Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 6

Due: 11/18 in 2024 (100 points)

Wenbo Qian

https://github.com/wenboqian/CS-6220/tree/main/A6

Naive Bayes

```
Training set Accuracy: 0.890
Training set Precision: 0.892
Training set Recall: 0.887
Testing set Accuracy: 0.8805250743513485
Testing set Precision: 0.8805474906677727
Testing set Recall: 0.8781799379524302
```

Decision Tree

```
Training set Accuracy: 0.998
Training set Precision: 0.996
Training set Recall: 0.999
Testing set Accuracy: 0.79
Testing set Precision: 0.79
Testing set Recall: 0.79
```

Random Forests

```
Training set Accuracy: 0.998
Training set Precision: 0.996
Training set Recall: 0.999
Testing set Accuracy: 0.8708850374320583
Testing set Precision: 0.9144645340751043
Testing set Recall: 0.8159255429162358
```

Logistic Regression

```
Training set Accuracy: 0.946
Training set Precision: 0.944
Training set Recall: 0.949
Testing set Accuracy: 0.8962157727412573
```

```
Testing set Precision: 0.890181669728516
Testing set Recall: 0.9019648397104447
```

Decision Tree, Random Forests, and Logistic Regression are overfitting.