**1.INTRODUCTION**

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.

 stock prices are a function of information and rational expectations, and that newly revealed information about a company's prospects is almost immediately reflected in the current stock price. This would imply that all publicly known information about a company, which obviously includes its price history, would already be reflected in the current price of the stock. Accordingly, changes in the stock price reflect release of new information, changes in the market generally, or random movements around the value that reflects the existing information set. Burton Malkiel, in his influential 1973 work A Random Walk Down Wall Street, claimed that stock prices could therefore not be accurately predicted by looking at price history. As a result, Malkiel argued, stock prices are best described by a statistical process called a "random walk" meaning each day's deviations from the central value are random and unpredictable. This led Malkiel to conclude that paying financial services persons to predict the market actually hurt, rather than helped, net portfolio return. A number of empirical tests support the notion that the theory applies generally, as most portfolios managed by professional stock predictors do not outperform the market average return after accounting for the managers' fees.

2.TECHNICAL ANALYSIS

This project deals with the IT companies(Tech Mahindra ,Infosys ,Oracle , Tata consultancy services) and analyzing there stock prices and using machine learning algorithms to predict the future price of the stock with the influence of macro factors like GDP, Inflation, Consumer Confidence, Unemployment.

2.1 packages used

**pandas**

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any language.

pandas is built on top of NumPy and is intended to integrate well within a scientific computing environment with many other 3rd party libraries.

**numpy**

Numpy is the core library for scientific computing in Python. It provides a high-performance multidimensional array object, and tools for working with these arrays. If you are already familiar with MATLAB, you might find this tutorial useful to get started with Numpy

**matplotlib.pyplot**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.

**sklearn.svm**

**sklearn.linear\_model**

logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

**2.2 machine learning algorithm**

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

**2.2.1Support vector machine**

In machine learning, support-vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall. When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups.

SVMs can be used to solve various real-world problems:

SVMs are helpful in text and hypertext categorization, as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.[citation needed] Some methods for shallow semantic parsing are based on support vector machines.

Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also true for image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.

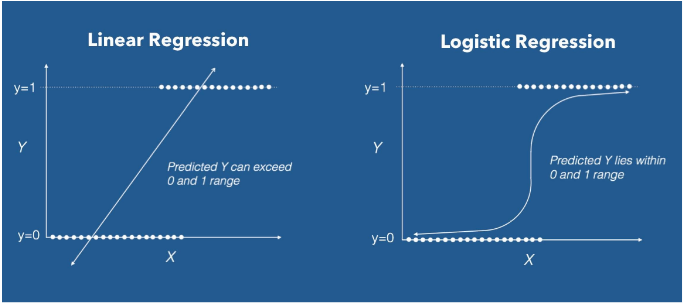
Hand-written characters can be recognized using SVM.

The SVM algorithm has been widely applied in the biological and other sciences. They have been used to classify proteins with up to 90% of the compounds classified correctly. Permutation tests based on SVM weights have been suggested as a mechanism for interpretation of SVM models. Support-vector machine weights have also been used to interpret SVM models in the past. Post hoc interpretation of support-vector machine models in order to identify features used by the model to make predictions is a relatively new area of research with special significance in the biological sciences.

2.2.2 Logistic regression

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.

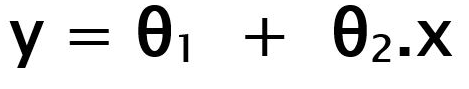
We can call a Logistic Regression a Linear Regression model but the Logistic Regression uses a more complex cost function, this cost function can be defined as the ‘Sigmoid function’ or also known as the ‘logistic function’ instead of a linear function.



Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.



2.2 Interpolation

Interpolation is the process of finding a value between two points on a line or a curve. To help us remember what it means, we should think of the first part of the word, 'inter,' as meaning 'enter,' which reminds us to look 'inside' the data we originally had. This tool, interpolation, is not only useful in statistics, but is also useful in science, business, or when there is a need to predict values that fall within two existing data points.

Pandas dataframe.interpolate() function is basically used to fill NA values in the dataframe or series. But, this is a very powerful function to fill the missing values. It uses various interpolation technique to fill the missing values rather than hard-coding the value.

Syntax

DataFrame.interpolate(method=’linear’, axis=0, limit=None, inplace=False, limit\_direction=’forward’)

There are directions in interpolation

1.forward

2.backward

3.both

1.forward interpolation

In this interpolation is done in the forward direction that it fills all NA values

2.backward interpolation

This method is used when the in the middle and different to fill the NA values.

3.both direction interpolation

This is the best method in interpolation it interpolates using above and below values according to that it fill the NA values.

RESAMPLING

Pandas dataframe.resample() function is primarily used for time series data.

A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. It is a Convenience method for frequency conversion and resampling of time series. Object must have a datetime-like index (DatetimeIndex, PeriodIndex, or TimedeltaIndex), or pass datetime-like values to the on or level keyword.

Resampling generates a unique sampling distribution on the basis of the actual data. We can apply various frequency to resample our time series data. This is a very important technique in the field of analytics.

Syntax:

DataFrame.resample(rule, how=None, axis=0, fill\_method=None, closed=None, label=None, convention=’start’, kind=None, loffset=None, limit=None, base=0, on=None, level=None)

Most commonly used time series frequency are –

W : weekly frequency

M : month end frequency

SM : semi-month end frequency (15th and end of month)

Q : quarter end frequency

There are many other types of time series frequency available.

There are some methods in resampling are:

Bfill: Backward fill

Count:Count of values

Ffill: Forward fill

First: First valid data value

last:Last valid data value

max:Maximum data value

mean: Mean of values in time range

Tech Mahindra

|  |
| --- |
|  |

Tech Mahindra Limited is a specialist in digital transformation, consulting and business re-engineering solutions and is a part of the Mahindra Group. Started its life in the year of 1986 as Mahindra-British Telecom Limited, for over two decades, Tech Mahindra has been the chosen transformation partner for wireline, wireless and broadband operators in Europe, Asia-Pacific and North America. As a leading provider of IT Solutions to the Telecom industry, the company committed to quality, Tech Mahindra adds value to client businesses through well-established methodologies, tools and techniques backed by its stringent quality processes. It is ISO 9001:2000 certified and is assessed at SEI-CMMI Level 5. Also awarded the ISO 20000-1 (IT Service Management standard) and ISO 27001 (Security Management standard) certification for its development centers across India and UK. Tech Mahindra is certified at PCMM Level 5 for its people-care practices and is the third company in the world to have been appraised for SSE-CMM Level 3. The company's global footprint spans 24 locations in 14 countries including 11 state-of-the-art development centres and 13 sales offices in Americas, Europe, Middle East, Africa and Asia-Pacific.  
  
  
On 23 January 2018, Tech Mahindra announced that it will now make available AT&T FlexWare, a transformative, global network infrastructure solution from AT&T Inc., to its global clients as well as use it internally. Tech Mahindra intends to combine AT&T FlexWare with its System Integration and Services Portfolio, and offer the solutions to its global clientele who are undergoing digital transformation.  
  
Mahindra Racing, the only Indian team to compete in the ABB FIA Formula E Championship, announced the formation of a strategic partnership with Pininfarina and Tech Mahindra on 24 January 2018. This agreement gives Mahindra Racing access to Pininfarina's globally renowned design expertise and Tech Mahindra's digital technology prowess.  
  
On 22 February 2018, Tech Mahindra announced a strategic investment of CAD 100 million over 5 years to establish a new Center of Excellence' (COE) in Canada. This strategic initiative will focus on major technologies such as Artificial Intelligence (AI) and Blockchain, which are driving innovation across industries and will cater to the exponentially growing need for AI and Blockchain application especially in the Fintech' and Smart Cities' spaces.

Infosys

|  |
| --- |
|  |

Infosys Ltd is a global technology services firm that defines, designs and delivers information technology (IT)-enabled business solutions to their clients. The company provides end-to-end business solutions that leverage technology for their clients, including technical consulting, design, development, product engineering, maintenance, systems integration, package-enabled consulting, and implementation and infrastructure management services.  
  
The company also provides software products to the banking industry. They have developed Finacle, a universal banking solution to large and medium size banks across India and overseas. Infosys BPO is a majority owned subsidiary. Through Infosys BPO, the company provides business process management services, such as offsite customer relationship management, finance and accounting, and administration and sales order processing. The company is having marketing and technical alliance with FileNet, IBM, Intel, Microsoft, Oracle and System Application Products.  
  
  
On 5 January 2018, Infosys announced that it had won a contract from Proximus, the largest telecommunications company in Belgium, to implement Excite - a business transformation program aimed at delivering superior digital customer experiences for its enterprise clients. The multi-year program will strengthen Proximus' leadership in the professional services market by replacing legacy IT systems, streamlining processes and deploying advanced tools for quoting, selling, ordering, billing, invoicing and more.  
  
  
On 9 January 2018, Infosys announced the successful conclusion of an Advance Pricing Agreement (APA) with the U.S. Internal Revenue Service (IRS). Under the APA, Infosys and the IRS have agreed on the methodology to allocate revenues and compute the taxable income of the company's US operations. This agreement covers financial years from 2011 to 2021. The APA will enhance predictability of the company's tax obligations in respect of its US operations.

Oracle

|  |
| --- |
|  |

Oracle provides products and services that address enterprise information technology (IT) environments. Co.'s products and services include applications and infrastructure offerings. Co.'s cloud and license business engages in the sale, marketing and delivery of its applications and infrastructure technologies through cloud and on-premise deployment models including its cloud services and license support offerings; and its cloud license and on-premise license offerings. Co.'s hardware business provides Oracle Engineered Systems, servers, storage, industry-specific hardware, operating systems, virtualization, management and other hardware-related software to support diverse IT environments.

Oracle Corporation is a global provider of enterprise cloud computing and is empowering businesses of all sizes on their journey of digital transformation. Oracle Cloud provides leading-edge capabilities in Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Data as a Service (DaaS). Oracle's security solutions enable organizations to implement and manage consistent security policies across the hybrid data center. Oracle security cloud services make leading security technologies available everywhere to organizations large and small. Oracle Cloud Platform enables developers, IT professionals, and business leaders to develop, extend, connect, and secure cloud applications, share data, and gain insights across applications and devices. Oracle's application suites, platforms, and infrastructure leverage both the latest technologies and emerging ones, including artificial intelligence (AI), machine learning, blockchain, and Internet of Things (IoT).

TCS

|  |
| --- |
|  |

Tata Consultancy Services Limited (TCS) is a subsidiary of the Tata Group, an Indian information technology consulting and business solutions company which operates in 46 countries worldwide. TCS Limited was founded in 1968 by a division of Tata Sons Limited.  Its early contracts included punched card services to TISCO (now Tata Steel), working on an Inter-Branch Reconciliation System for the Central Bank of India. In 1975 TCS made an electronic depository and trading system called SEMCOM for Swiss company. TCS also established India's first software research and development center called Tata Research Development and Design Centre in Pune, Maharashtra. On 25 August 2004, TCS became a Publicly Listed Company.

TCS is a global leader in technology and consulting services. It enables clients in 46 countries to create and execute strategies for their digital transformation.

A part of the Tata group, TCS has 3,95,000 associates(including subsidiaries) representing  131 nationalities, spanning across 46 countries as of March 31, 2018. The company generated consolidated revenues of  US $19.09 billion( a growth of 8.6% over the previous year)  for the year ended on March 31, 2018, and is listed on the National Stock Exchange and Bombay Stock Exchange in India. The company was ranked as the 57th leading brand in the US by Brand Finance. During the year the company Announces Premier Partnership with Adobe for Digital Marketing Solutions and Services. The company also announces Global Strategic Partnership with FICO during the year under review.In 2017 TCS China was set up as a joint venture with the Chinese government and other partners.