

## **Forecasting a Stock Price - IBM Advanced Data Science Capstone Project.**

*by Wany Fourreau.*

### **Forecasting a Stock Price using Linear Models such as LinearRegression, K-Nearest Neighbor(KNN), Support Vector Machine(SVM) and using RNN-LSTM as a Time Series Forecasting Model.**

*Disclaimer: The reader is not encouraged to invest in the stock market or attempt to trade based upon anything that is said or any models that the author has discussed in the paper. The equity market is a risky place and randomness is the best guarantee. This is simply an attempt to understand Advanced Data Science Modeling Process.*

## **Introduction**

### **A short understanding of my interpretation of market psyche.**

Why should we try to predict a stock price?

In an attempt to minimize risk and maximize profit investors and traders alike want to have an edge. Within that philosophy predictive models are created to serve as forecasting indicators which are assumed to provide that edge. In the financial world or to be more precise, at a trading firm, the term often used is “technical analysis” for predicting short term moves, for example in a stock, an index, a commodity or others. The term “Fundamental analysis” is used as well but it is usually applied to understanding the health of a company and for long term investments. Technical analysis is often used as short-term indicators to give the trader an edge. In my personal view even when a model does not work a trader wants to have one to follow because it makes the transaction appear less random. It also serves as a tool to keep emotions out of the trade. After all trading is an emotional affair but do pretend I did not tell you that since this is a scientific project for a data science class.

### **The Modeling Aspect - The Edge**

Predictive models not always but are often applied as short-term indicators. Many models are available in Machine Learning which can serve as stock market forecasting tools. What we must understand is no models can take the place of good governance over money management. Here however in my Advanced Data Science Capstone Project I would like to see how a few Machine Learning predictive models work in comparison to a stock’s actual price performance. My goal is to submit a Linear Regression model, a K-Nearest Neighbor model, a Support Vector Machine model and an RNN-LSTM Time Series Forecasting model using the company Amazon.com stock’s historical price data.

## **An overall understanding about the models I have selected.**

I will make the answers as short and as simple as possible. If you desire deeper input please go to [wikipedia.com](https://www.wikipedia.com) or a data science/machine learning website like [Towardsdatascience.com](https://towardsdatascience.com) or [machinelearningmastery.com](https://machinelearningmastery.com)

### **What is a Linear Regression Model?**

To give you a simple answer. Linear regression makes an attempt at modeling the relationship between two variables by fitting a linear equation to observed data where one variable is considered to be an explanatory variable and the other is considered to be a dependent variable. It is the most basic and commonly used type of predictive analysis.

### **What is a K-Nearest Neighbor (KNN) Model?**

Medium.com put it in the simplest term possible. KNN is a supervised learning classification and regression algorithm that uses nearby points in order to generate a prediction. It is said to be one of the most basic yet essential classification algorithms in Machine Learning.

### **What is a Support Vector Machine(SVM) model?**

Similar to the KNN model, SVM is a supervised machine learning model that can be used for classification and regression. SVM is effective in high dimensional spaces. It is effective in cases where the number of dimensions is greater than the number of samples. It is memory efficient and is versatile.

### **What is a RNN-LSTM Time Series Forecasting model?**

RNN stands for recurrent neural network. LSTM stands for long short-term memory network. Since time series prediction problems are difficult types of predictive modeling problems, recurrent neural networks are used in deep learning to train very large architectures successfully.

## **Files Organization**

- **Forecasting a Stock Move Steps\_1\_to\_5.inpg** - The file contains steps 1 through step 5. It shows the viewer where the Source of data is. How it is Extracted, Transformed and Loaded. What feature engineering took place before modeling is created and compiled.
- **Forecasting a Stock Move Linear Models.inpg** - The file focuses on the modeling process of the Linear Models. It provides the viewer with steps 1 to 5 in case the viewer escapes that file then it continues to model building, evaluation, performance and observation.
- **Forecasting a Stock Move with LSTM Time Series For.inpg** - The interesting thing with this deep learning model is that you can actually use raw data. Whereas with the linear model certain modifications with the date had to be made, here that does not apply. In model building, training and evaluation access to tensor flow for backend and keras for the compiling is needed.
- **Forecasting a Stock Move with LSTM Time Series For copy1.inpg - Can I improve the Performance of the RNN- LSTM Model ?** - This file is where I focus on trying to improve the LSTM model by making changes on how it is compiled to train. The major change is doubling the number of epochs from 50 to 100.

- **Forecasting a Stock Move with LSTM Time Series For copy2.inpg** - Since I am recommending the RNN-LSTM model here I am using a new set of recent data to test the model again. I used that new data as well in the file above which is **Forecasting a Stock Move with LSTM Time Series For copy1.inpg**. I want to see how the new test set compares from the original model to the performance it shows in the improved model.
- **Model Evaluation Metrics.pdf** - This file shows the evaluation metrics for the four models.
- **Model Deployment.pdf** - Deploying my preferred model
- **Data Product.pdf** - An overall summarization of the project.
- **The ADD** - Work documentation.
- **Video Presentation**