

MY APPROACH TO THE PROBLEM

I started off with a clean slate, taking the data set in hand and making it useful for training. The Fashion-MNIST is a dataset of Zalando's (<http://www.zalando.com>) article images —consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 gray scale image, associated with a label from 10 classes.

Fashion-MNIST serves as a direct drop-in replacement for the original MNIST dataset for benchmarking machine learning algorithms. It shares the same image size and structure of training and testing splits.

filePath = '/cxldata/datasets/project/fashion-mnist/'

I had many training models at my disposal so I tried out multiple models on the dataset and obtained the results of each. After comparing I found that the Ensemble(Voting Classifier) was the most superior of all of the other techniques and hence I selected that model for our Fashion MNIST Classifier.

I also took the cross validation of each model and came to know that the Decision Tree Classifier and the Random Forest Classifier tend to overfit the data and hence not suitable for the model training purpose.

After training and selecting the optimum model , I went a step further in fine tuning the hyper parameters to obtain even better score on the testing parameters such as Accuracy. Finally after all the steps I obtained the final model ready to be tested on a test data.

The final score results were really encouraging , which could even further be improved by a little tweaking of the hyper parameters.