IST 707 Applied Machine Learning

By Prof. Kelvin King

Assignment 6

**Model Comparison – MNB and DT**

Submitted by:

Samarth Sandesh Mengji  
SUID: 718473878

NetID: [smengji@syr.edu](mailto:smengji@syr.edu)

Date of Submission: 10/15/2023

Syracuse University – School of Information Studies

**Section 1: Introduction**

The objective of this report is to compare the performance of two classification algorithms, Decision Tree, and Multinomial Naïve Bayes, on the task of handwritten digit recognition.

**Section 2: Decision Tree**

The following steps are taken to build a decision tree:

**Data Preprocessing:**

The training dataset is loaded and split into training and validation sets. The pixel features and labels are separated. Labels is stored in another variable.

**Model Building:**

A Decision Tree classifier is constructed with tuned hyperparameters.

Hyperparameter tuning is performed using GridSearchCV to find the best combination of parameters of max\_depth and min\_samples\_split.

A 3-fold cross-validation and sklearn.metric accuracy score is implemented to evaluate the model's accuracy.

**Following are the results:**

Decision Tree Validation Accuracy: 0.7238095238095238

Decision Tree 3-Fold CV Accuracy: 0.7248380250223413

**Section 3: Multinomial Naïve Bayes**

In this section, we build a Multinomial Naïve Bayes model for digit recognition:

**Data Preprocessing:**

Like the Decision Tree section, the training dataset is loaded and split into training and validation sets.

**Model Building:**

A Multinomial Naïve Bayes classifier is constructed for discrete features commonly found in image data.

Hyperparameter tuning is performed by changing the smoothing alpha parameter. These were the following values: 'alpha': [0.01, 0.02, 0.05, 0.1,0.5]

A 3-fold cross-validation and sklearn.metrics accuracy score is implemented to assess the model's accuracy.

Following are the results:

Multinomial Naïve Bayes Validation Accuracy: 0.8285714285714286

Multinomial Naïve Bayes 3-Fold CV Accuracy: 0.828466636878165

**Section 4: Algorithm Performance Comparison**

In this section, we compare the results obtained from the Decision Tree and Multinomial Naïve Bayes models:

**Accuracy Comparison:**

Multinomial Naïve Bayes achieved a higher 3-fold cross-validation accuracy of approximately **82.85%,** while the Decision Tree achieved a cross-validation accuracy of approximately **72.48%.**

**Results on Testing Data:**

To test the models on real world data, we evaluated them on the given testing data.

The Decision Tree model achieved a testing accuracy of approximately **75.49%,** which shows that it maintained a reasonable level of accuracy on unseen data.

The Multinomial Naïve Bayes model also performed well on the testing data, with an accuracy of approximately **81.68%,** which shows that it can perform well on digit recognition data.

**Algorithm Performance Analysis:**

**Multinomial Naïve Bayes** outperformed the Decision Tree both in cross-validation and testing accuracy.

These results suggest that **Multinomial Naïve Bayes is better** suited for digit recognition. It effectively captures patterns in the data, leading to improved accuracy.

The testing accuracy of Multinomial Naïve Bayes indicates that it can be reliable for unseen/real world data and is a strong choice for this task.

**Conclusion:**

The results confirm that Multinomial Naïve Bayes is a suitable choice for digit recognition, offering higher accuracy than the Decision Tree model.

I certify that this assignment represents my work. I have not used any unauthorized or unacknowledged assistance or sources in completing it, including free or commercial systems or services offered on the internet.

Reference:

<https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>