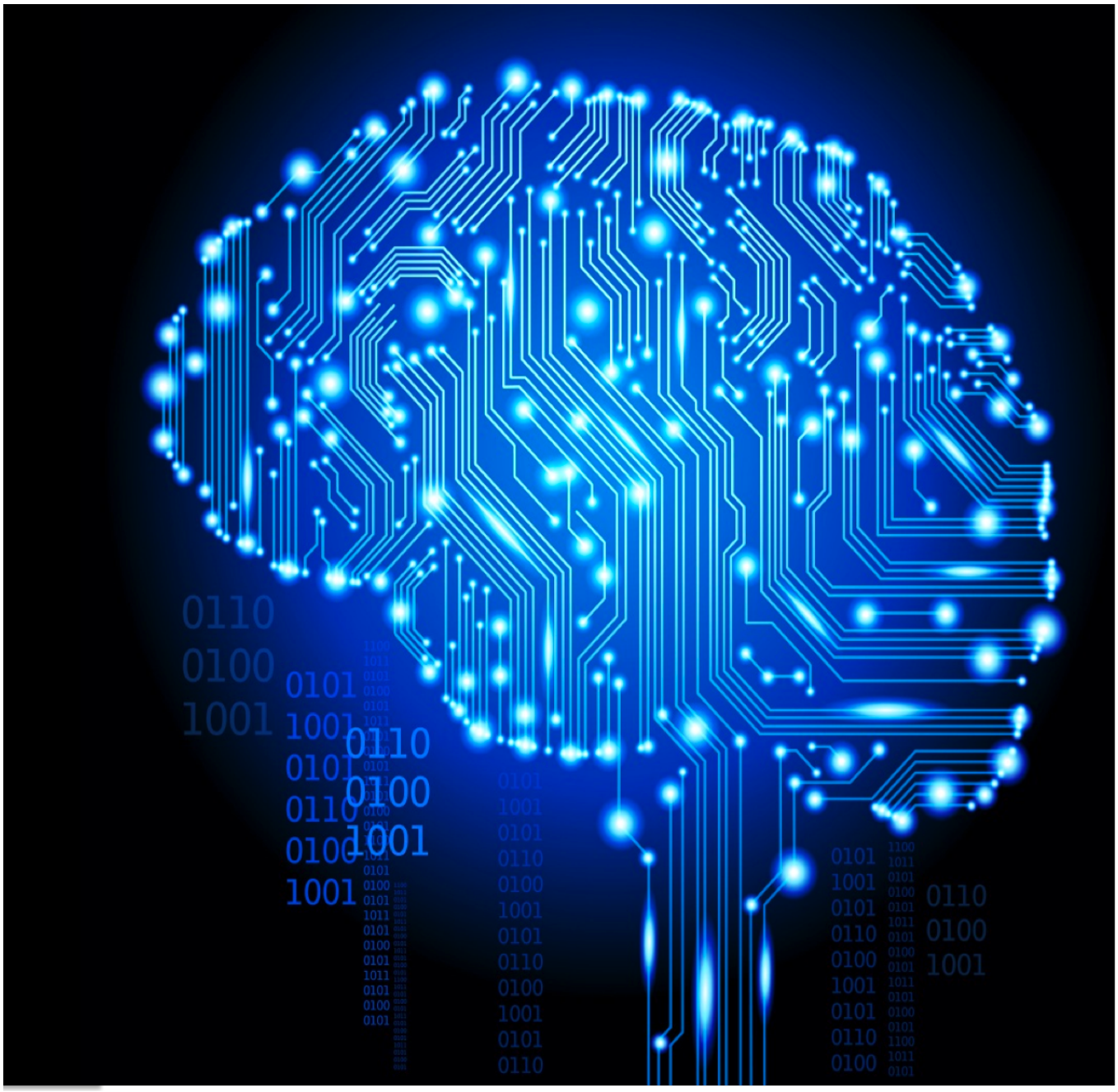

MSBD 6000B PROJECT 1

REPORT FOR MODEL USING AND PRE-/POST-PROCESSING

MSBD 6000B

PROJECT 1

REPORT FOR MODEL USING AND PRE-/POST-PROCESSING

TASK

This project is to train a model on the training data set with any classifier such as linear regression or SVM. With the trained model, predict the class labels for the test data set and put the result into a csv file.

Training data: traindata.csv

Class label: trainlabel.csv

Testing data: testdata.csv

Data Preprocessing

As the data in the “traindata.csv” is just so raw numerical data without specified the feature meaning, I added f1 – f57 to the first row of the csv file to make a complete and to be easier to read.

Then I used the python panda package to import the “traindata.csv” and the “traindata.csv”.

And before training the model, I had normalized the train data by subtracted the mean for each feature and Divide each feature by its Standard deviation.

Model Training

First I split to data randomly into 10 parts to cross train the model using K-fold. Then I used SVM with the Radial Basis Function kernel to train the model.

Results

After training the model, I used the model to predict the result for the “testdata.csv” and saved the result in the “project1_20057566.csv” file.