Experimental Results and Analysis

We used a data set consisting of 3162 records for testing our models.

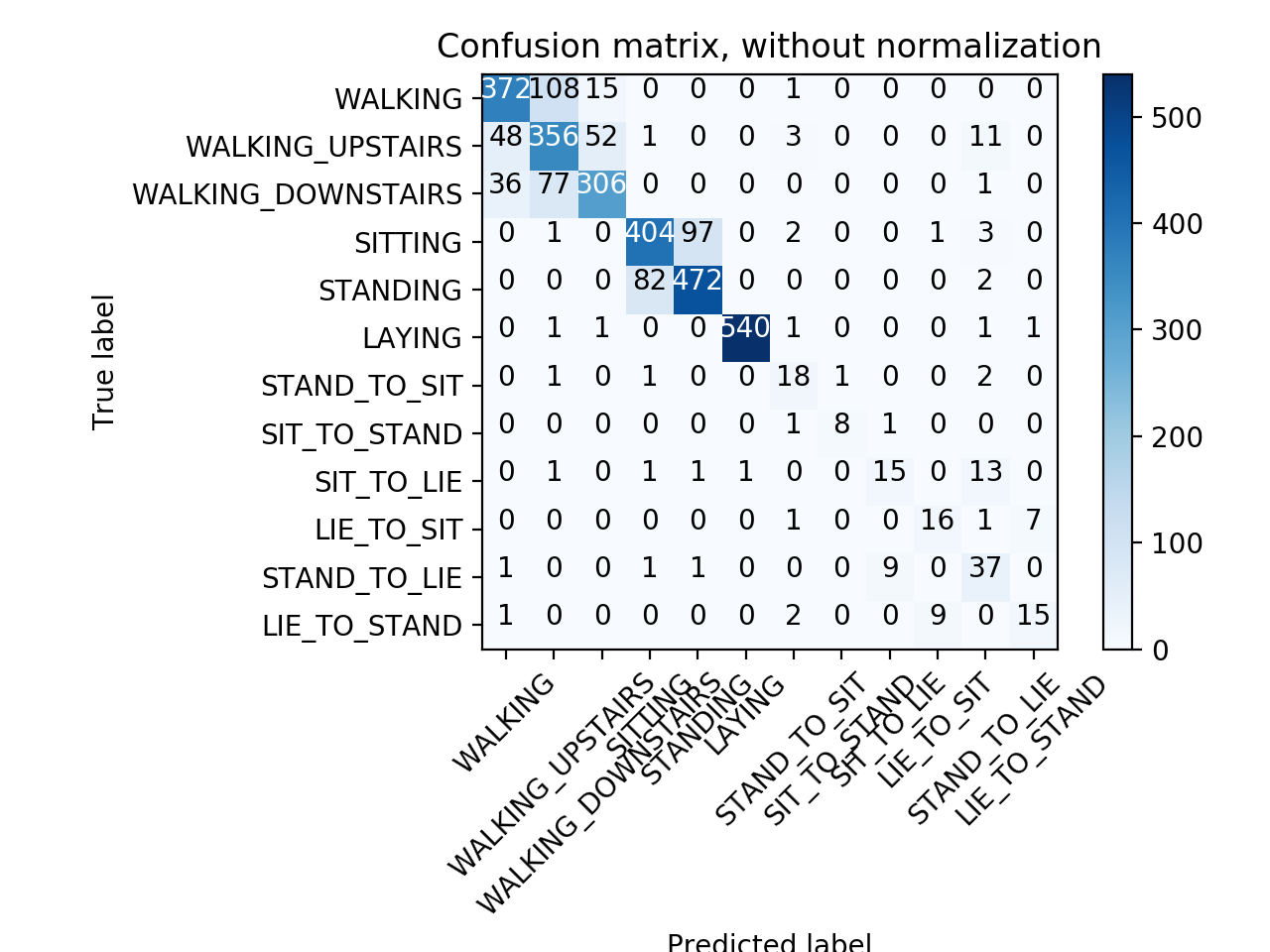
Model 1: Decision Tree

Number of correctly classified records: 2562

Accuracy: 81.025%

Mean Squared Error: 0.577

The confusion matrix obtained for this model is shown below:



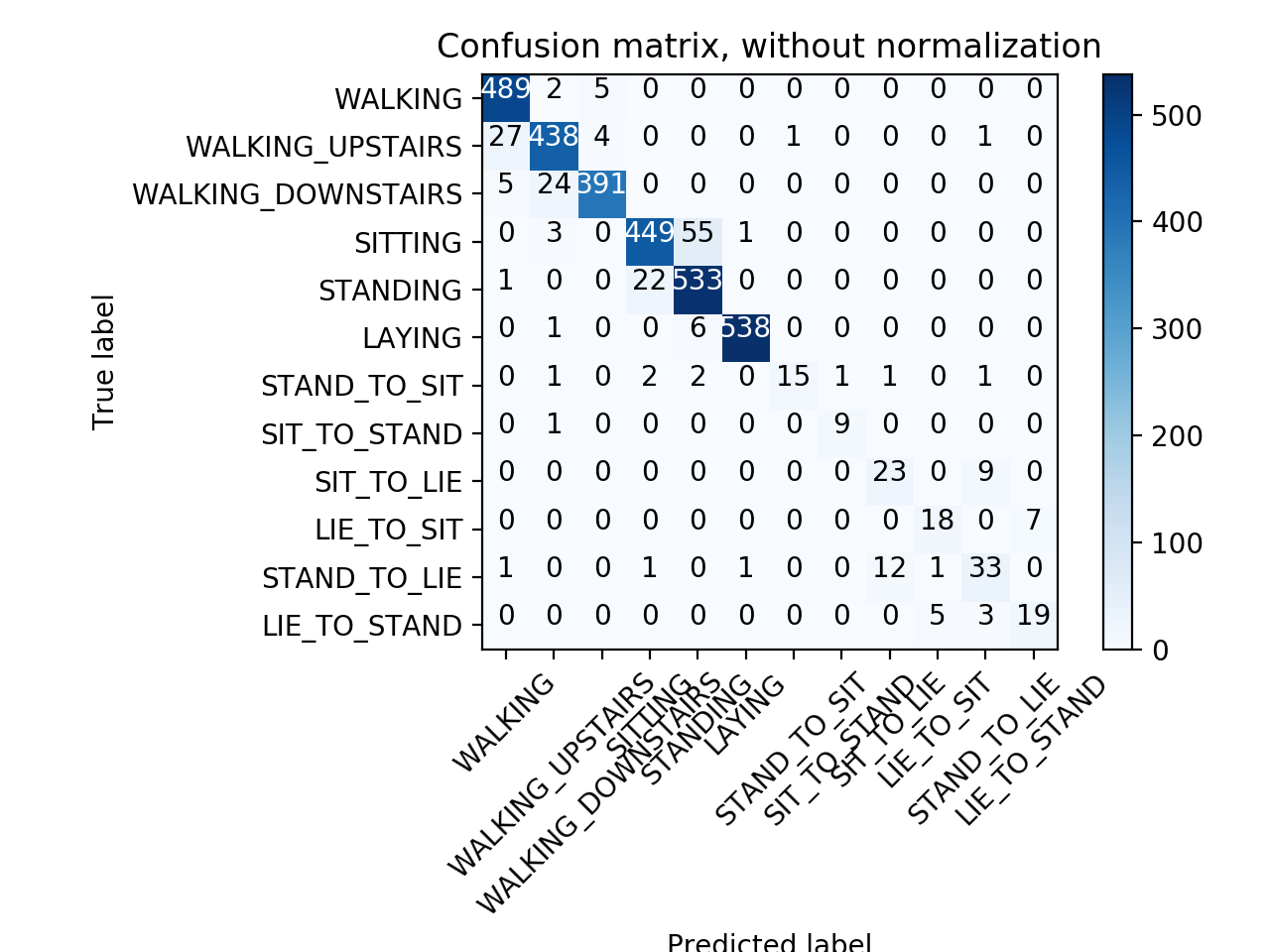
Model 2: Multi-Layer Perceptron

Number of correctly classified records: 2914

Accuracy: 92.16%

Mean Squared Error: 0.285

The confusion matrix obtained for this model is shown below:



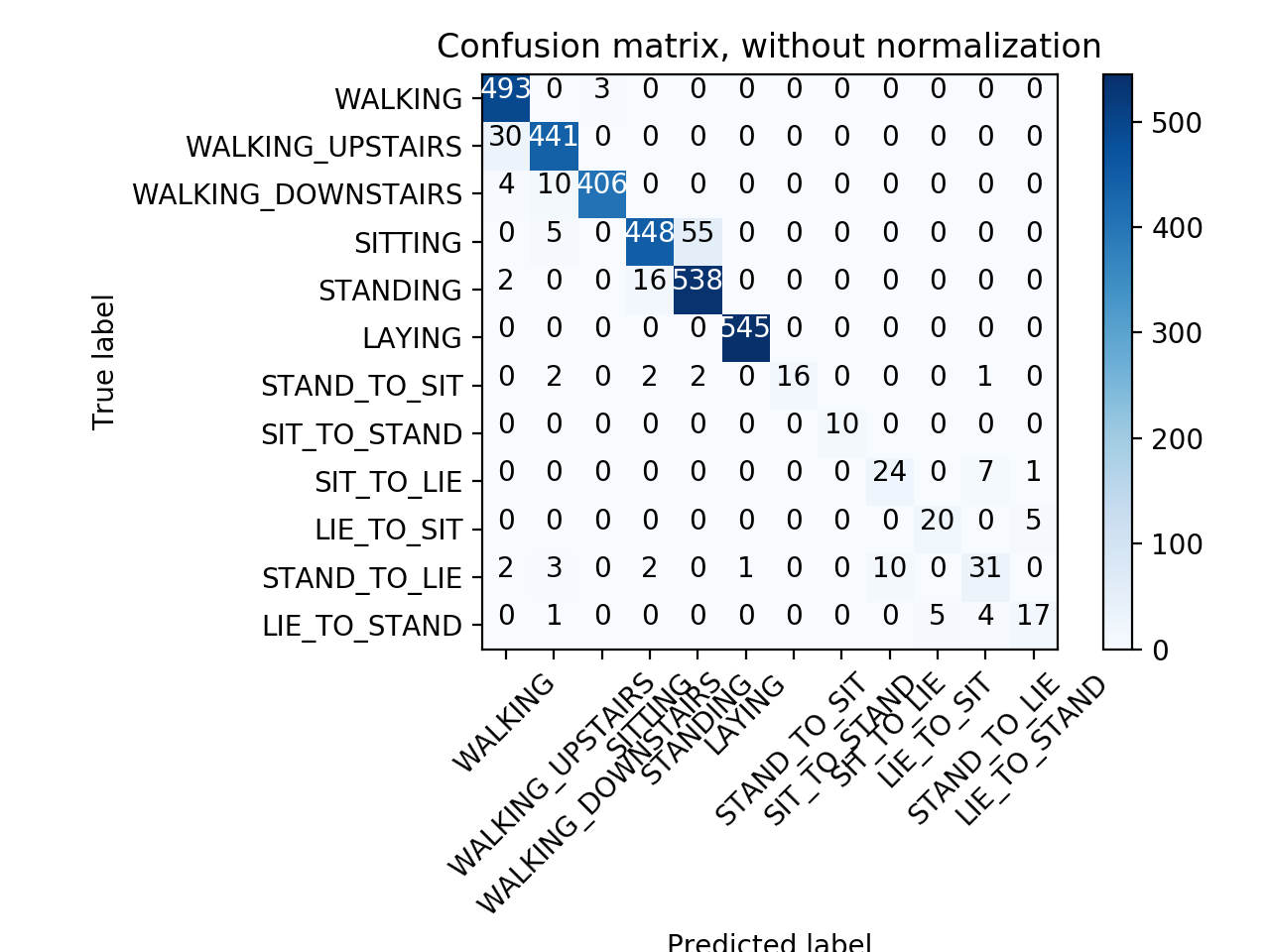
Model 3: Logistic Regression

Number of correctly classified records: 2989

Accuracy: 94.53%

Mean Squared Error: 0.338

The confusion matrix obtained for this model is shown below:



We started with the process of creating a Decision Tree for which we calculated Mean Squared Error, Confusion Matrix and Accuracy. Since, the accuracy was low we went ahead and implemented Multi-Layer Perceptron Classifier and calculated the metrics for the Neural Network. There was a boost in the accuracy by around more than 10% and the Mean Squared Error was decreased a lot. We got a model which was around ~93% accurate and both models took around similar amount of time to train the model. After that we implemented Logistic regression wherein we got an even more accurate model with accuracy more than 94% but surprisingly the Mean Squared error rose and the time for training this model was more than twice the time taken to train the other two models.

Conclusion:

By analyzing the above data, we concluded that for the given data we can use either MLP Classifier or Logistic Regression depending on the requirement i.e. if we need more accuracy, we can use Logistic Regression and when we have less time to train the model, we can use MLP Classifier.

Another conclusion we made is that the Mean Squared Error is not suitable for classification models as we can see that it is not a good parameter to determine the accuracy of our models.