

A series of four teal-colored curved lines of varying lengths, arranged in a loose, upward-curving arc above the title.

Project Completion Assignment

Astha Kaushik



Problem Statement:

Stock-Time-Series-Analysis

Mathematical modeling for financial time series data.

Use LSTMs for predicting the price of stocks of IBM for the year 2017

The data is presented in a couple of formats keeping in mind the computational limitations. There are files containing 13 years of stock data (in the all_stocks_2006-01-01_to_2018-01-01.csv and corresponding folder) and a smaller version of the dataset (all_stocks_2017-01-01_to_2018-01-01.csv) with only the past year's stock data for those wishing to use something more manageable in size.

The folder individual_stocks_2006-01-01_to_2018-01-01 contains files of data for individual stocks, labelled by their stock ticker name. The all_stocks_2006-01-01_to_2018-01-01.csv and all_stocks_2017-01-01_to_2018-01-01.csv contain this same data, presented in merged .csv files.

Depending on the intended use (graphing, modelling etc.) the user may prefer one of these given formats.

All the files have the following columns:

Date - in format: yy-mm-dd

Open - price of the stock at market open (this is NYSE data so all in USD)

High - Highest price reached in the day

Low Close - Lowest price reached in the day

Volume - Number of shares traded

Name - the stock's ticker name