**IST718 – Big Data Analytics**

**Lab 3**

**David Doman**

**Introduction**

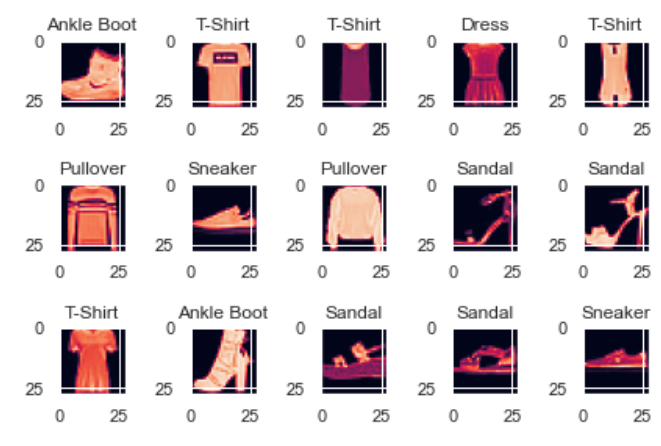
Being able to identify certain objects using algorithms seems like something that may not provide much use. Although, being able to do so can assist companies in all different industries helping to provide insight and potentially provide advantages over competitors. For example, a clothing company may want to identify clothing items for research and development. There are many different types of approaches that could be taken from Naïve Bayes, Neural Networks, Linear Classification, etc. In this report, a few of these approaches will be taken to determine which is the most successful at identifying clothing items based on the accuracy of the results.

**Dataset**

The dataset for this study comes from a MNIST fashion dataset contained within a Tensorflow Keras library in python. Training and testing datasets will be used to run the models and help to determine accuracy of different models.

Below is an initial view of the types of clothing items:

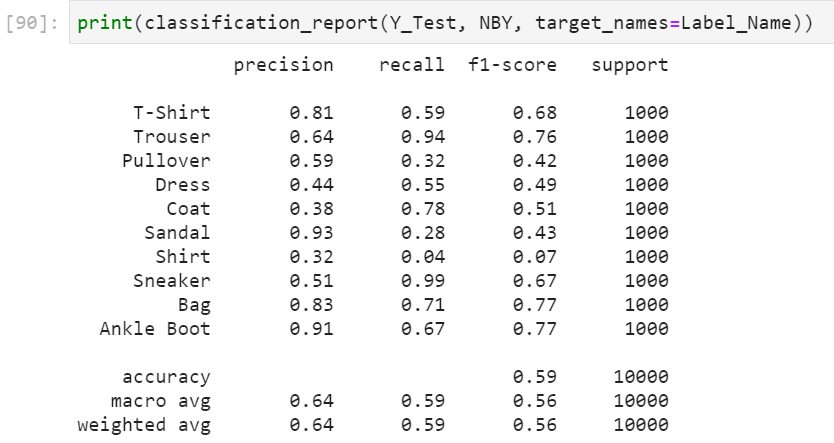


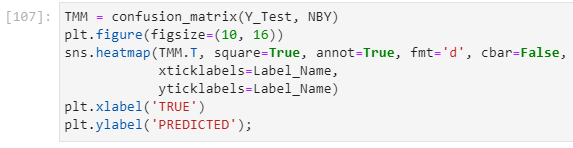


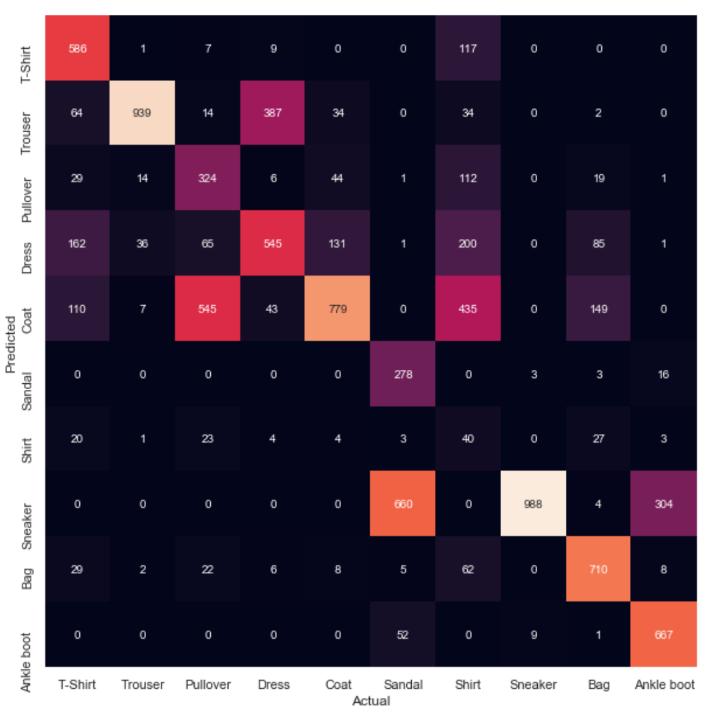
As you can see above, all different types of clothing from shoes to shirts will be examined and attempted to be identifies using algorithms.

**Models**

The first algorithm used will be the Naïve Bayes model. Results are shown below:

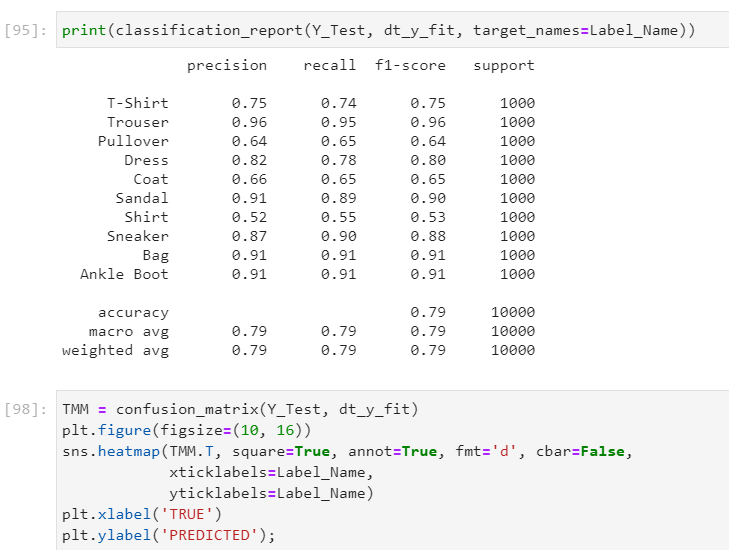


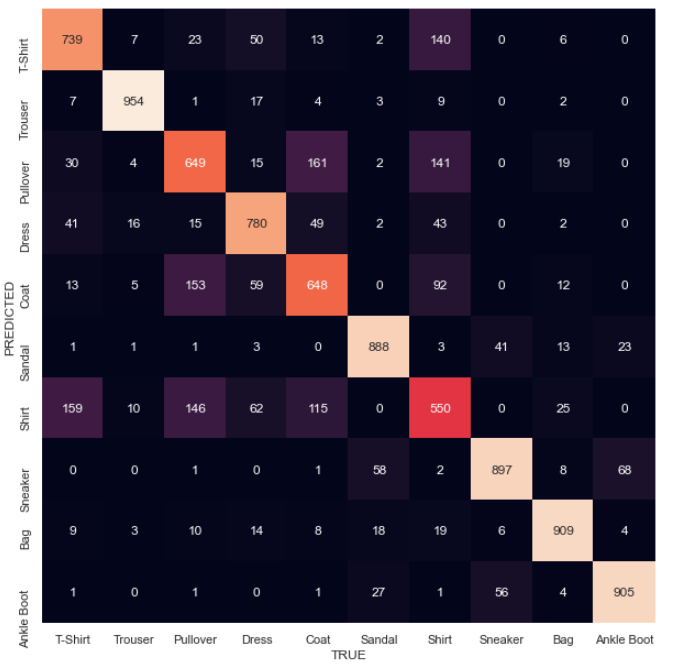




These results seem good but 2 other models will be tested to see if they can outperform this model.

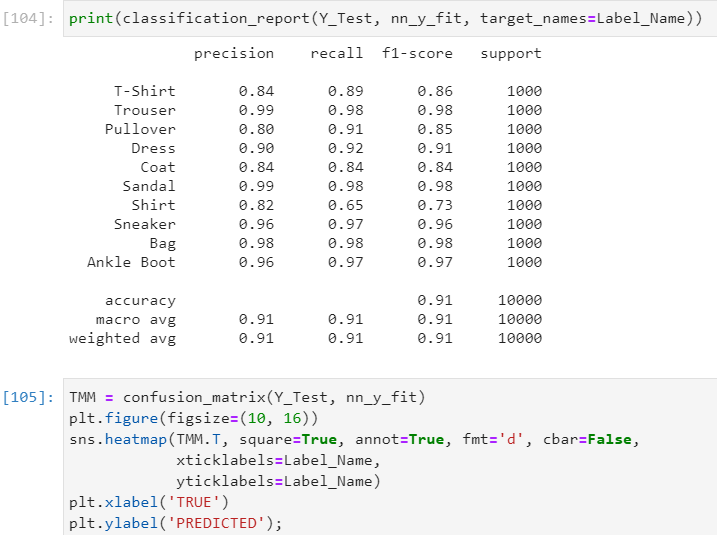
The second model used is Decision Trees. Below are the results.

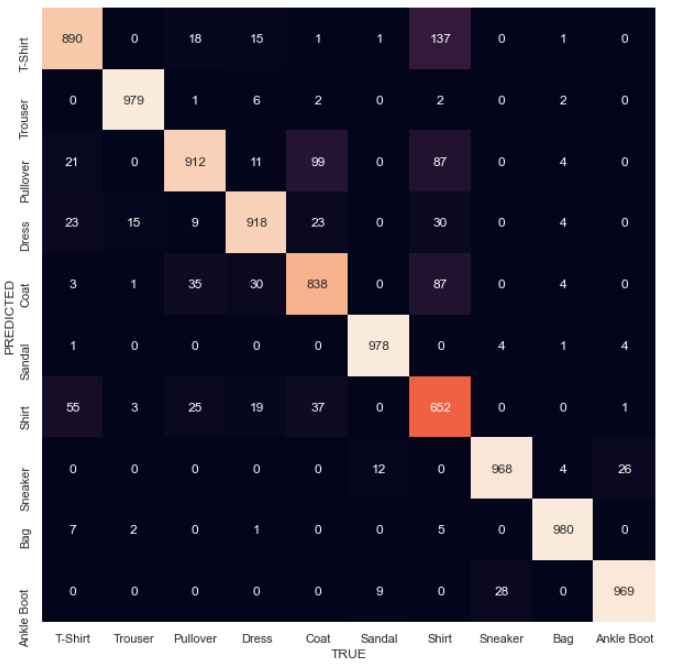




Accuracy was better with this model resulting in 79% accuracy compared to the 59% accuracy achieved by the Naïve Bayes model.

The last model run is the Neural Network model. Below are the results.





The final model resulted in 91% accuracy, beating out both Naïve Bayes and Decision Trees.

**Conclusions**

As described above, the 3 models run were the Naïve Bayes, Decision Trees and Neural Network. Results showed that the Neural Network model provided the optimal results with 91% accuracy, followed by Decision Trees at 79% accuracy and lastly the Naïve Bayes model at 59% accuracy. All these models did take a decent amount of type to process, which should also be taken into consideration when being considered by companies, as certain industries may need answers more quickly than others who have more time to allow for processing. Further analysis would need to be performed before any company would want to invest in utilize any of these models.