Student Name – Aishvarya Krishnan
Course Name – CS7646 Machine Learning for Trading
Project Name – 2014Fall7646 Project 4

My Method:

Indicators/Features used
Assuming the data is close to sine wave form, I have used the following features to forecast the daily
returns of stock prices after five days – Phase, Amplitude and Frequency. To perform fitting of the daily
returns data to a sine curve I have used FFT(Fast Fourier Transform) and obtained the amplitude and
frequency for global maxima of the frequency spectrum. Using these features, I have constructed the
sine curve for the data.

Learning method used
I have used KNN learning method in this implementation since it performs relatively well will good
coorrelation (>0.9) and low RMSE for the dataset. Random Forest Learner also leads to good results
but it takes more execution time and considering the size of dataset, I preferred KNN Learner.
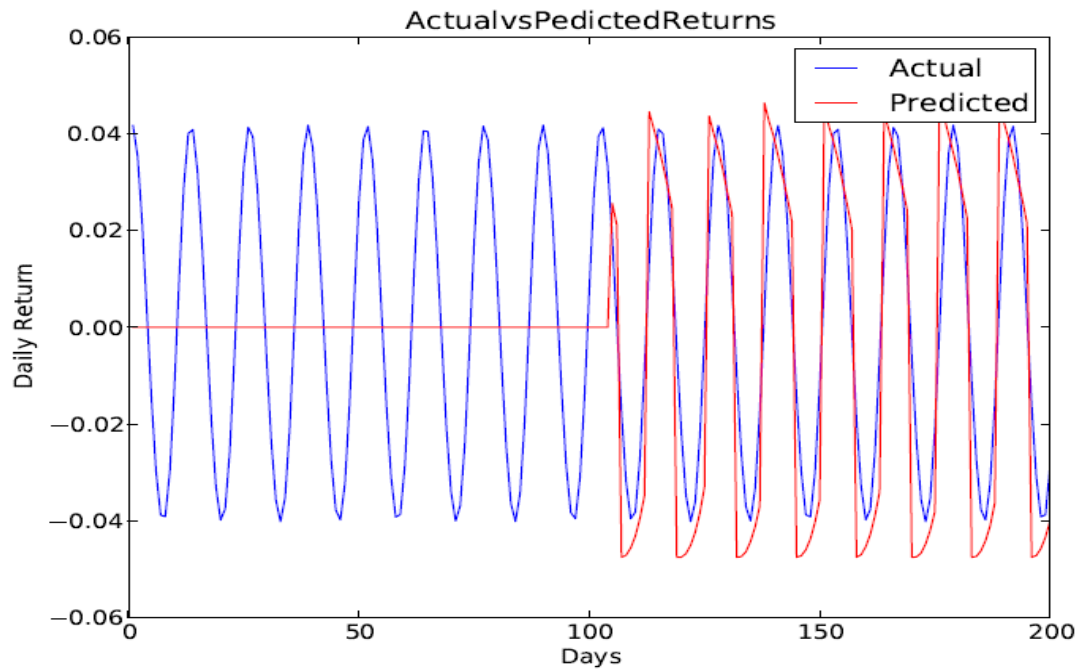
Charts:

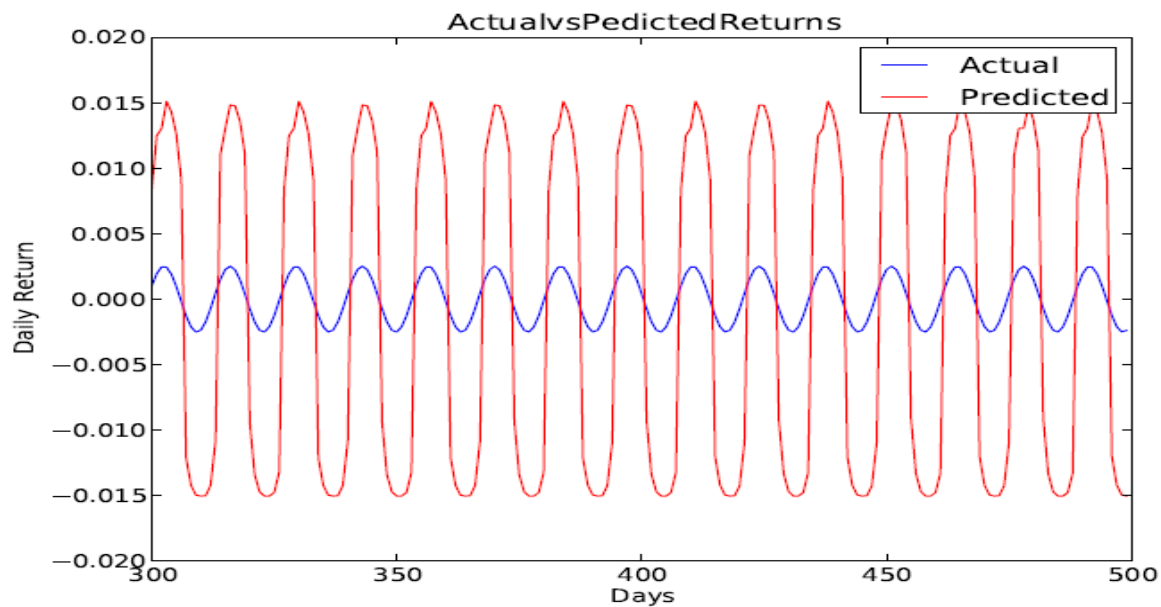1) First 200 days of Yactual vs Ypredict

ML4T-292
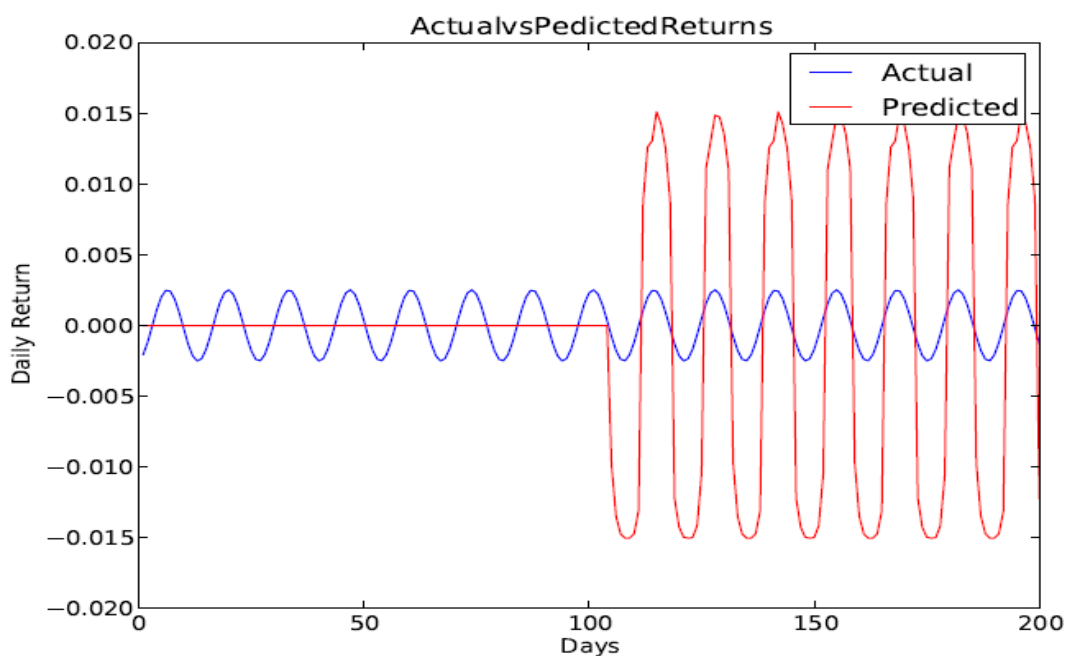
ML4T-312

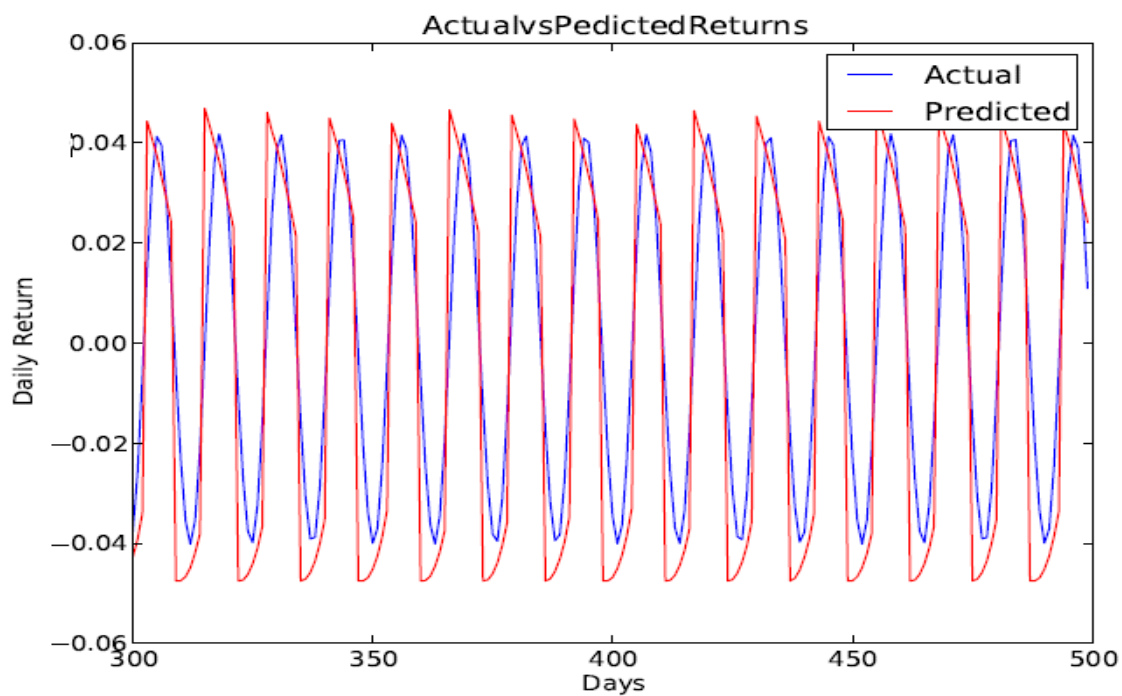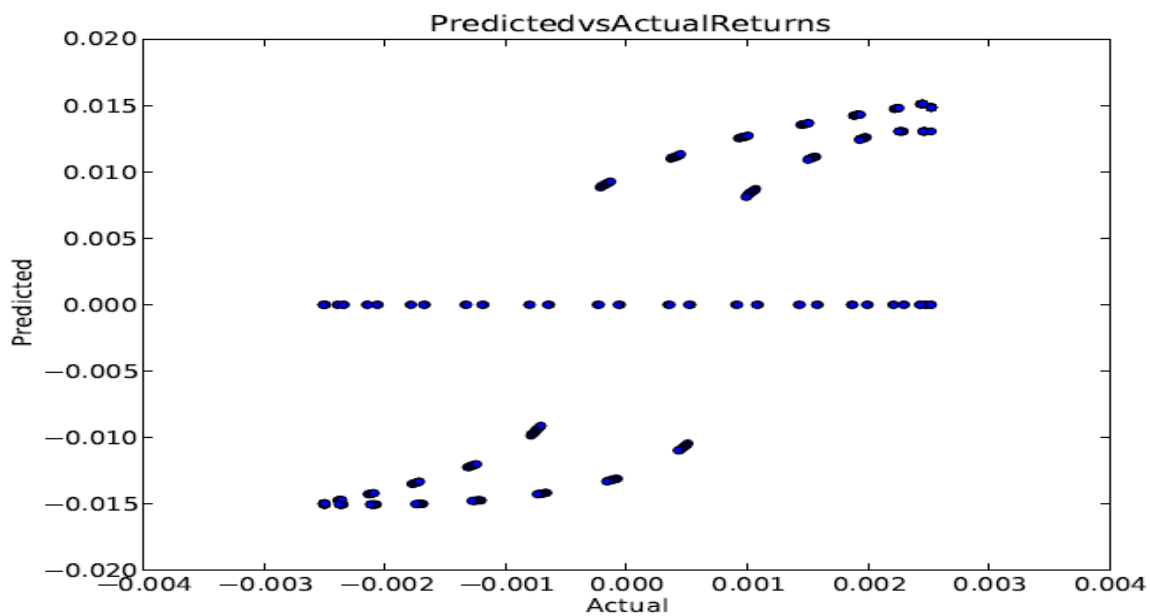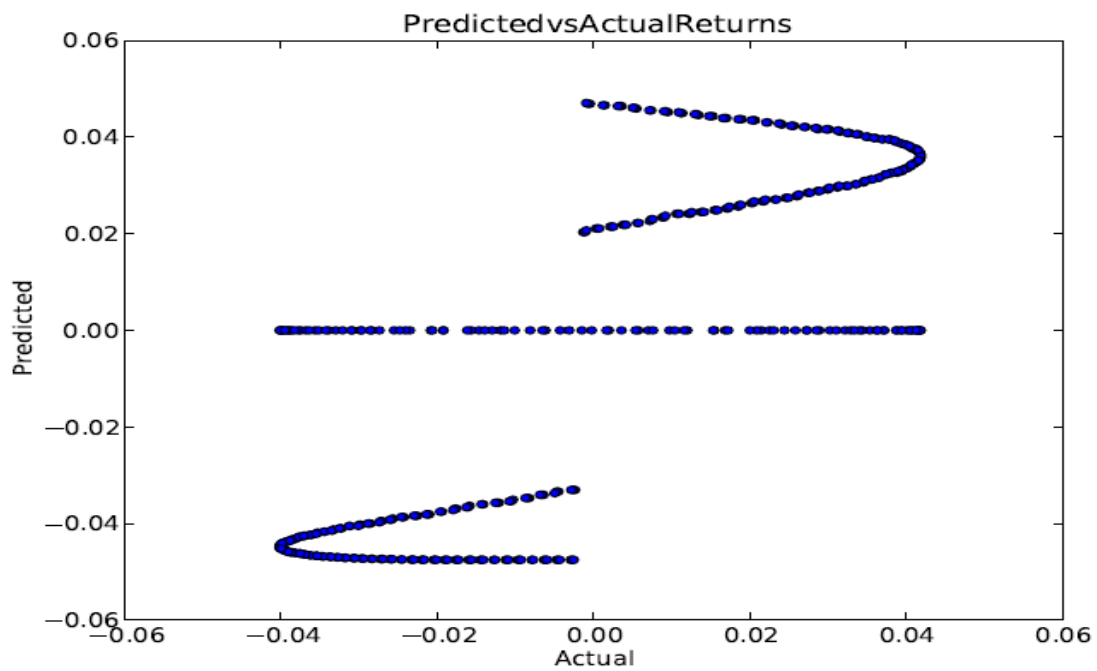2) Last 200 days of Yactual and Ypredict

ML4T-292

ML4T-312

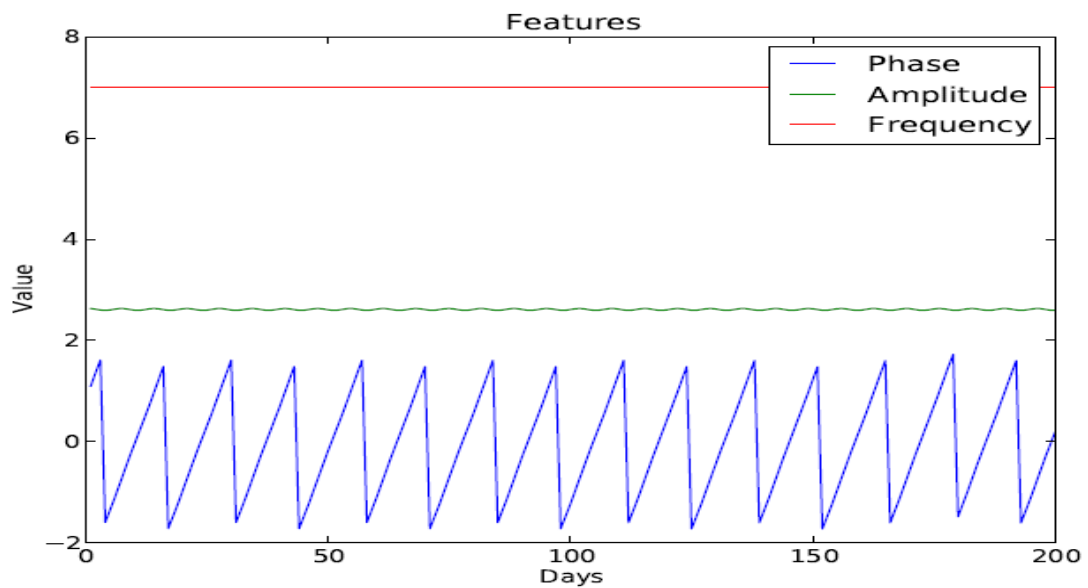3) Scatter plot of Ypredict vs Yactual

ML4T-292

ML4T-312

4) Plot of features used – Phase, Amplitude, Frequency of data

ML4T-292

<u>RMSE and Correlation for data:</u>

ML4T-292:
Correlation: 0.917254285455
RMS: 0.0116893540283
ML4T-310:
Correlation: 0.90122737953
RMS: 0.0187460682843

<u>Why the method worked well</u>

The features selected, that is, Phase, Amplitude and Frequency has impactful and sufficient information about the curve formed by the data. I have not used mean of data as a feature since we are predicting daily returns, that is independent of the mean value. For fitting of data to sine waveform, I have used FFT since it gives the frequency spectrum which best fits the dataset. I took the approach of choosing the maxima of frequency spectrum and enhancing amplitude which has lead to good prediction results for the data sets. KNN learner is more time efficient compared to Random Forest learner and also has given good prediction for the returns after five days (small time window) when using a relatively large number for k (I have used k = 50) to reduce chances of overfitting. In this way, my method for predicting the daily returns of stock prices after five days has produced good and speedy results even compared to Random Forest learner.