|  |  |  |
| --- | --- | --- |
| **S.No** | **TOPICS** | **PAGE NO** |
| **1** | **INTRODUCTION** | **4** |
| 2 | **PROBLEM DEFINITION** | **4** |
| 3 | **SEARCH FOR BEST MODEL** | **5** |
| **4** | **DATA PREPARATION** | **5** |
| **5** | **SELECTION OF AN ALGORITHM & PARAMETERS** | **7** |
| **6** | **DATA PRE-PROCESSING** | **8** |
| **7** | **EXPERIMENT RESULT** | **11** |
| **8** | **RESULT ANALYSIS** | **17** |
| **9** | **SUMMARY & CONCLUSION** | **17** |
| **10** | **SAMPLE SCREEN SHOTS** | **19** |

Dataset Name CPU

Source of the dataset https://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html

Dataset description The data set contains 3 classes of 50 instances each where each class refers to a type of iris plant.

No of attributes 5

No of instances 150

* 1. The inputs and output are:

|  |  |  |
| --- | --- | --- |
| **Data Set** | **Input** | **Output** |
| IRIS | 1.Sepal Length  2.Sepal Width  3.Petal Length  4.Petal Width  5.Class | Sepal length(3): class  Sepal width(3): class  Petal length(3): class  Petal width(3): class |

**SEARCH FOR BEST MODEL**

**3. Data Preparation**

**4. Selecting an Algorithm**

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Result from Weka** | |
| **Correctly Classified Instances** | **Mean Absolute Error** |
| **Weka Classifiers Bayes** | | |
| Bayes Net | 139 | 0.0454 |
| Naïve Bayes | 144 | 0.0342 |
| Naïve Bayes Updateable | 144 | 0.0342 |
| **Weka Classifiers Functions** | | |
| Logistic | 144 | 0.287 |
| SMO | 144 | 0.2311 |
| SimpleLogistic | 141 | 0.0439 |
| **Weka Classifiers Lazy** | | |
| IB1 | - | - |
| IBk | 143 | 0.0399 |
| LWL | 140 | 0.1041 |
| **Weka Classifiers Meta** | | |
| Bagging | 141 | 0.0529 |
| Ada Boost M1 | 143 | 0.069 |
| Multi Class Classifier | 144 | 0.0987 |
| **Weka Classifier Misc** | | |
| Hyper Pipes | - | - |
| VFI | - | - |
| **Weka Classifier Trees** | | |
| J48 | 144 | 0.035 |
| AD Tree | - | - |
| NB Tree | - | - |
| **Weka Classifier Rules** | | |
| Zero R | 50 | 0.4444 |
| J Rip | 143 | 0.0454 |
| Ridor | - | - |

**Selecting Algorithm Training Parameters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm & Evaluation Method** | Weka classifiers meta Bagging | | |
| **Training Parameter** | | **Result from Weka** | |
| classifiers | | Correctly Classified Instances | Mean Absolute Error |
| REP Tree | | 141 | 0.0563 |
| AD Tree | | - | - |
| Decision Stump | | 100 | 0.2222 |
| J48 | | 144 | 0.035 |
| Random Tree | | 138 | 0.0533 |
| Random Forest | | 143 | 0.0408 |

**Data Preprocessing**

|  |  |
| --- | --- |
| **Attribute Evaluator** | Classifier Subset Eval (Classifier: meta.Bagging with training parameter: classifiers = J48 and the others in their default) |
| **Search Method** | BestFirst (Direction: Forward) |
| **Evaluation Mode** | 10-fold cross-validation (Folds = 10, Seed =1) |
| **Attribute and Number of Folds (%)** | |
|  | |

Results for different cut-off number of folds are summarized as follows:

| **Algorithm** | weka.classifiers.meta.Bagging | |
| --- | --- | --- |
| **Training Parameter** | classifier = J48 (Other parameters in default) | |
| **Rule** | **Attribute Selection** | **Result from Weka** |
| Cut-off Number of Folds | Removed Attribute | Correctly Classified Instances, Mean Absolute Error |
| 2 |  |  |
| 4 |  |  |
| 6 |  |  |
| 8 |  |  |
| 10 |  |  |

**Iteration**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm & Evaluation Method** | | weka.classifiers.trees.J48 | |
| **Training Parameter** | | **Result from Weka** | |
| binarySplits | Unpruned | Correctly Classified Instances | Mean Absolute Error |
| False | False |  |  |
| True | False |  |  |
| False | True |  |  |
| True | True |  |  |

Selecting algorithm training parameters:

Based on the attribute list from Weka:

|  |  |
| --- | --- |
| **Attribute Evaluator** | ClassifierSubsetEval (Classifier: trees.J48 with training parameters: binarySplits = False, Unpruned = True and the others in their default) |
| **Search Method** | BestFirst (Direction: Forward) |
| **Evaluation Mode** | 10-fold cross-validation (Folds = 10, Seed =1) |
|  |  |
| **Attribute and Number of Folds (%)** | |
| - | |

Reducing input size:

| **Algorithm** | weka.classifiers.trees.J48 | |
| --- | --- | --- |
| **Training Parameter** | binarySplits = False, unpruned = True (Other parameters in default) | |
| **Rule** | **Attribute Selection** | **Result from Weka** |
| Cut-off Number of Folds | Removed Attribute | Correctly Classified Instances, Mean Absolute Error |
| 2 |  |  |
| 4 |  |  |
| 6 |  |  |
| 8 |  |  |
| 10 |  |  |

Selecting algorithm training parameters:

|  |  |  |
| --- | --- | --- |
| **Algorithm & Evaluation Method** | weka.classifiers.trees.ADTree | |
| **Training Parameter** | **Result from Weka** | |
| searchPath | Correctly Classified Instances | Mean Absolute Error |
| Expand all paths |  |  |
| Expand the heaviest path |  |  |
| Expand the best z-pure path |  |  |
| Expand a random path |  |  |

Based on the attribute list from Weka:

|  |  |
| --- | --- |
| **Attribute Evaluator** | ClassifierSubsetEval (Classifier: trees.ADTree with training parameters: searchPath = Expand a random Path and the others in their default ) |
| **Search Method** | BestFirst (Direction = Forward) |
| **Evaluation Mode** | 10-fold cross-validation (Folds = 10, Seed =1) |
| **Attribute and Number of Folds (%)** | |
|  | |

Reducing input size:

|  |  |  |
| --- | --- | --- |
| **Algorithm** | weka.classifiers.trees.ADTree | |
| **Training Parameter** | searchPath = Expand a random path (Other parameters in default) | |
| **Rule** | **Attribute Selection** | **Result from Weka** |
| Cut-off Number of Folds | Removed Attribute | Correctly Classified Instances, Mean Absolute Error |
| 2 | - |  |
| 4 |  |  |
| 6 |  |  |
| 8 |  |  |

Selecting algorithm training parameters:

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm & Evaluation Method** | | weka.classifiers.rules.JRip | |
| **Training Parameter** | | **Result from Weka** | |
| checkErrorRate | usePruning | Correctly Classified Instances | Mean Absolute Error |
| True | True |  |  |
| False | True |  |  |
| False | False |  |  |
| True | False |  |  |

.

Based on the attribute list from Weka:

|  |  |
| --- | --- |
| **Attribute Evaluator** | ClassifierSubsetEval (Classifier: rules.JRip with training parameters: checkErrorRate = False, usePruning = True and the others in their default) |
| **Search Method** | BestFirst (Direction: Forward) |
| **Evaluation Mode** | 10-fold cross-validation (Folds = 10, Seed =1) |
| **Attribute and Number of Folds (%)** | |
|  | |

Reducing input size:

|  |  |  |
| --- | --- | --- |
| **Algorithm** | weka.classifiers.rules.JRip | |
| **Training Parameter** | checkErrorRate = False, usePruning = True (Other parameters in default) | |
| **Rule** | **Attribute Selection** | **Result from Weka** |
| Cut-off Number of Folds | Removed Attribute | Correctly Classified Instances, Mean Absolute Error |
| 2 |  |  |
| 4 |  |  |
| 6 |  |  |
| 8 |  |  |
|  | | |

**SUMMARY**

**Result Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **weka.classifiers.meta.Bagging** | **weka.classifiers.trees.J48** | **weka.classifiers.trees.ADTree** | **weka.classifiers.rules.JRip** |
| **Training Parameter** |  |  |  |  |
|  |  |  |  |
|  |  |
| **Removed**  **Input**  **Attribute** |  |  |  |  |
| **Correctly Classified Instances** |  |  |  |  |

*↑ highest %*

has the highest percentage of instances being classified correctly (---------%) and is considered to be the final best performing model; this is our recommended model.

SCREENSHOTS:



