Machine Learning: Lab 2 – Exploratory Analysis for Time Series Data

Prerequisites: Python basics, numpy, pandas, matplotlib, seaborn, etc.

Download the CSV files of stocks.csv. The dataset contains the time series information about stock prices of 4 companies.

1. Import the data from the stocks.csv file. Keep the first row and first column as column and row index labels, respectively. Display the top 10 and bottom 10 rows and the total number of rows and columns.

2. How many rows are NA / NULL values.

3. Display the summary of the dataset (using describe()). Check the data types of each column, and see if the dataframe is in chronological order (w.r.t. “date” column).

4. Sort the dataframe into reverse chronological order (recent first). Do not sort on the String “date” type. You may convert into appropriate type and save the sorted dataframe into another variable.

5. Display the average stock prices for the different years and for all the 4 stocks. Plot it on a graph for comparison with other companies. Create appropriate legend by keeping different stocks represented by different coloured lines.

6. Compute moving average values (3 day moving average and 10 day moving average) for the different stocks. Create separate columns for each stock and moving average. Concatenate the new columns with the original dataframe after naming the columns. For 4 initial stocks, the resultant dataframe will have 8 new columns plus the original columns.

7. Find the average stock prices for each month for IBM and plot the trend between the years 1995 – 2010. Compare the trend with the Microsoft Corporation stocks (MSFT). Create appropriate legend and axis labels.

8. Plot the moving average values with actual values for the IBM stock for the year 2010. Show the moving averages as dashed line with appropriate labels and legend.