Report

**CSC 4139: Data Warehouse and Mining**

**Part 1 : k-nearest neighbor**

From the given dataset we imported the training features: **X\_train**, training labels: **y\_train**, and testing features: **X\_test** into matlab and fitted k-nearest neighbor model with the function **fitcknn()**. We used the value **k = 5** and the simple **Euclidean** distance measure to compute the distance between two samples.

The labels for **X\_test** was predicted using the **predict()** function. The accuracy was measured with the predicted labels and the **y\_test** data using the **classperf()** function.

The accuracy for kNN is **98.60%.**

**Part 2 : feedforward Artificial Neural Network**

From the given dataset we imported the training features: **X\_train**, training labels: **y\_train**, and testing features: **X\_test** into matlab. We transformed the **y\_train** into a sparse matrix of vectors, with 1 in each column, as indicated by ind. ind2vec(ind,N) returns an N -by- M matrix.

We used two methods to train the model.

1. **Feedforward** : We used the **feedforward()** function with 25 neurons to train the model. It takes long time to train the model and the accuracy rate is –  
   Feedforward Accuracy = **95.90%**
2. **Patternnet**: We used the **patternnet()** function with 25 neurons to train the model. It takes very short time to train the model and the accuracy rate is –  
   Patternnet Accuracy = **98.60%**

**Part 3: SVM**

From the given dataset we imported the training features: **X\_train**, training labels: **y\_train**, and testing features: **X\_test** into matlab.

We trained the model using the **fitcsvm()** function. We used the Kernel function = **'polynomial'** and Polynomial order = **2**.

The accuracy for SVM is = **99.80%**