

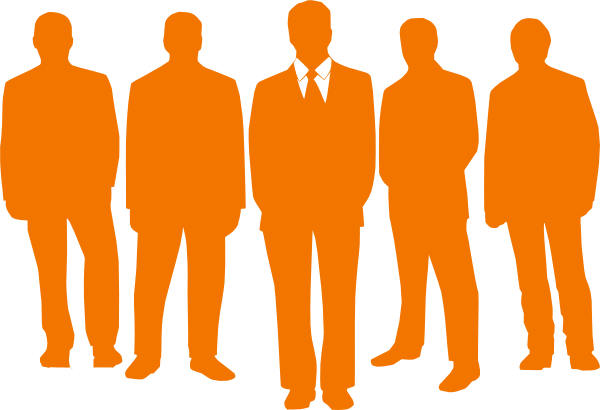
STOCK PRICE PREDICTION

NIVT

Mentor: Tania Chakraborty

summer training  KOLKATA

Group



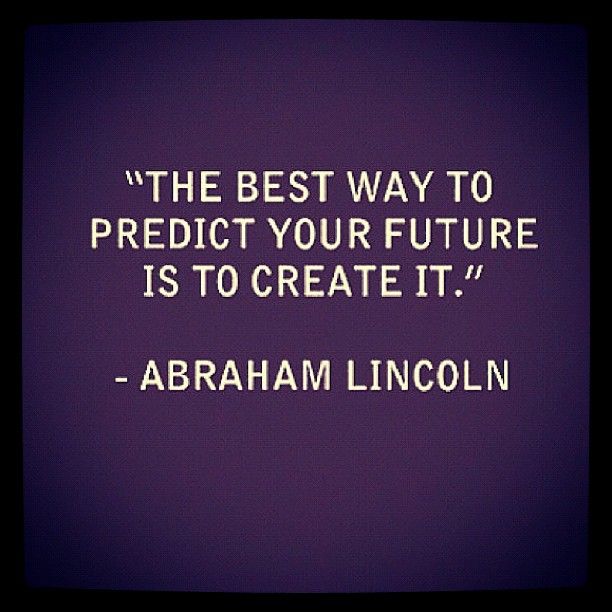
* Prasenjit Karmakar
* Sagar Roy
* Yasar Reza Rahaman

ACKNOWLEDGEMENT

We would like to express my special thanks of gratitude to my teacher Tania Chakraborty as well as our institute NIVT who gave us the golden opportunity to do this wonderful project on the topic STOCK PRICE PREDICTIONS which also helped us to learn so many new things. I am really thankful to them.Secondly, I would also like to thank our friends & group members who helped us a lot in finalizing this project within the limited time frame.

CONTENTS

* INTRODUCTION
* HISTORY OF STOCK PRICE PREDICTIONS
* WHAT IS DATA SCIENCE?
* MEANING OF DATA SCIENCE
* DEFINE DATA SCIENCE COMPONENTS
* DATA
* TYPES OF DATA
* DATA ANALYTICS
* TYPES OF DATA ANALYTICS
* PROCESS OF ANALYZING THE DATA IN DATA SCIENCE
* DATA MANAGEMENT TOOLS
* TOP 10 PYTHON LIBRARIES
* JUPETER NOTEBOOK
* DATA SCIENCE IN STOCK ANALYSIS
* USE OF DATA SCIENCE IN STOCK MARKET
* DTYPES OF GIVEN RAW DATA
* DATA DESCRIPTION
* OUR APPROACH TO THE PROBLEM
* TRAINING AND TESTING DATA SET
* OUR RNN MODEL
* TRAINING OR LOADING OUR MODEL
* TESTING OUR MODEL
* WHAT IS BETTER APPROACH?
* DECISION MAKING
* BENEFITS
* CONCLUSION

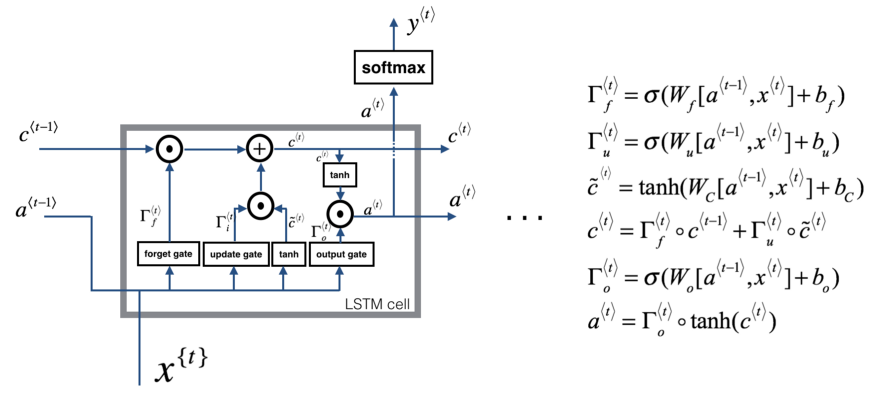


INTRODUCTION

There are a lot of complicated financial indicators and also the fluctuation of the stock market is highly violent. However, as the technology is getting advanced, the opportunity to gain a steady fortune from the stock market is increased and it also helps experts to find out the most informative indicators to make a better prediction. The prediction of the market value is of great importance to help in maximizing the profit of stock option purchase while keeping the risk low. Recurrent neural networks (RNN) have proved one of the most powerful models for processing sequential data. Long Short-Term memory is one of the most successful RNNs architectures. LSTM introduces the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network. With these memory cells, networks are able to effectively associate memories and input remote in time, hence suit to grasp the structure of data dynamically over time with high prediction capacity.

HISTORY OF STOCK PREDICTIONS

*The art of forecasting stock prices has been a difficult task for many of the researchers and analysts. In fact, investors are highly interested in the research area of stock price prediction. For a good and successful investment, many investors are keen on knowing the future situation of the stock market. Good and effective prediction systems for stock market help traders, investors, and analyst by providing supportive information like the future direction of the stock market. In this work, we present a recurrent neural network (RNN) and Long Short-Term Memory (LSTM) approach to predict stock market indices.*

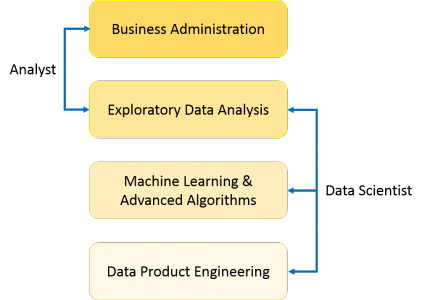


WHAT IS DATA SCIENCE?

To manipulate data and extract the important part of data is called Data Science.

Data science is a multidisciplinary blend of **data inference, algorithm development, and technology** in order to solve analytically complex problems.

At the core is data. Troves of raw information, streaming in and stored in enterprise data warehouses. Much to learn by mining it. Advanced capabilities we can build with it.



MEANING OF DATA SCIENCE

Data science is the field of study that combines domain expertise, programming skills, and knowledge of math and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems that perform tasks which ordinarily require human intelligence. In turn, these systems generate insights that analysts and business users translate into tangible business value.



DEFINE DATA SCIENCE COMPONENTS

Basically, here three component of data science-

* Data Management
* Data Analytics
* Machine Learning

**Data Management:**

**Data Management** is a comprehensive collection of practices, concepts, procedures, processes, and a wide range of accompanying systems that allow for an organization to gain control of its **data** resources.

* **Data** Storage and Big **Data**.
* Business Intelligence and Analytics.
* Metadata **Management**.

**Data Analytics:**

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.

**Machine Learning:**

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of ML is to allow computer systems to learn from experience without being explicitly programmed or human intervention.

DATA

Data is a set of values of qualitive or quantitative variables. It is information in raw or unorganized form. It may be fact, figure, character symbols etc.

* Qualitative -Categorical or Nominal:
  + Discrete Data
  + Continuous Data
* Quantitative -Measurable or Countable:
  + Attribute
  + Nominal
  + Ordinal

Information:

Meaningful or organized data is information.

Big Data:

Extremely large data sets that may be analyzed computationally to reveal patterns, trends and associations, especially relating to human behavior and interactions.

Design distributed systems that manage big data using HADOOP and related technologies.

Big data cannot manage by RDBMS

TYPES OF DATA

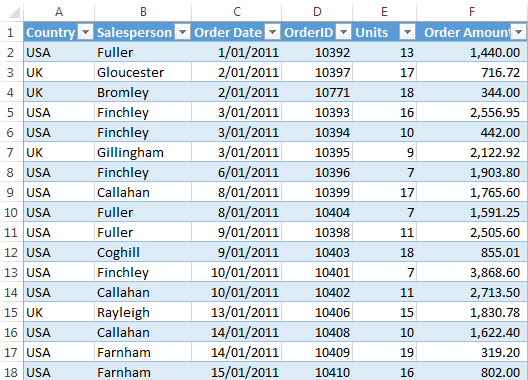
There are three types of dataset-

1. STRUCTURED DATA
2. SEMI-STRUCTURED DATA
3. UNSTRUCTURED DATA

**Structured Data:**

Data which can be stored in database SQL in table with rows and columns are called structured data.

Only 5 to 10 % of all informatics data.

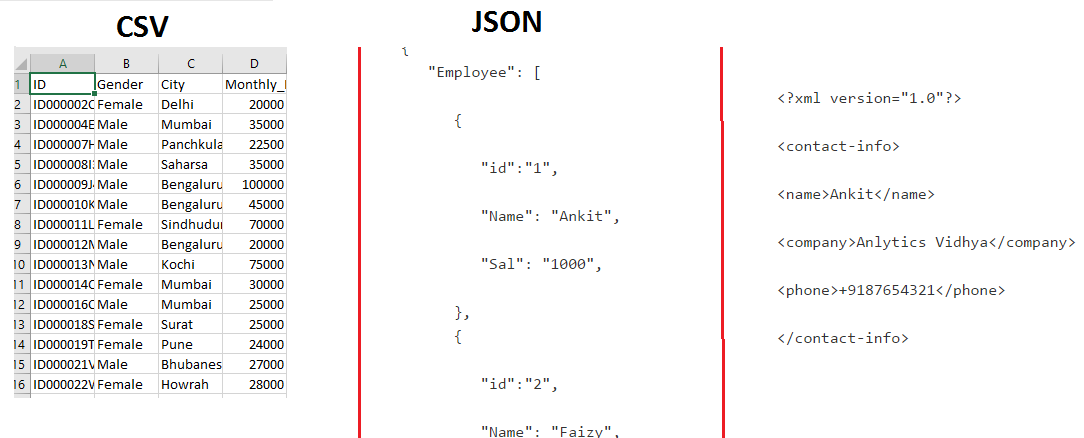


**Semi-Structured Data:**

Doesn’t reside in a relational database but that does have some organizational properties that make it easier to analyze.

CSV, XML AND JSON documents are semi-structured documents, NoSQL databases are considered as semi-structured.

A few parts of data(5 to 10 %).



**Unstructured Data:**

Unstructured data represent around 80% of data. It often includes text and multimedia content.

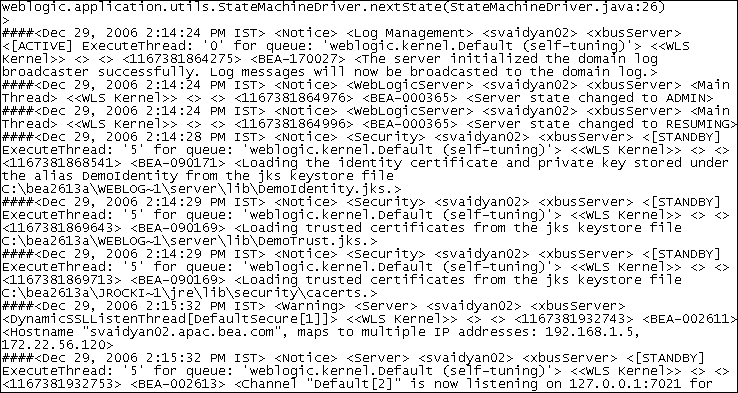
Example: e-mail messages, word processing documents, videos, photos, audio files, presentation, webpages and many other kinds of business documents.

Some example of machine-generated unstructured data :

* Satellite images
* Scientific data
* Photographs and video
* Rader or Sonar data

Some example of human-generated unstructured data:

* Text internal
* Social media data
* Mobile data
* Website content



DATA ANALYTICS



Data Analytics is the process of examining data sets in order to draw conclusion about the information it contains.

increasingly with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses.

Analytics is not a tool of technology, rather it is the way of thinking and acting on data.

TYPES OF DATA ANALYTICS

Data analytics are of three types those are: -

1. Descriptive Analytics
2. Predictive Analytics
3. Prescriptive Analytics

**Descriptive Analytics:**

90% of organizations today use descriptive analytics which is the most basic form of analytics. The simplest way to define descriptive analytics is that, it answers the question “What has happened?”. This type of analytics, analyses the data coming in real-time and historical data for insights on how to approach the future. The main objective of descriptive analytics is to find out the reasons behind precious success or failure in the past. The ‘Past’ here, refers to any particular time in which an event had occurred and this could be a month ago or even just a minute ago. The vast majority of big data analytics used by organizations falls into the category of descriptive analytics.

**Predictive Analytics:**

The subsequent step in data reduction is predictive analytics. Analyzing past data patterns and trends can accurately inform a business about what could happen in the future. This helps in setting realistic goals for the business, effective planning and restraining expectations. Predictive analytics is used by businesses to study the data and ogle into the crystal ball to find answers to the question “What could happen in the future based on previous trends and patterns?”

**Prescriptive Analytics:**

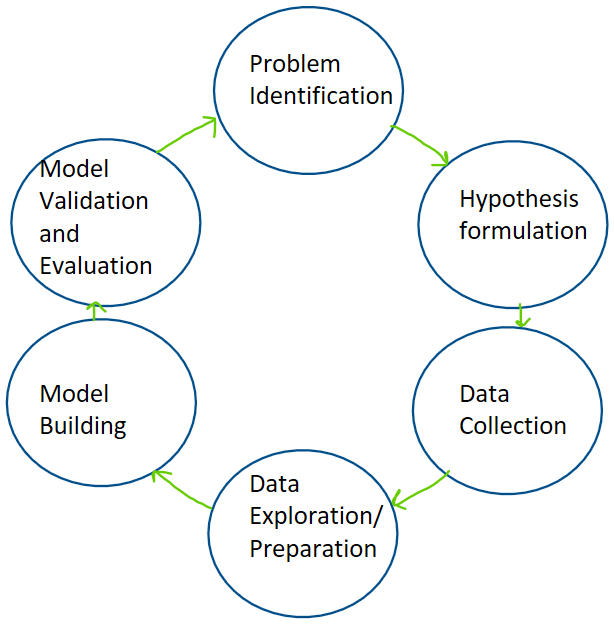
Big data might not be a reliable crystal ball for predicting the exact winning lottery numbers but it definitely can highlight the problems and help a business understand why those problems occurred. Businesses can use the data-backed and data-found factors to create prescriptions for the business problems, that lead to realizations and observations.

PROCESS OF ANALYZING THE DATA IN DATA SCIENCE

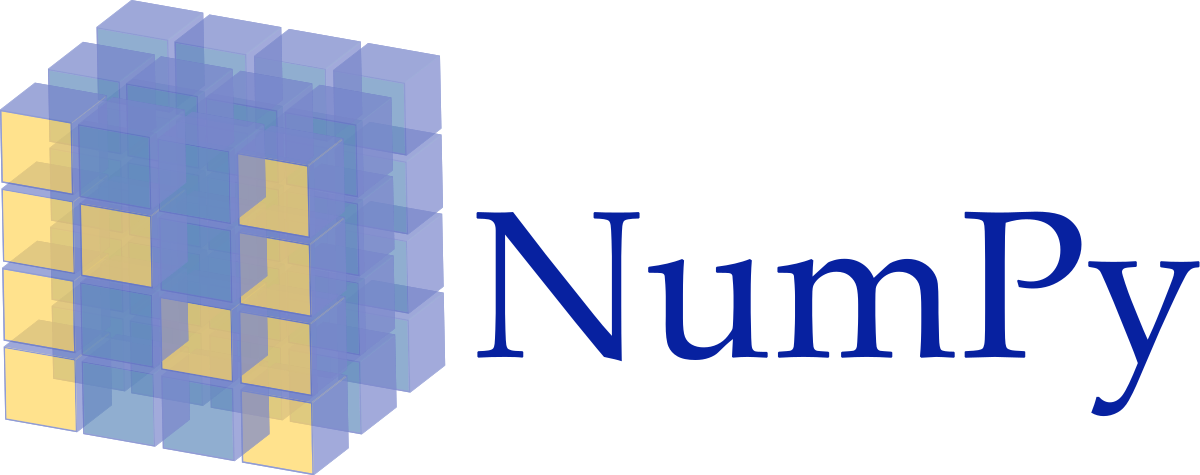
Data goes through various process during data analytics in data Science. Those are given below

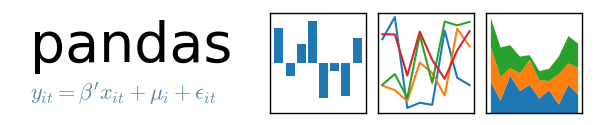
**DATA SCIENCE PROCESS OR LIFE CYCLE**

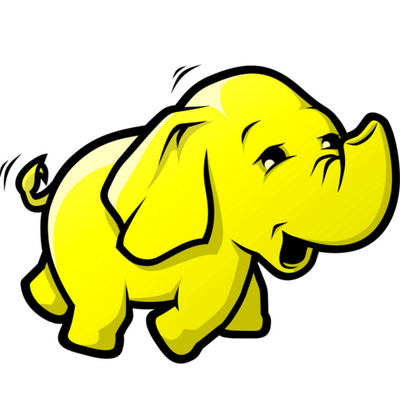
1. Problem Identification
2. Hypothesis formulation
3. Data Collection
4. Data Exploration/Preparation
5. Model Building
6. Model Validation and Evaluation



DATA MANAGEMENT TOOLS

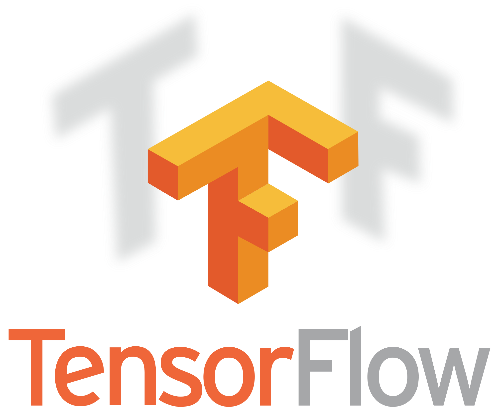


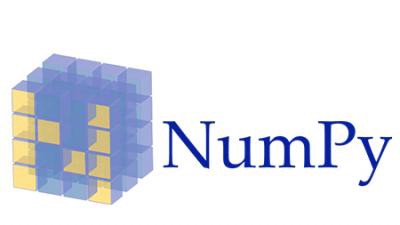


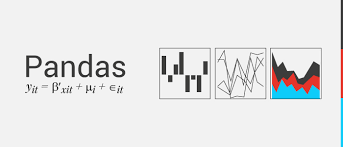


TOP 10 PYTHON LIBRARIES

One of the reason PYTHON Is mostly popular among developers is its wide range of libraries. We can’t even count how many libraries it has but We will be considering the following 10 libraries:

* NUMPY
* PANDAS
* MATPLOTLIB
* SEABORN
* IPYTHON
* TENSORFLOW
* SCIKIT-LEARN
* KERAS
* PYTORCH
* SCIPY

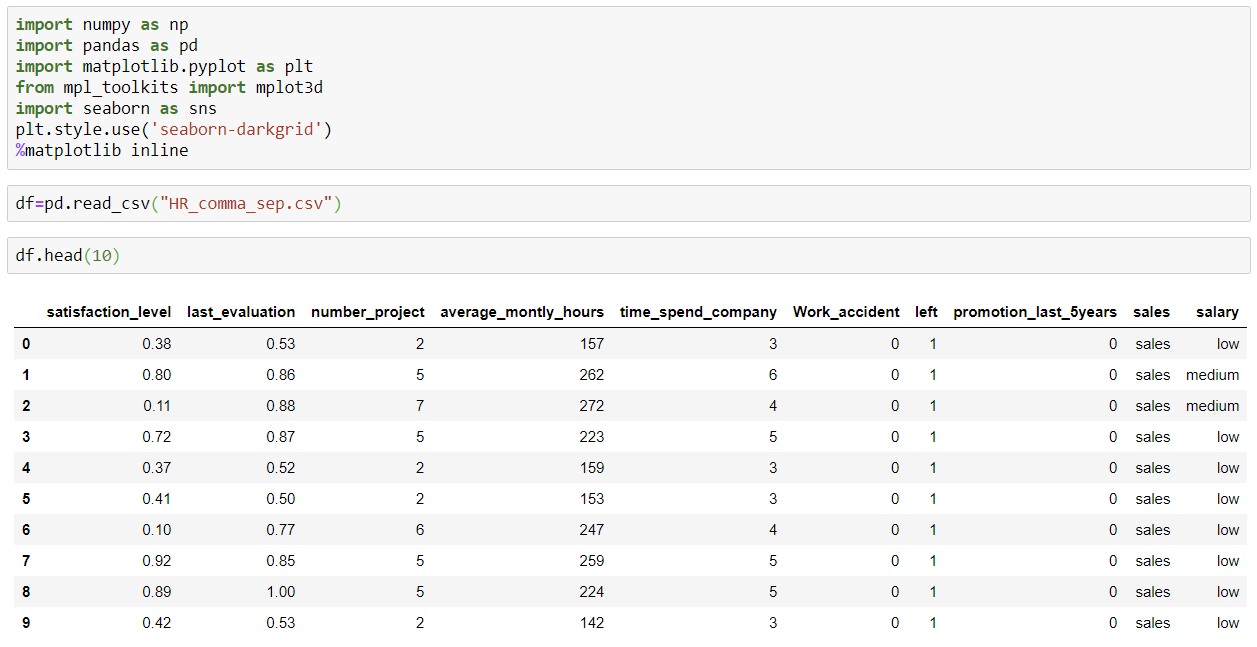
 



JUPETER NOTEBOOK

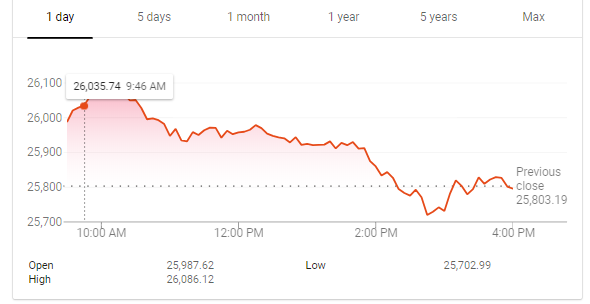
Jupiter Notebook is an interface we use to write python programs. It is Widely popular because most of the popular python libraries are pre-installed in it.

And we can write and run part of a code written in a code cell. So, it is very helpful for debugging.



DATA SCIENCE IN STOCK ANALYSIS

Trying to predict the stock market is an enticing prospect to data scientists motivated not so much as a desire for material gain, but for the challenge. We see the daily up and downs of the market and imagine there must be patterns we, or our models, can learn in order to beat all those day traders with business degrees. Naturally, when I started using additive models for time series prediction, I had to test the method in the proving ground of the stock market with simulated funds. Inevitably, I joined the many others who have tried to beat the market on a day-to-day basis and failed. However, in the process, I learned a ton of Python including object-oriented programming, data manipulation, modeling, and visualization. I also found out why we should avoid playing the daily stock market without losing a single dollar (all I can say is play the long game)!



One day vs 30 years: which would you rather put your money on?

When we don’t experience immediate success — in any task, not just data science — we have three options:

1. Tweak the results to make it look like we were successful
2. Hide the results so no one ever notices
3. Show all our results and methods so that others (and ourselves) can learn how to do things better

USE OF DATA SCIENCE IN STOCK MARKET

## Machine Learning and Data Analytics are making trading much more efficient. Together, they complement each other and act as catalysts towards improved ability to identify opportunities and reduce trading costs.

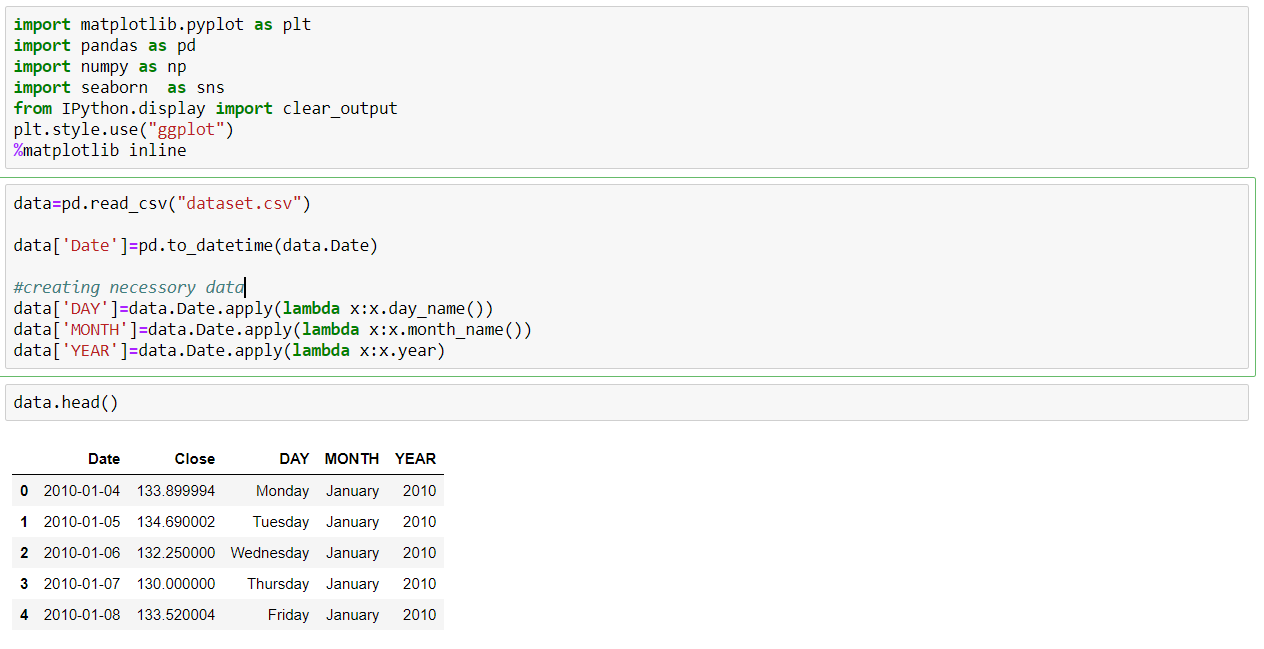


During the research phase, the ability to analyze and learn effectively from the vast amount of data gives the strategies an edge.

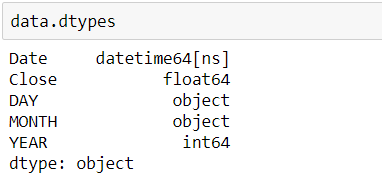
Technology is scaling at an exponential rate and today we are processing vast loads of data if numbers are to be believed. A recent report reveals that the total data existing in the world will grow at a CAGR of 61% to 175 zettabytes by 2025 from 33 zettabytes in 2018. These numbers are huge!

Today, Machine Learning and Data Analytics are making trading much more efficient. Together, they complement each other and act as catalysts towards improved ability to identify opportunities and reduce trading costs. Leveraging data in trading at present comes mostly in two flavours. During the research phase, the ability to analyse and learn effectively from the vast amount of data gives the strategies an edge. Here the analysis is mostly exploratory and speed is important but not critical. During the execution phase, these same tools can be used to gain the ability to quickly react to changing market conditions. Here speed is critical to success.

DTYPES OF GIVEN RAW DATA

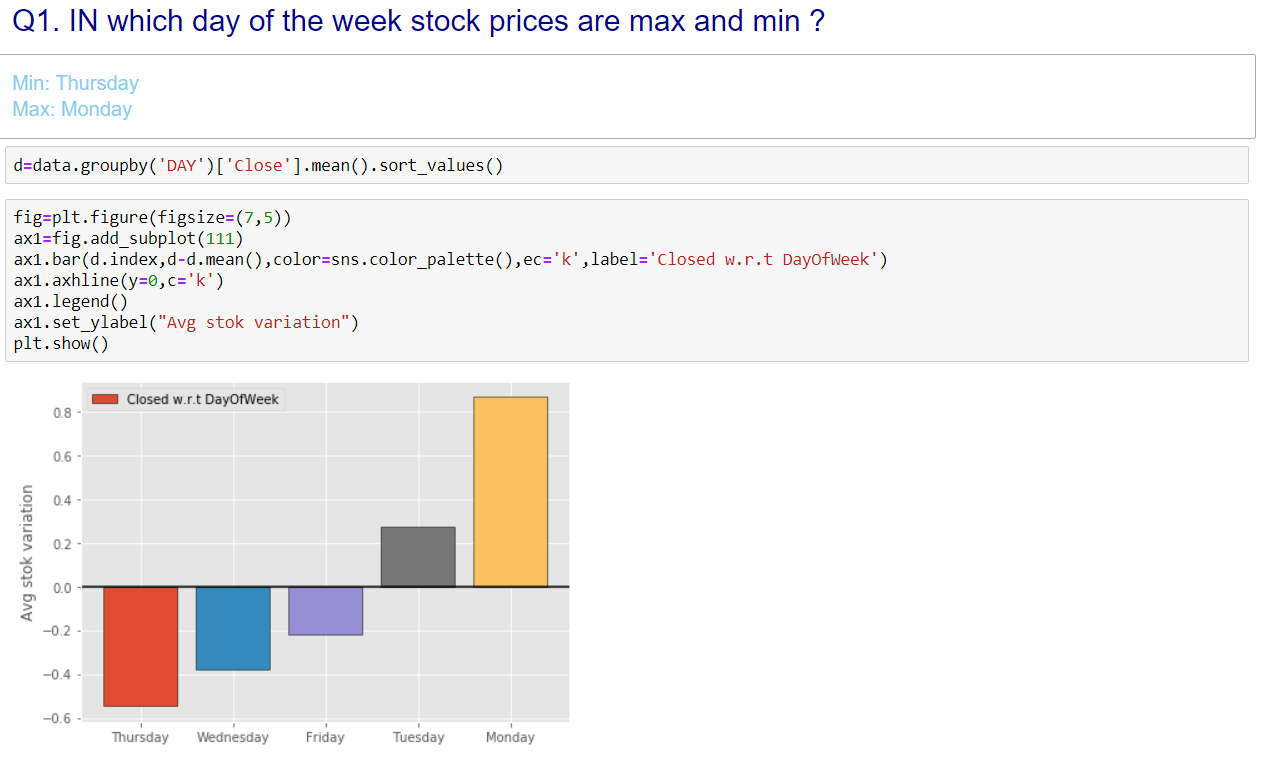
This dataset originally belongs to AMAZON.

It has 5 columns and 1762 rows for their employees.



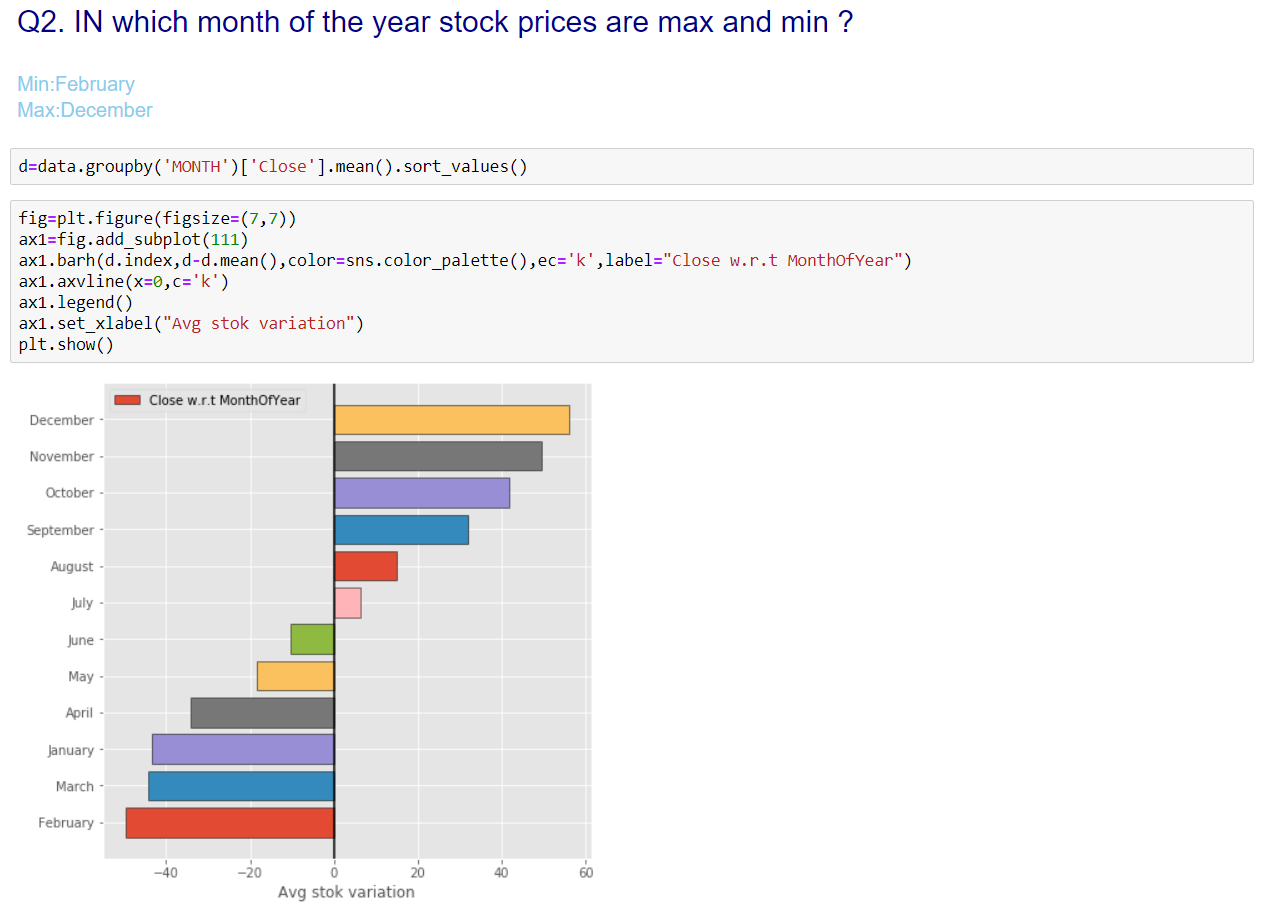
The columns represent date, closing stock price, day of weak month, year of a given working day in amazon’s stock market data.

DATA DESCRIPTION

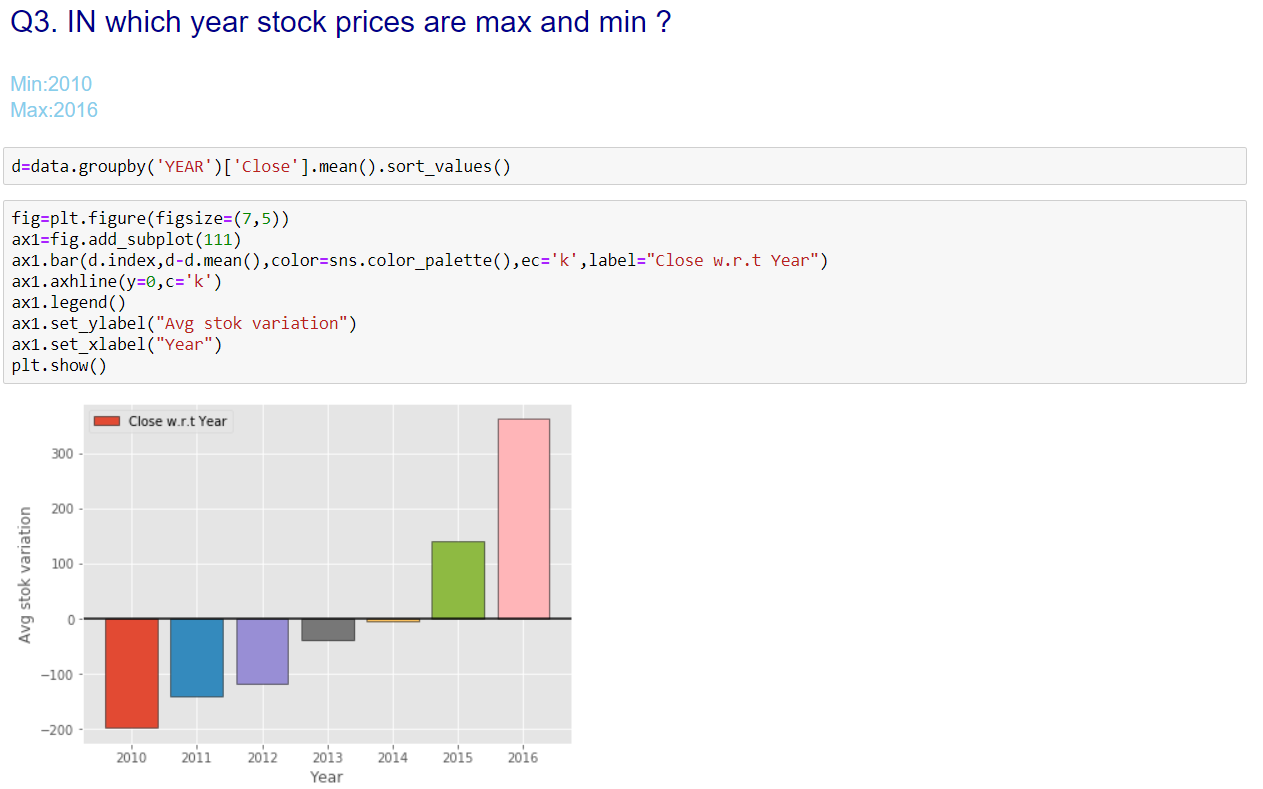


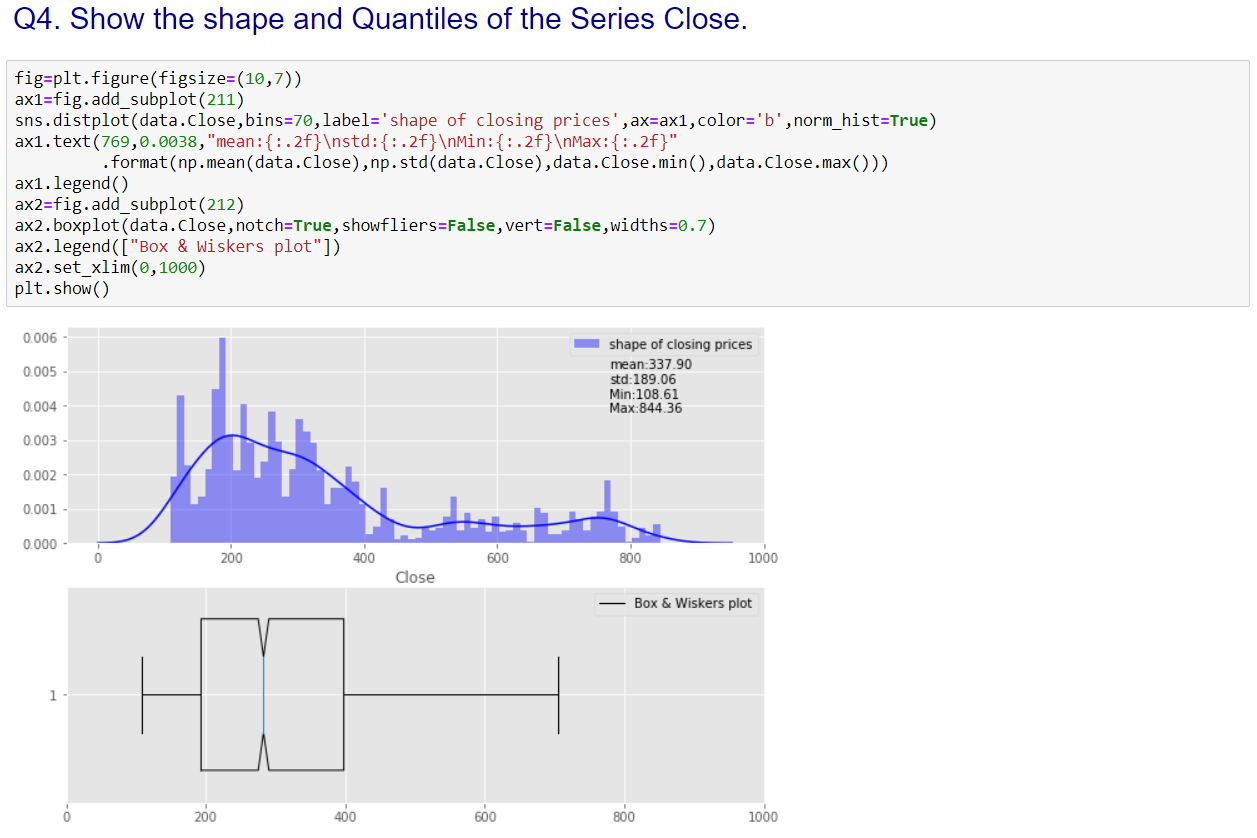
In this graph we can see that in Monday the sock prices are maximum and in Thursday it is minimum. So, it will be optimal for a person to buy stocks in Thursday and sell them in Monday.

The bar chart shows the corresponding chance of increasing or decreasing chance of every working day in Amazon.

In this graph we can see that in December the sock prices are maximum and in February it is minimum. So, it will be optimal for a person to buy stocks in February and sell them in December.

The bar chart shows the corresponding chance of increasing or decreasing chance of every month in Amazon.

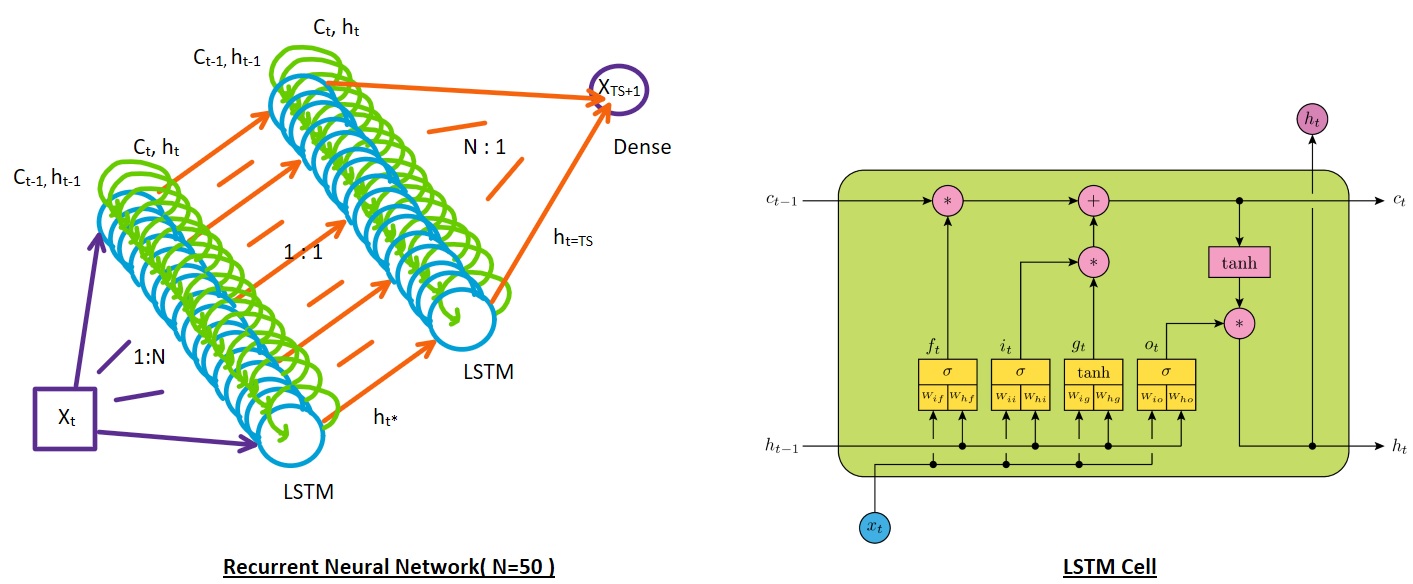
this graph shows the gradual increase of stock prices over the year range 2010 to 2016 for Amazon. So, we can conclude that the company is improving its economic conditions.



This graph represents how the stock data is distributed and give a general understanding of the expectation and variance of the stocks. We can see that its mean is 337.90 and standard deviation is 189.06.

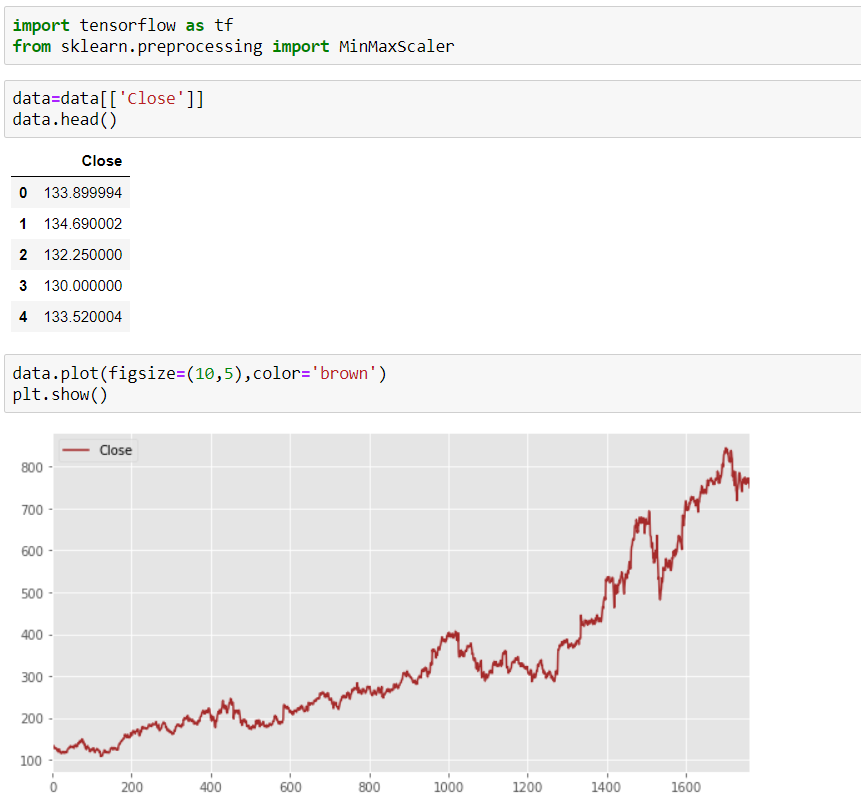
OUR APPROACH TO THE PROBLEM

LSTM cell and RNN:

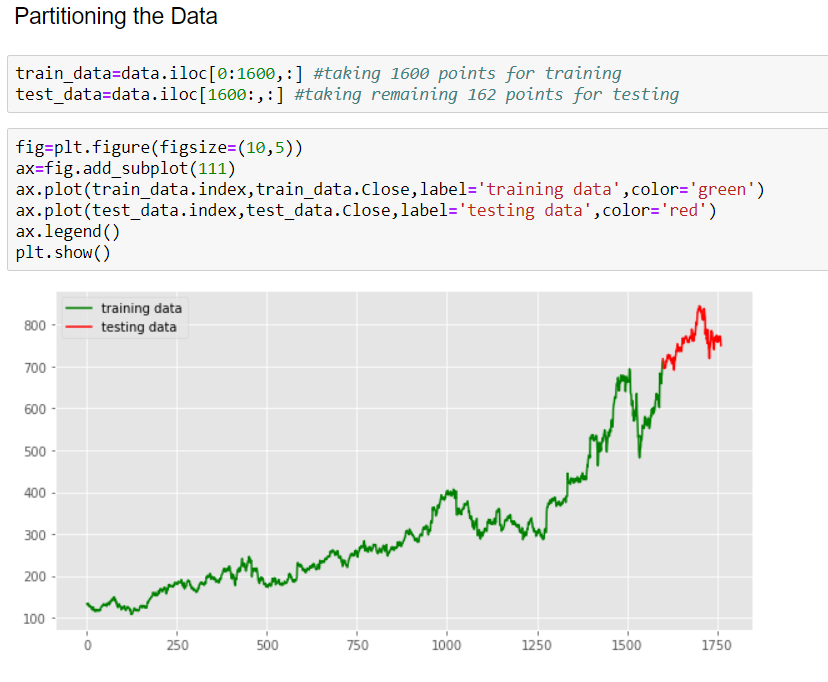


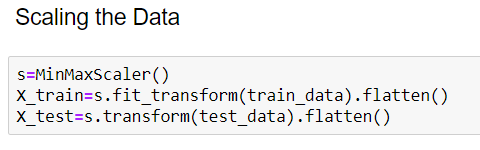
Goal: Our goal is to predict the next stock price from previous 10 days stock prices.

TRAINING AND TESTING DATA SET

whole data:

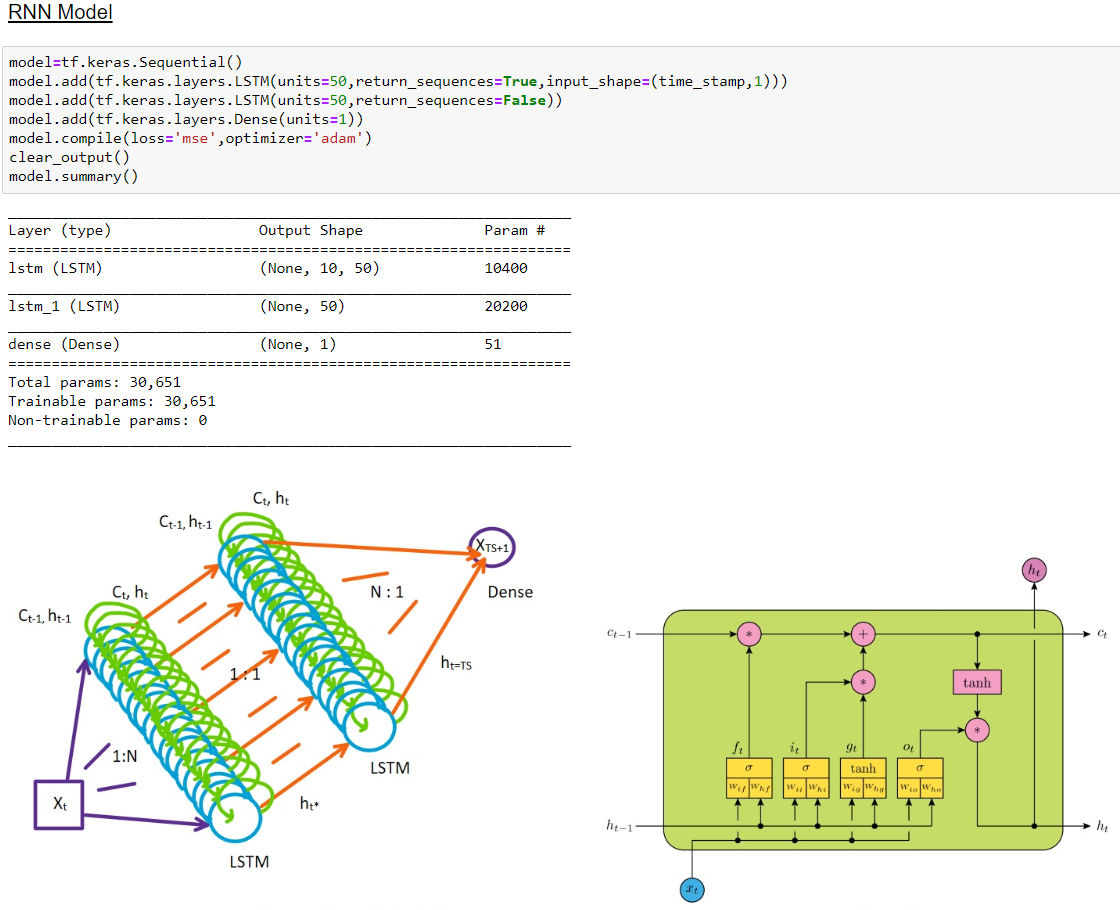
SPLITING THE DATA:

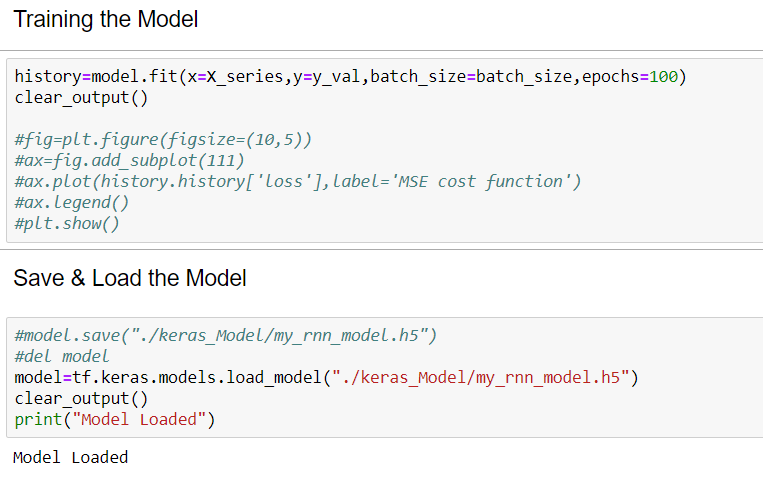






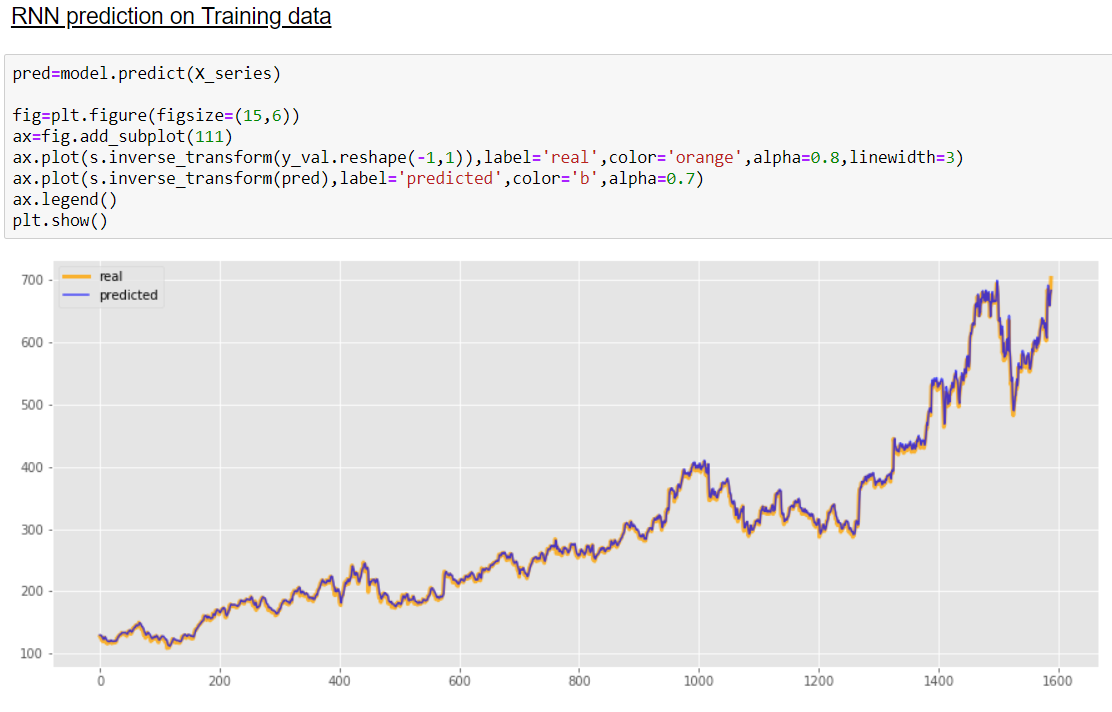
OUR RNN MODEL

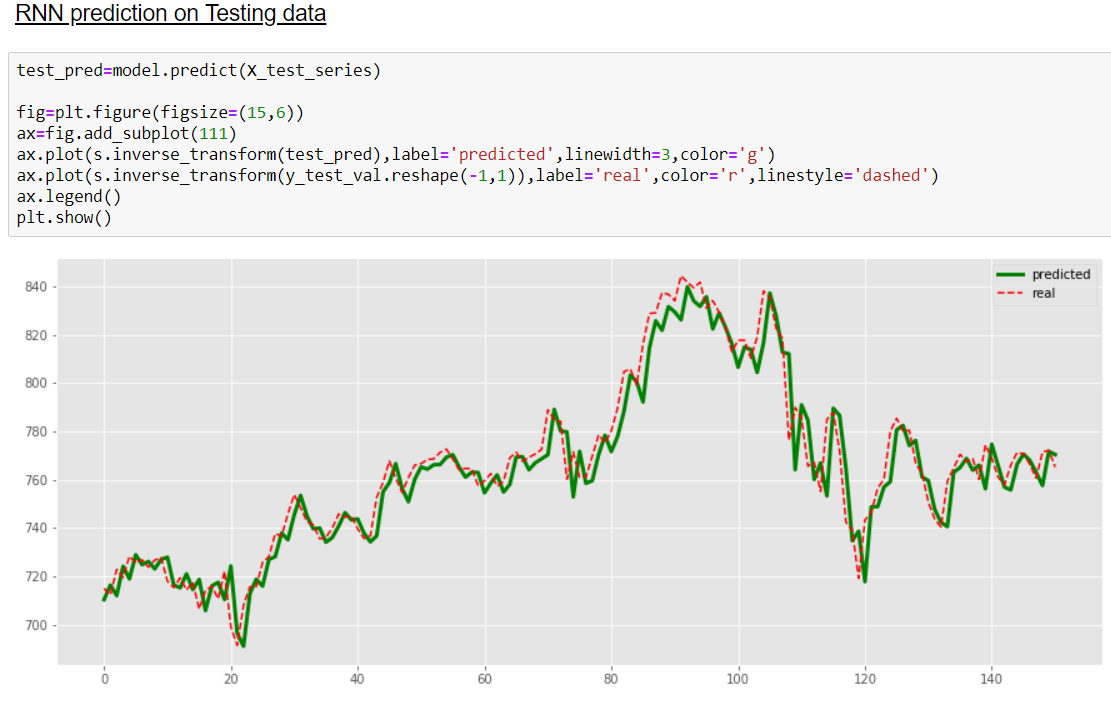


TRAINING OR LOADING OUR MODEL

TESTING OUR MODEL

ON TRAINING DATA:



ON TESTING DATA:

WHAT IS BETTER APPROACH?

RNNs in general and LSTMs in particular have received the most success when working with sequences of words and paragraphs, generally called natural language processing.

This includes both sequences of text and sequences of spoken language represented as a time series. They are also used as generative models that require a sequence output, not only with text, but on applications such as generating handwriting.

**Use RNNs for:**

Text data

Speech data

Classification prediction problems

Regression prediction problems

Generative models

Recurrent neural networks are not appropriate for tabular datasets as you would see in a CSV file or spreadsheet. They are also not appropriate for image data input.

**Don’t Use RNNs for:**

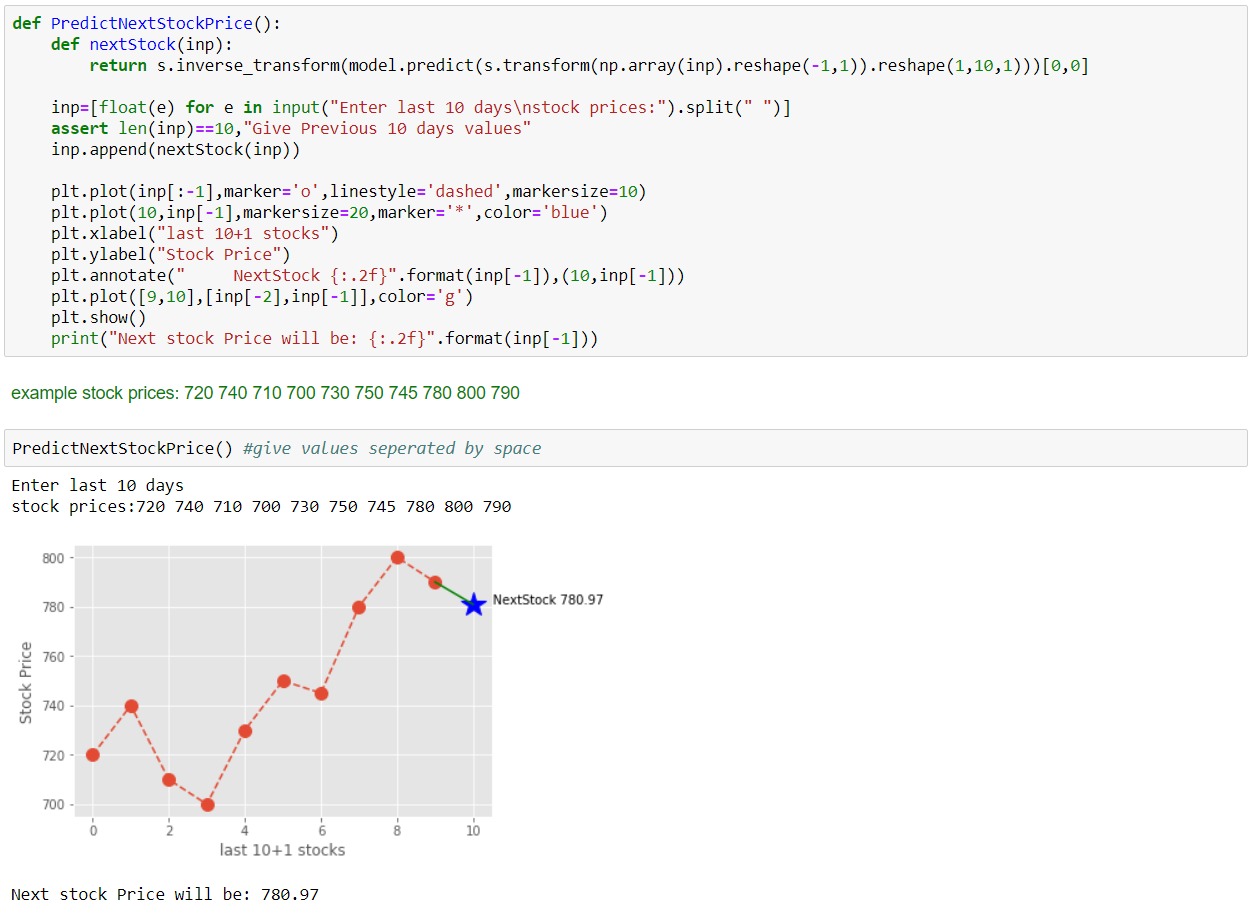
Tabular data

Image data

RNNs and LSTMs have been tested on time series forecasting problems, but the results have been poor, to say the least. Autoregression methods, even linear methods often perform much better. LSTMs are often outperformed by simple MLPs applied on the same data.

DECISION MAKING

Now from the RNN we can predict next day stock prices from the previous 10 days stock price value. This have a great accuracy we see on the testing data. So, we are confident that it will work on current real data also. So, by using this advanced statistical approach we can successfully predict stocks and increase our economic assets.



BENEFITS

* Using This we can successfully predict stocks of a given company.
* We can invest wisely and make money.
* The company can also take economic decisions based on the prediction.
* With this we can successfully simulate a company’s economic behavior.
* We can create jobs under this.
* Help to make educated gushes of the future.

CONCLUSION

*The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required. In this work, we have used one of the most precise forecasting technologies using Recurrent Neural Network and Long Short-Term Memory unit which helps investors, analysts or any person interested in investing in the stock market by providing them a good knowledge of the future situation of the stock market.*

END