**BRAND PREDICTION ANALYSIS REPORT**

* + **EVALUATED CLASSIFIERS**

Two classifiers were used to predict the brand preference: IBK nearest-neighbor and J48 (decision Tree)

**IBK nearest-neighbor Training Set Parameters**:

IBK with Cross Validation: Survey\_Responses\_Complete.arff was used as the training set with no preprocess applied to attributes. Cross-Validation of 10 folds was applied. K values applied: 1,2,3,10,20,30,40,35,28,24,23. The best performance was observed for K=24 as follows:

Correctly classified instances: 8239 / 82.3%

Incorrectly classified instances: 1770 / 17.7%

Kappa Coefficient: 0.616

Mean Absolute Error: 0.3519

Root Mean Squared Error: 0.3852

Relative Absolute Error: 74.8016%

Root Relative Absolute Error: 79.4382%

IBK K=1 without Cross-Validation: Survey\_Responses\_Complete.arff was used as the training set with no preprocess applied to attributes.

Correctly classified instances: 10000 / 100%

Incorrectly classified instances: 0 / 0%

Kappa Coefficient: 1

Mean Absolute Error: 0.0001

Root Mean Squared Error: 0.0001

Relative Absolute Error: 0.0213%

Root Relative Absolute Error: 0.0206%

**J48 Tree Training Set Parameters (See Performance in table below)**:

J48 Run 1: Survey\_Responses\_Complete.arff was used as training set with no preprocess applied to attributes. Default cross-validation of 10 folds was applied. Overfitted tree of 329 leaves and tree size of 562.

J48 Run 2: Survey\_Responses\_Complete.arff was used as training set with discretized salary. Default cross-validation of 10 folds was applied. Best performance metrics of all J48 models. Prunned tree of 28 leaves and tree size of 44.

J48 Run 3: Survey\_Responses\_Complete.arff was used as training set with discretized salary and age. Default cross-validation of 10 folds was applied. Prunned tree of 87 leaves.

J48 Run 4: Survey\_Responses\_Complete.arff was used as training set with discretized age. Default cross-validation of 10 folds was applied. Prunned tree of 52 leaves.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Performance** | **J48 Run 1** | **J48 Run 2** | **J48 Run 3** | **J48 Run 4** |
| Correctly classified instances | 8075 /80.75% | 9059/90.59% | 8992/89.92% | 9019/90.19% |
| Incorrectly classified instances | 1925/19.25% | 941/9.41% | 1008/10.08% | 981/9.81% |
| Kappa Statistic | 0.5919 | 0.8022 | 0.788 | 0.7914 |
| Mean Absolute Error | 0.218 | 0.1421 | 0.1365 | 0.1327 |
| Root Relative Error | 0.3594 | 0.2714 | 0.2637 | 0.2668 |
| Relative Absolute Error | 46.3541% | 30.2097% | 29.0211% | 28.2077% |
| Root Relative Squared Error | 74.1156% | 55.9622% | 54.3774% | 55.0056% |
| Leaves | 329 | 28 | 87 | 52 |

* + **SELECTED CLASSIFIER INCLUDING A RATIONALE FOR SELECTING THE METHOD YOU DID AND THE LEVEL OF CONFIDENCE IN THE PREDICTIONS.**

IBK K=1 with no cross-validation was selected as the best prediction model due to its performance. The Test set was re-evaluated using this model.

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances 10000 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0001

Root mean squared error 0.0001

Relative absolute error 0.0213 %

Root relative squared error 0.0206 %

Total Number of Instances 10000

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 0 1 1 1 1 0

1 0 1 1 1 1 1

Weighted Avg. 1 0 1 1 1 1

=== Confusion Matrix ===

a b <-- classified as

3783 0 | a = 0

0 6217 | b = 1

IBK Test set WEKA output details can be checked in “IBK BRAND TEST TS” Text File

* + **The predicted answers to the brand preference question for the instances of survey results that are missing that answer.**
* Predicted answers can be checked in “IBK BRAND TEST TS” Text File and “Complete and Brand Predictions” Excel File
  + A chart that displays the customer preference for each brand based on the combination of the actual answers and the predicted answers to the brand preference survey question.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Surveyed preference** | **%** | **Predicted preference** | % |
| **Acer** | 3783 | 37.83% | 2089 | 41.78% |
| **Sony** | 6217 | 62.17% | 2911 | 58.22% |

It can be observed that brand preference percentage differ by approximately 3% between surveyed and predicted values.

* The results of each classifier were exported from WEKA and included in the ZIP file.