]' Pb_5 = '(-inf-0.038576]' Cs_5 = '(-inf-0.121621]' Zn_9 = '(-inf-0.013859]': C1 (73.0) Zn_9 = '(0.013859-inf)' MnO_5 = '(-inf-0.07377]': C1 (3.0) MnO_5 = '(0.07377-inf)': Not C1 (2.0) Cs_5 = '(0.121621-inf)': Not C1 (3.0/1.0) Pb_5 = '(0.038576-inf)': Not C1 (2.0) S_9 = '(0.041379-inf)': Not C1 (5.0) |
| Number of Leaves : 7 Size of the tree : 13 |

Discriminant rules:

Rule 1: If CaO+MgO is less than 0.478646, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 2: If CaO+MgO is more than 0.478646, S is less than 0.041379, Pb is less than 0.038576, Cs is less than 0.121621, Zn is less than 0.013859, then the type of rock belongs to C1, i.e R. Carbonatees and R. Carbonatees impures.

Rule 3: If CaO+MgO is more than 0.478646, S is less than 0.041379, Pb is less than 0.038576, Cs is less than 0.121621, Zn is more than 0.013859, MnO is less than 0.07377, then the type of rock belongs to C1, i.e R. Carbonatees and R. Carbonatees impures.

Rule 4: If CaO+MgO is more than 0.478646, S is less than 0.041379, Pb is less than 0.038576, Cs is less than 0.121621, Zn is more than 0.013859, MnO is more than 0.07377, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 5: If CaO+MgO is more than 0.478646, S is less than 0.041379, Pb is less than 0.038576, Cs is more than 0.121621, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 6: If CaO+MgO is more than 0.478646, S is less than 0.041379, Pb is more than 0.038576, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 7: Rule 4: If CaO+MgO is more than 0.478646, S is more than 0.041379, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

PreprocessClassifyClusterAssociateSelect attributesVisualize

Weka Explorer

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Folds10

%66

More options...

(Nom) Type de roche

StartStop

Result list (right-click for options)

18:19:46 - meta.FilteredClassifier

18:19:57 - meta.FilteredClassifier

18:20:29 - meta.FilteredClassifier

Classifier output

| | | Cs_5 = '(0.121621-inf)': Not C1 (3.0/1.0)

| | Pb_5 = '(0.038576-inf)': Not C1 (2.0)

| S_9 = '(0.041379-inf)': Not C1 (5.0)

Number of Leaves : 7

Size of the tree : 13

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances8889.7959 %

Incorrectly Classified Instances1010.2041 %

Kappa statistic0.697

Mean absolute error0.1097

Root mean squared error0.3113

Relative absolute error32.2068 %

Root relative squared error75.8044 %

Total Number of Instances98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|--------|
| | 0.935 | 0.238 | 0.935 | 0.935 | 0.935 | 0.697 | 0.850 | 0.930 | C1 |
| | 0.762 | 0.065 | 0.762 | 0.762 | 0.762 | 0.697 | 0.850 | 0.664 | Not C1 |
| Weighted Avg. | 0.898 | 0.201 | 0.898 | 0.898 | 0.898 | 0.697 | 0.850 | 0.873 | |

=== Confusion Matrix ===

a b <-- classified as

72 5 | a = C1

5 16 | b = Not C1

Status

OKLogx 0

c. Supervised discretization

| |
|--|
| Classifier Model J48 pruned tree |
| Fe2O3*_1 = '(-inf-0.010288]': C1 (74.0/1.0) Fe2O3*_1 = '(0.010288-inf)' CaO_2 = '(-inf-0.700343]': Not C1 (20.0/1.0) CaO_2 = '(0.700343-inf)': C1 (4.0/1.0) |
| Number of Leaves : 3 Size of the tree : 5 |

Discriminant rules:

Rule 1: If Fe2O3 is less than 0.010288, then the type of rock is C1, i.e R. Carbonatees and R. Carbonatees impures.

Rule 2: If Fe2O3 is more than 0.010288, CaO is less than 0.700343, then the type of rock is not C1, i.e not R. Carbonatees and R. Carbonatees impures.

Rule 3: If Fe2O3 is more than 0.010288, CaO is more than 0.700343, then the type of rock is C1.

The screenshot shows the Weka Explorer interface. The 'Classifier' tab is selected, and the 'FilteredClassifier' is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following information:

```
Fe2O3*_1 = '(-inf-0.010288-inf)'  
| CaO_2 = '(-inf-0.700343]': Not C1 (20.0/1.0)  
| CaO_2 = '(0.700343-inf)': C1 (4.0/1.0)
```

Number of Leaves : 3
Size of the tree : 5

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 94 | 95.9184 % |
| Incorrectly Classified Instances | 4 | 4.0816 % |
| Kappa statistic | 0.8744 | |
| Mean absolute error | 0.0663 | |
| Root mean squared error | 0.2096 | |
| Relative absolute error | 19.4561 % | |
| Root relative squared error | 51.0388 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|--------|
| | 0.987 | 0.143 | 0.962 | 0.987 | 0.974 | 0.876 | 0.862 | 0.920 | C1 |
| | 0.857 | 0.013 | 0.947 | 0.857 | 0.900 | 0.876 | 0.862 | 0.811 | Not C1 |
| Weighted Avg. | 0.959 | 0.115 | 0.959 | 0.959 | 0.958 | 0.876 | 0.862 | 0.897 | |

=== Confusion Matrix ===

| | | |
|----|----|-----------------|
| a | b | ← classified as |
| 76 | 1 | a = C1 |
| 3 | 18 | b = Not C1 |

The 'Result list' on the left shows three entries for 'meta.FilteredClassifier' with timestamps 18:19:46, 18:19:57, and 18:20:29. The 'Status' bar at the bottom shows 'OK' and 'Log' buttons.

For non-descriptive classifier- Neural network:

| |
|--|
| Learning rate-0.3 |
| Momentum - 0.2 |
| Epoch-500 |
| Hidden layers- a (average of input and output) |
| Cross validation 10 folds |
| Accuracy - 88.7755% |

The screenshot shows the Weka Explorer interface with the MultilayerPerceptron classifier selected. The 'Test options' section on the left shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section on the right displays the following results:

```
Attrib Th 0.0012024683984603087
Attrib U 0.04777213745239161
Class C1
Input
Node 0
Class Not C1
Input
Node 1

Time taken to build model: 7.28 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      87      88.7755 %
Incorrectly Classified Instances   11      11.2245 %
Kappa statistic                    0.6051
Mean absolute error                 0.1185
Root mean squared error             0.3037
Relative absolute error             34.8069 %
Root relative squared error         73.9619 %
Total Number of Instances          98

=== Detailed Accuracy By Class ===
```

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|--------|
| | 0.987 | 0.476 | 0.884 | 0.987 | 0.933 | 0.639 | 0.920 | 0.975 | C1 |
| | 0.524 | 0.013 | 0.917 | 0.524 | 0.667 | 0.639 | 0.920 | 0.827 | Not C1 |
| Weighted Avg. | 0.888 | 0.377 | 0.891 | 0.888 | 0.876 | 0.639 | 0.920 | 0.944 | |

```
=== Confusion Matrix ===
 a b  <-- classified as
76 1 | a = C1
10 11 | b = Not C1
```

The 'Result list' on the left shows the command: `17:33:25 - functions.MultilayerPerceptron`. The 'Status' bar at the bottom shows 'OK' and a 'Log' button.

Experiment 3: Classification using important attributes only: Here, we have only used the 8 attributes, important as per the experts - S, Zn, Pb, Cu, CaO+MgO, CaO, MgO, Fe₂O₃.

Descriptive classifier:

a. Equal width discretization using only important attributes

| |
|---|
| Classifier Model J48 pruned tree |
| CaO_4 = '(-inf-0.4]': Sediments terrigenes (9.0/1.0) CaO_4 = '(0.4-inf)' Cu_1 = '(-inf-0.1]' Zn_1 = '(-inf-0.1]': R. Carbonatees and R. Carbonatees impures (85.0/8.0) Zn_1 = '(0.1-inf)': Galene (2.0/1.0) Cu_1 = '(0.1-inf)': Chalcopyrite (2.0) |
| Number of Leaves : 4 Size of the tree : 7 |

Discriminant rules:

Rule 1: If CaO is less than 0.4, then the type of rock is Sediments terrigenes.

Rule 2: If CaO is more than 0.4, Cu is less than 0.1, Zn is less than 0.1, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 3: If CaO is more than 0.4, Cu is less than 0.1, Zn is more than 0.1, then the type of rock is Galene.

Rule 4: If CaO is more than 0.4, Cu is more than 0.1, then the type of rock is Chalcopyrite.

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Weka Explorer

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds10

Percentage split

%66

More options...

(Nom) Type de roche

Start

Stop

Result list (right-click for options)

17:50:39 - meta.FilteredClassifier

17:51:05 - meta.FilteredClassifier

17:52:06 - meta.FilteredClassifier

17:55:26 - meta.FilteredClassifier

17:55:52 - meta.FilteredClassifier

17:56:12 - meta.FilteredClassifier

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances8586.7347 %

Incorrectly Classified Instances1313.2653 %

Kappa statistic0.5479

Mean absolute error0.0695

Root mean squared error0.2053

Relative absolute error52.0104 %

Root relative squared error82.1994 %

Total Number of Instances98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 1.000 | 0.476 | 0.885 | 1.000 | 0.939 | 0.681 | 0.729 | 0.872 | R. Carbonatees and R. Carbonatees impures |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.330 | 0.046 | Pyrate |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.500 | 0.020 | Charcopyrite |
| | 0.333 | 0.011 | 0.500 | 0.333 | 0.400 | 0.393 | 0.391 | 0.187 | Galene |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.223 | 0.027 | Spahlerite |
| | 0.778 | 0.022 | 0.778 | 0.778 | 0.778 | 0.755 | 0.790 | 0.595 | Sediments terrigenes |
| Weighted Avg. | 0.867 | 0.377 | ? | 0.867 | ? | ? | 0.689 | 0.749 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <== classified as |
|----|---|---|---|---|---|---|---|
| 77 | 0 | 0 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | c = Charcopyrite |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | d = Galene |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | e = Spahlerite |
| 2 | 0 | 0 | 0 | 0 | 0 | 7 | f = Sediments terrigenes |

Status

OK

Log

x 0

b. Equal frequency discretization using only important attributes

| |
|--|
| Classifier Model J48 pruned tree |
| Pb_6 = '(-inf-0.053537]' CaO+MgO_1 = '(-inf-0.478646]': Sediments terrigenes (9.0/1.0) CaO+MgO_1 = '(0.478646-inf)' S_9 = '(-inf-0.041379]' Zn_9 = '(-inf-0.013859]': R. Carbonatees and R. Carbonatees impures (76.0/2.0) Zn_9 = '(0.013859-inf)' Fe2O3*_5 = '(-inf-0.004962]': R. Carbonatees and R. Carbonatees impures (2.0) Fe2O3*_5 = '(0.004962-inf)': Spahlerite (4.0/1.0) S_9 = '(0.041379-inf)': Pyrate (4.0/1.0) Pb_6 = '(0.053537-inf)': Galene (3.0) |
| Number of Leaves : 6 Size of the tree : 11 |

Discriminant rules:

Rule 1: If Pb is less than 0.053537, CaO+MgO is less than 0.478646, then the type of rock is Sediments terrigenes.

Rule 2: If Pb is less than 0.053537, CaO+MgO is more than 0.478646, S is less than 0.041379, Zn is less than 0.013859, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 3: If Pb is less than 0.053537, CaO+MgO is more than 0.478646, S is less than 0.041379, Zn is more than 0.013859, Fe2O3 is less than 0.004962, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 4: If Pb is less than 0.053537, CaO+MgO is more than 0.478646, S is less than 0.041379, Zn is more than 0.013859, Fe2O3 is more than 0.004962, then the type of rock is Spahlerite.

Rule 5: If Pb is less than 0.053537, CaO+MgO is more than 0.478646, S is more than 0.041379, then the type of rock is Pyrate .

Rule 6: If Pb is more than 0.053537, then the type of rock is Galene.

PreprocessClassifyClusterAssociateSelect attributesVisualize

Weka Explorer

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 --- -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds10

Percentage split

%66

More options...

(Nom) Type de roche

Start

Stop

Result list (right-click for options)

17:50:39 - meta.FilteredClassifier

17:51:05 - meta.FilteredClassifier

17:52:06 - meta.FilteredClassifier

17:55:26 - meta.FilteredClassifier

17:55:52 - meta.FilteredClassifier

17:56:12 - meta.FilteredClassifier

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

8889.7959 %

Incorrectly Classified Instances

1010.2041 %

Kappa statistic

0.7122

Mean absolute error

0.0501

Root mean squared error

0.1819

Relative absolute error

37.4806 %

Root relative squared error

72.8303 %

Total Number of Instances

98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|--------|----------|----------|---|
| | 0.974 | 0.190 | 0.949 | 0.974 | 0.962 | 0.813 | 0.891 | 0.935 | R. Carbonatees and R. Carbonatees impures |
| | 0.750 | 0.021 | 0.600 | 0.750 | 0.667 | 0.655 | 0.779 | 0.410 | Pyrate |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.422 | 0.020 | Charcopyrite |
| | 1.000 | 0.011 | 0.750 | 1.000 | 0.857 | 0.861 | 0.989 | 0.750 | Galene |
| | 0.000 | 0.011 | 0.000 | 0.000 | 0.000 | -0.018 | 0.663 | 0.053 | Spahlerite |
| | 0.778 | 0.022 | 0.778 | 0.778 | 0.778 | 0.755 | 0.888 | 0.717 | Sediments terrigenes |
| Weighted Avg. | 0.898 | 0.153 | ? | 0.898 | ? | ? | 0.873 | 0.842 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 75 | 0 | 0 | 0 | 1 | 1 | | a = R. Carbonatees and R. Carbonatees impures |
| 1 | 3 | 0 | 0 | 0 | 0 | | b = Pyrate |
| 0 | 1 | 0 | 0 | 0 | 1 | | c = Charcopyrite |
| 0 | 0 | 0 | 3 | 0 | 0 | | d = Galene |
| 2 | 0 | 0 | 1 | 0 | 0 | | e = Spahlerite |
| 1 | 1 | 0 | 0 | 0 | 7 | | f = Sediments terrigenes |

Status

OK

Log

x 0

c. Supervised important attributes

| |
|---|
| Classifier Model J48 pruned tree |
| Fe2O3*_1 = '(-inf-0.010288]': R. Carbonatees and R. Carbonatees impures (74.0/1.0) Fe2O3*_1 = '(0.010288-inf)' Zn_2 = '(-inf-0.040932]' CaO_1 = '(-inf-0.624637]': Sediments terrigenes (12.0/3.0) CaO_1 = '(0.624637-inf)' S_1 = '(-inf-0.00611]': R. Carbonatees and R. Carbonatees impures (4.0) S_1 = '(0.00611-inf)': Pyrate (3.0) Zn_2 = '(0.040932-inf)' Pb_2 = '(-inf-0.053537]': Spahlerite (2.0) Pb_2 = '(0.053537-inf)': Galene (3.0) |
| Number of Leaves : 6 Size of the tree : 11 |

Discriminant rules:

Rule 1: If Fe2O3 is less than 0.010288, then the type of rock is R. Carbonatees and R. Carbonatees impures..

Rule 2: If Fe2O3 is more than 0.010288, Zn is less than 0.040932, CaO is less than 0.624637, then the type of rock is Sediments terrigenes.

Rule 3: If Fe2O3 is more than 0.010288, Zn is less than 0.040932, CaO is more than 0.624637, S is less than 0.00611, then the type of rock is R. Carbonatees and R. Carbonatees impures.

Rule 4: If Fe2O3 is more than 0.010288, Zn is less than 0.040932, CaO is more than 0.624637, S is more than 0.00611, then the type of rock is Pyrate.

Rule 5: If Fe2O3 is more than 0.010288, Zn is more than 0.040932, Pb is less than 0.053537, then the type of rock is Spahlerite.

Rule 6: If Fe2O3 is more than 0.010288, Zn is more than 0.040932, Pb is more than 0.053537, then the type of rock is Galene.

PreprocessClassifyClusterAssociateSelect attributesVisualize

Weka Explorer

Classifier

ChooseFilteredClassifier -F "weka.filters.supervised.attribute.Discretize -D -R first-last -precision 6" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Folds10

Percentage split

%66

More options...

(Nom) Type de roche

Start

Stop

Result list (right-click for options)

17:50:39 - meta.FilteredClassifier

17:51:05 - meta.FilteredClassifier

17:52:06 - meta.FilteredClassifier

17:55:26 - meta.FilteredClassifier

17:55:52 - meta.FilteredClassifier

17:56:12 - meta.FilteredClassifier

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

8788.7755 %

Incorrectly Classified Instances

1111.2245 %

Kappa statistic

0.6829

Mean absolute error

0.0426

Root mean squared error

0.1844

Relative absolute error

31.8895 %

Root relative squared error

73.8165 %

Total Number of Instances

98

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | RDC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|--------|----------|----------|---|
| | 0.987 | 0.143 | 0.962 | 0.987 | 0.974 | 0.876 | 0.923 | 0.961 | R. Carbonatees and R. Carbonatees impures |
| | 0.750 | 0.011 | 0.750 | 0.750 | 0.750 | 0.739 | 0.859 | 0.573 | Pyrate |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.698 | 0.073 | Charcopyrite |
| | 0.333 | 0.021 | 0.333 | 0.333 | 0.333 | 0.312 | 0.821 | 0.344 | Galene |
| | 0.000 | 0.021 | 0.000 | 0.000 | 0.000 | -0.026 | 0.604 | 0.132 | Spahlerite |
| | 0.778 | 0.034 | 0.700 | 0.778 | 0.737 | 0.710 | 0.909 | 0.555 | Sediments terrigenes |
| Weighted Avg. | 0.888 | 0.117 | ? | 0.888 | ? | ? | 0.902 | 0.845 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 76 | 0 | 0 | 0 | 0 | 1 | | a = R. Carbonatees and R. Carbonatees impures |
| 0 | 3 | 0 | 0 | 0 | 1 | | b = Pyrate |
| 0 | 1 | 0 | 0 | 0 | 1 | | c = Charcopyrite |
| 0 | 0 | 0 | 1 | 2 | 0 | | d = Galene |
| 1 | 0 | 0 | 2 | 0 | 0 | | e = Spahlerite |
| 2 | 0 | 0 | 0 | 0 | 7 | | f = Sediments terrigenes |

Status

OK

Log

x 0

For the **non-descriptive classifier**, the neural network, we have used the following parameters:

| |
|--|
| Learning rate-0.3 |
| Momentum - 0.2 |
| Epoch-500 |
| Hidden layers- a (average of input and output) |
| Cross validation 10 folds |
| Accuracy - 89.7959% |

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Type de roche

Start Stop

Result list (right-click for options)

17:18:51 - functions.MultilayerPerceptron

17:20:55 - functions.MultilayerPerceptron

17:22:51 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 4.36 seconds

=== Stratified cross-validation ===

=== Summary ===

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 88 | 89.7959 % |
| Incorrectly Classified Instances | 10 | 10.2041 % |
| Kappa statistic | 0.6614 | |
| Mean absolute error | 0.041 | |
| Root mean squared error | 0.1821 | |
| Relative absolute error | 30.7179 % | |
| Root relative squared error | 72.9147 % | |
| Total Number of Instances | 98 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 1.000 | 0.429 | 0.895 | 1.000 | 0.945 | 0.715 | 0.925 | 0.972 | R. Carbonatees and R. Carbonatees impures |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.660 | 0.083 | Pyrate |
| | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | Charcopyrite |
| | 0.333 | 0.000 | 1.000 | 0.333 | 0.500 | 0.571 | 0.688 | 0.511 | Galene |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.923 | 0.205 | Spahlerite |
| | 0.889 | 0.011 | 0.889 | 0.889 | 0.889 | 0.878 | 0.988 | 0.942 | Sediments terrigenes |
| Weighted Avg. | 0.898 | 0.338 | ? | 0.898 | ? | ? | 0.914 | 0.896 | |

=== Confusion Matrix ===

| | a | b | c | d | e | f | <-- classified as |
|----|---|---|---|---|---|---|---|
| 77 | 0 | 0 | 0 | 0 | 0 | 0 | a = R. Carbonatees and R. Carbonatees impures |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | b = Pyrate |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | c = Charcopyrite |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | d = Galene |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | e = Spahlerite |
| 1 | 0 | 0 | 0 | 0 | 8 | 0 | f = Sediments terrigenes |

Status

OK Log x 0

CONCLUSION

We have successfully used WEKA to classify our data and found the following accuracies for the descriptive and non-descriptive classifiers:

Analysis: For Decision Trees -

| | Unsupervised Equal Width Binning | Unsupervised Equal Frequency Binning | Supervised Discretization |
|--------------|--|--|------------------------------|
| Experiment 1 | 84.6939% | 87.7551% | 87.7551% |
| Experiment 2 | 89.7959% | 89.7959% | 95.9% |
| Experiment 3 | 86.7347% | 89.7959% | 88.7755% |

For Neural Network -

| | Accuracy |
|--------------|----------|
| Experiment 1 | 88.7755% |
| Experiment 2 | 88.7755% |
| Experiment 3 | 89.7959% |