

Amazon Sentiment Review Analysis:

In each of the following models, the following parameters are used for tuning using cross-validation and the final accuracy is calculated.

1) Decision Tree Classification:

The decision tree implementation in Assignment 1 obtained its best accuracy, after pruning the decision tree with a maximum depth of 10 on the Amazon Review dataset.

The training accuracy obtained is: 88.5% and Testing Accuracy obtained is: 83.9%

2) Neural Network Classification:

The Neural Network implementation in Assignment 2 obtained a best accuracy by setting the following parameters: iterations = 500, number of hidden layers = 600, Activation function = RELU, Solver = SGD, Learning Rate = 0.001, Regularization Parameter alpha=0.0001.

The above settings gave a training accuracy of 91.81% and Testing accuracy of 73.89%

3) K-Nearest Neighbors Classification Model:

For K-Nearest Neighbor classification model, after hyperparameter tuning of K, the value of K =6 gives more accuracy on the Testing dataset. Hence, K = 6 is considered as the tuned parameter. Hence it obtained an accuracy of 86.62% on the Training dataset, 85.31% on the testing dataset. (Reference Assignment 3)

4) Adaboosting Classification model:

After tuning the parameters of the n_estimators, the value n_estimator=900 gives a better accuracy which reduces the error on both the training as well as a testing dataset.

The above settings gave an accuracy of 94.63% on the training dataset, 74.66% on the testing dataset.

5) SVM Classification Model:

For the Support Vector Machine Classification model, the values gamma, C are tuned and comparison of various kernels is done. After tuning, the Poly Kernel seems to be the best fit and the degree used is 3.

The above parameters on the SVM Classification model gave an accuracy of 85.54% on the Training dataset, 85.53% on the testing dataset.

6) Naïve Bayes Classification Model:

For the Naïve Bayes classification model, the priors are not needed to train the learning model. The Laplacian smoothing of 1.0 is applied to the Training dataset to estimate the value. The Multinomial NB gave a better training accuracy of 87.6% and a testing accuracy of 82.79%.

Performance Measure Comparison on all Classification Models:

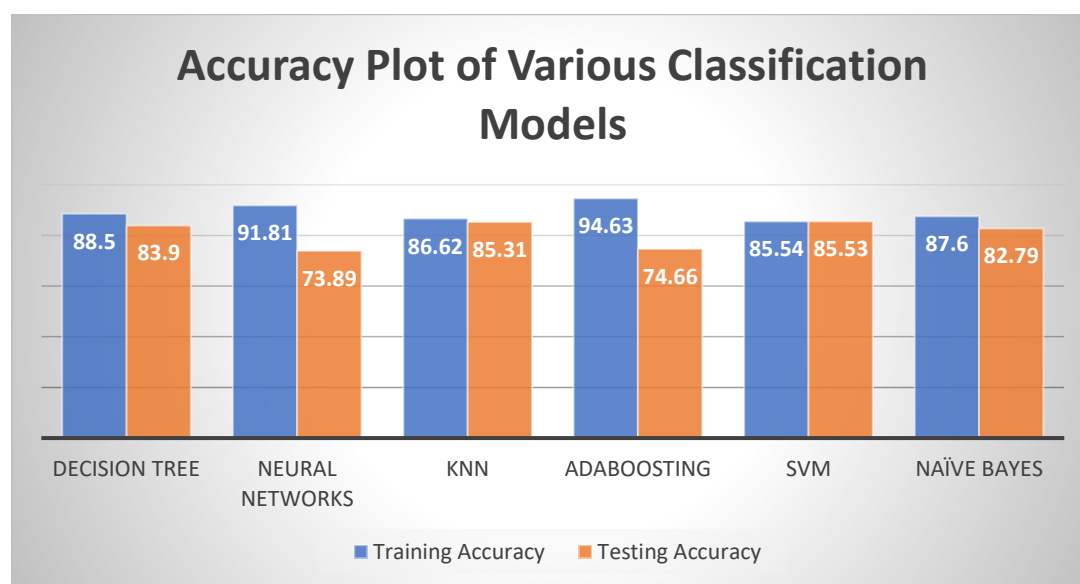
The below table consists of all the results of all the classification models that we have implemented till now. Also, the precision, recall, and F1-score of the Testing dataset are calculated. Also, sample complexity is used as performance measures to compare the results.

Color coding depicts which are the best model and worse models on Amazon Dataset.

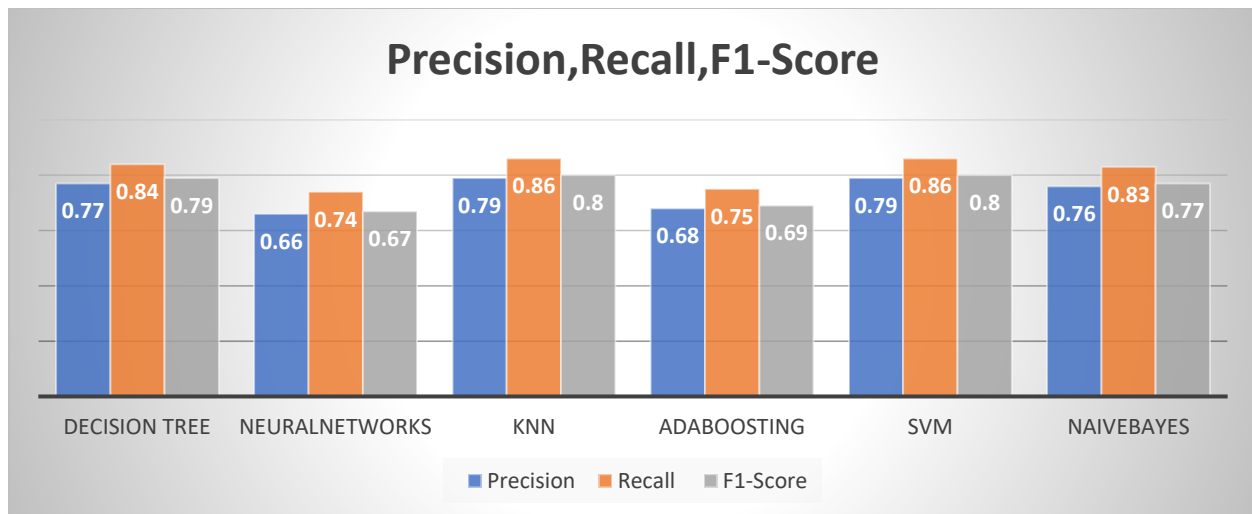
Classification Model	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
Decision Tree	88.5	83.9	0.77	0.84	0.79
Neural Networks	91.81	73.89	0.66	0.74	0.67
KNN	86.62	85.31	0.79	0.86	0.8
Adaboosting	94.63	74.66	0.68	0.75	0.69
SVM	85.54	85.53	0.79	0.86	0.8
Naïve Bayes	87.6	82.79	0.76	0.83	0.77

Table Demonstrating various values of performance of each model.

The plot for accuracy comparison of each classification models:



Plot demonstrating other performance measures F1-Scores, Precision, Recall:



From the above results and tables, one of the best classification model that perfectly classifies the given Amazon Review dataset is both the K-Nearest Neighbor classification model which gives a training accuracy of 86.62% and Testing Accuracy of 85.31%.

Also, Support Vector Machines with Poly Kernel is also giving a better accuracy with degree 3 having training accuracy of 85.54% and Testing accuracy of 85.53%.

Also, these two classification models contain the highest Precision, Recall and F1-score values which are some of the other measures for estimating the performance of the dataset. Hence from the above values, we can conclude KNN and SVM as the best classification model for the Amazon Sentiment Review Analysis.