**FOREX PREDICTION USING MACHINE LEARNING TECHNIQUES.**

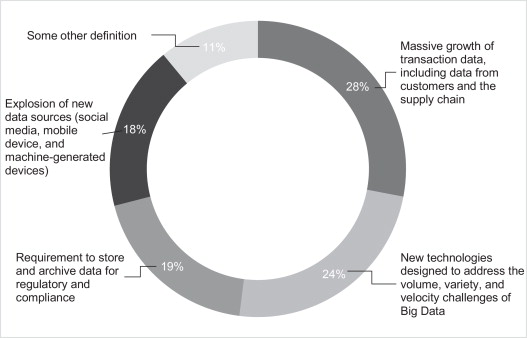
**INTRODUCTION**

Stock prediction has recently grown to be a huge research area in the field of predictive analytics, big data analytics and statistical analysis. The field of stock prediction has used machine learning techniques as well as recently predictive analytics to predict stock prices. . Stock prediction is the science of determining the future value of a stock. The basic methodologies of stock prediction are classified into fundamental analysis and technical analysis. Fundamental analysis does not take into account the previous time series data and rather takes into account the intrinsic worth of the stock such as the earning potential, the portfolio of the company and it is in other words like a survey an analyst would do. Technical analysis on the other hand uses previous data of stock, which can be taken on a daily, weekly or monthly basis, and this time series data is analyzed to predict the future price of the stock. The random walk hypothesis on the other hand states that past data cannot be used to forecast the future data because it simply is independent of each other although it has been successfully predicted in the recent times, which our paper explores.

Analysis and Prediction of forex has gained immense value in today’s economy. The stock price prediction is a difficult process owing to the irregularities in stock prices. Every trader wants to know if the pattern has been repeated in past to know what the possible output of the current situation will be. The primary objective is to propose a methodology that will use a historical dataset and provide a more accurate prediction on stock price. In this paper, we will be using machine learning pattern recognition algorithm on forex tick dataset. The learned model then will produce pattern from the given dataset and on the pattern of increasing or decreasing, the buyer will initiate a buy or sell the stock respectively. We will use python coding to execute the algorithm in jupyter notebook. Matplot library will help us to perform graphing in the process and Numpy will be helpful in doing statistical analysis of data.

**Defining Big Data:**

Big Data has gained much attention from the academia. In the digital and computing world, information is generated and collected at a rate that rapidly exceeds the boundary range. Currently, over 2 billion people worldwide are connected to the Internet, and over 5 billion individuals own mobile phones. By 2020, 50 billion devices are expected to be connected to the Internet. At this point, predicted data production will be 44 times greater. As information is transferred and shared at light speed on optic fiber and wireless networks, the volume of data and the speed of market growth increase. However, the fast growth rate of such large data generates numerous challenges, such as the rapid growth of data, transfer speed, diverse data, and security. Nonetheless, Big Data is still in its infancy stage, and the domain has not been reviewed in general. Future research directions in this field are determined based on opportunities and several open issues in Big Data domination. These research directions facilitate the exploration of the domain and the development of optimal techniques to address Big Data. Big data definitions have evolved rapidly, which has raised some confusion. The below figure shows how executives differed in their understanding of big data, where some definitions focused on what it is, while others tried to answer what it does.



Clearly, the first question comes to mind is “what are the characteristics of big data?” However, other characteristics of big data have emerged recently. For instance, [Laney](https://www.sciencedirect.com/science/article/pii/S0268401214001066" \l "bib0115) suggested that Volume, Variety, and Velocity (or the Three V's) are the three dimensions of challenges in data management. The Three V's have emerged as a common framework to describe big data. We describe the Three V's below:

**Volume** refers to the magnitude of data. Big data sizes are reported in multiple terabytes and petabytes. A survey conducted by IBM in mid-2012 revealed that just over half of the 1144 respondents considered datasets over one terabyte to be big data. One terabyte store as much data as would fit on 1500 CDs or 220 DVDs, enough to store around 16 million Facebook photographs. It is reported that Facebook processes up to one million photographs per second. One petabyte equals 1024 terabytes. Earlier estimates suggest that Facebook stored 260 billion photos using storage space of over 20 petabytes. Definitions of big data volumes are relative and vary by factors, such as time and the type of data. What may be deemed big data today may not meet the threshold in the future because storage capacities will increase, allowing even bigger data sets to be captured. In addition, the type of data, discussed under variety, defines what is meant by ‘big’. Two datasets of the same size may require different data management technologies based on their type, e.g., tabular versus video data. Thus, definitions of big data also depend upon the industry. These considerations therefore make it impractical to define a specific threshold for big data volumes.

**Variety** refers to the structural heterogeneity in a dataset. Technological advances allow firms to use various types of structured, semi-structured, and unstructured data. Structured data, which constitutes only 5% of all existing data, refers to the tabular data found in spreadsheets or relational databases. Text, images, audio, and video are examples of unstructured data, which sometimes lack the structural organization required by machines for analysis. Spanning a continuum between fully structured and unstructured data, the format of semi-structured data does not conform to strict standards. Extensible Mark-up Language (XML), a textual language for exchanging data on the Web, is a typical example of semi-structured data. XML documents contain user-defined data tags which make them machine-readable.

A high level of variety, a defining characteristic of big data, is not necessarily new. Organizations have been hoarding unstructured data from internal sources (e.g., sensor data) and external sources (e.g., social media). However, the emergence of new data management technologies and analytics, which enable organizations to leverage data in their business processes, is the innovative aspect. For instance, facial recognition technologies empower the brick-and-mortar retailers to acquire intelligence about store traffic, the age or gender composition of their customers, and their in-store movement patterns. This invaluable information is leveraged in decisions related to product promotions, placement, and staffing. Clickstream data provides a wealth of information about customer behaviour and browsing patterns to online retailers. Clickstream advises on the timing and sequence of pages viewed by a customer. Using big data analytics, even small and medium-sized enterprises (SMEs) can mine massive volumes of semi-structured data to improve website designs and implement effective cross-selling and personalized product recommendation systems.

**Velocity** refers to the rate at which data are generated and the speed at which it should be analysed and acted upon. The proliferation of digital devices such as smartphones and sensors has led to an unprecedented rate of data creation and is driving a growing need for real-time analytics and evidence-based planning. Even conventional retailers are generating high-frequency data. Wal-Mart, for instance, processes more than one million transactions per hour. The data emanating from mobile devices and flowing through mobile apps produces torrents of information that can be used to generate real-time, personalized offers for everyday customers. This data provides sound information about customers, such as geospatial location, demographics, and past buying patterns, which can be analysed in real time to create real customer value.

Given the soaring popularity of smartphones, retailers will soon have to deal with hundreds of thousands of streaming data sources that demand real-time analytics. Traditional data management systems are not capable of handling huge data feeds instantaneously. This is where big data technologies come into play. They enable firms to create real-time intelligence from high volumes of ‘perishable’ data.

In addition to the three V's, other dimensions of big data have also been mentioned. These include:

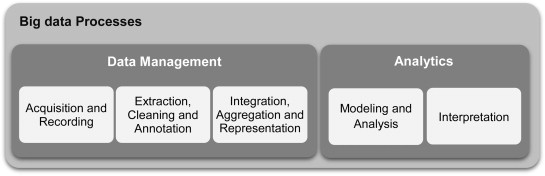
* Veracity - IBM coined Veracity as the fourth V, which represents the unreliability inherent in some sources of data. For example, customer sentiments in social media are uncertain in nature, since they entail human judgment. Yet they contain valuable information. Thus, the need to deal with imprecise and uncertain data is another facet of big data, which is addressed using tools and analytics developed for management and mining of uncertain data.
* Variability (and complexity) - SAS introduced Variability and Complexity as two additional dimensions of big data. Variability refers to the variation in the data flow rates. Often, big data velocity is not consistent and has periodic peaks and troughs. Complexity refers to the fact that big data are generated through a myriad of sources. This imposes a critical challenge: the need to connect, match, cleanse and transform data received from different sources.
* Value - Oracle introduced Value as a defining attribute of big data. Based on Oracle's definition, big data are often characterized by relatively “low value density”. That is, the data received in the original form usually has a low value relative to its volume. However, a high value can be obtained by analysing large volumes of such data.

The relativity of big data volumes discussed earlier applies to all dimensions. Thus, universal benchmarks do not exist for volume, variety, and velocity that define big data. The defining limits depend upon the size, sector, and location of the firm and these limits evolve over time. Also important is the fact that these dimensions are not independent of each other. As one-dimension changes, the likelihood increases that another dimension will also change as a result. However, a ‘three-V tipping point’ exists for every firm beyond which traditional data management and analysis technologies become inadequate for deriving timely intelligence. The Three-V tipping point is the threshold beyond which firms start dealing with big data. The firms should then trade-off the future value expected from big data technologies against their implementation costs.

**Big Data Analytics:**

Big data analytics is the method of analyzing large amounts of data to find patterns or correlations within the data. The kinds of data that are analyzed are not necessarily structured in their formats. The unstructured data is in the form of text and not in the formats of rows and columns, which makes it challenging to understand the text and mine the particular data required within the text. Hence, big data analytics requires the processes of data mining which could be from various sources such as historical data or real time social media data. Hence forecasting, optimization, text analytics and predictive analytics play a major role in big data analytics. Data mining is the method to compute patterns and correlations within a large amount of data. This may involve methods such as machine learning, artificial intelligence and statistical analysis. Predictive analysis is the process of collection of huge amounts of data to find the underlying trends and patterns in the data to make scientific decisions in the future. There have been recent advancements in stock prediction by predictive analytics of social media and at the same time it has been shown that machine learning techniques can be successfully used to predict stock prices. Stock prediction is a key research interest in recent times and there have been many improvements in the prediction methodologies of stock.

Big data potential value is unlocked only when leveraged to drive decision making. To enable such evidence-based decision making, organizations need efficient processes to turn high volumes of fast-moving and diverse data into meaningful insights. The overall process of extracting insights from big data can be broken down into five stages, in the below given figure. These five stages form the two main sub-processes: data management and analytics. Data management involves processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis. Analytics, on the other hand, refers to techniques used to analyse and acquire intelligence from big data. Thus, big data analytics can be viewed as a sub-process in the overall process of ‘insight extraction’ from big data.



Some of the big data analytical technique has been explained in detail below:

**Text analytics** (text mining) refers to techniques that extract information from textual data. Social network feeds, emails, blogs, online forums, survey responses, corporate documents, news, and call center logs are examples of textual data held by organizations. Text analytics involve statistical analysis, computational linguistics, and machine learning. Text analytics enable businesses to convert large volumes of human generated text into meaningful summaries, which support evidence-based decision-making.

**Information extraction** (IE) techniques extract structured data from unstructured text. For example, IE algorithms can extract structured information such as drug name, dosage, and frequency from medical prescriptions. Two sub-tasks in IE are Entity Recognition (ER) and Relation Extraction (RE). ER finds names in text and classifies them into predefined categories such as person, date, location, and organization. RE finds and extracts semantic relationships between entities (e.g., persons, organizations, drugs, genes, etc.) in the text.

**Text summarization** techniques automatically produce a succinct summary of a single or multiple document. The resulting summary conveys the key information in the original text(s). Applications include scientific and news articles, advertisements, emails, and blogs.

**Predictive analytics** comprise a variety of techniques that predict future outcomes based on historical and current data. In practice, predictive analytics can be applied to almost all disciplines – from predicting the failure of jet engines based on the stream of data from several thousand sensors, to predicting customers’ next moves based on what they buy, when they buy, and even what they say on social media. At its core, predictive analytics seek to uncover patterns and capture relationships in data. Predictive analytics techniques are subdivided into two groups. Some techniques, such as moving averages, attempt to discover the historical patterns in the outcome variable(s) and extrapolate them to the future. Others, such as linear regression, aim to capture the interdependencies between outcome variable(s) and explanatory variables, and exploit them to make predictions. Based on the underlying methodology, techniques can also be categorized into two groups: regression techniques (e.g., multinomial logit models) and machine learning techniques (e.g., neural networks). Another classification is based on the type of outcome variables: techniques such as linear regression address continuous outcome variables (e.g., sale price of houses), while others such as Random Forests are applied to discrete outcome variables (e.g., credit status).

**PROBLEM ASSESSMENT**

Prediction and analysis of GBP/USD market data have got an important role in today’s economy. We will be using the network architecture to design the flow of data. We will be using pattern recognition algorithm to complete the analysis. This approach allows the user to specify mathematical operations as elements in a graph of data, variables and operators.

In recent years, the financial industry has seen an upsurge of interest in Big Data. This comes as no surprise to finance experts, who understand the potential value of data in this field and are aware that no industry can benefit more from Big Data than the financial services industry. After all, the industry not only is driven by data but also thrives on data. Today, the data, characterized by the four Vs, which refer to volume, variety, velocity, and veracity, are prevalent at various levels of this field, ranging from capital markets to the financial services industry. The prevalence of electronic trading has spurred up growth in trading activity and HFT, which, among other factors, have led to the availability of very large-scale ultrahigh-frequency data (UHFD). These high-speed data are already having a huge impact in the field in several areas ranging from risk assessment and management to business intelligence. For example, the availability of UHFD is forcing the market participants to rethink the traditional ways of risk assessment and bringing up attention to more accurate, short-term risk assessment measures. Similar trends can be observed in the financial services sector, where Big Data is increasingly becoming the most significant, promising, and differentiating asset for the financial services companies.

In recent years, it is becoming increasingly important for trading firms to adopt a data centric perspective to handle the mounting regulatory pressures and succeed in today’s digital, global marketplace. In the past, trading organizations collected large amounts of data. However, these institutions depended primarily on the conventional ETL framework and lacked the ability to process the data and produce actionable knowledge out of it within realistic time frames. This approach prevented them from gaining a full perspective of their business insights and made it difficult for them to anticipate and respond to changing market conditions, business needs, and emerging opportunities, a few must-haves essential to thrive in today’s dynamic business environment. As a result, the firms relying on the traditional schemes have started to address the limitations inherent in their conventional systems. Today, a growing number of financial institutions are exploring new ways of unlocking the potential of available data to gain insights that can help them improve their performance and gain competitive advantage through factual and actionable knowledge, timely decisions, risk management and mitigation, and efficient operations in highly complex and often volatile business environments.

Every trader wants to find out the pattern of the forex before he makes any decision of making small or big investment in that forex. These changes in forex market reflect directly to the economy of the area. There is abundance of algorithms that could be found on the Internet that allows the user to predict the next change in forex market. Most of these are just a hoax and a way to manipulate people.

We are going to find out patterns by plotting together the lines of those patterns on graph, which are very much like one another. Then we will perform back test on these results. A forward test can be performed on the upcoming data that has been produced after the prediction of the data, but that data cannot be back tested to give the guarantee that it is indeed a suitable prediction.

**OBJECTIVE**

Analysis and Prediction of forex has gained immense value in today’s economy. Every trader wants to know if the pattern has been repeated in past to know what the possible output of the current situation will be. We will attempt to find out some patterns in the data and analyze if the pattern has been seen somewhere before and perform prediction technique based on the kind of pattern we found. We are going to use basic machine learning principles to locate actionable trading rules within stocks or forex. After this, we will back test the algorithm we got. We will use python to perform these steps. Matplot library will help us to perform graphing in the process and Numpy will be helpful in doing statistical analysis of data.

We are expecting that the prices of stocks or various forex ratios are a direct reflection of the psychology of the people trading, i.e.: the traders (either people or computers) are making decisions, based on a bunch of variables. The theory is that, when those same variables present themselves again, we get a repeat of actions that create a similar "pattern," then the outcome as well is likely to be similar, because the variables are almost the same.

So, what we're going to do here is to first create a machine-learned batch of what will end up being millions of patterns with their results, which can be used to predict future outcomes second is to test this.

Our entire system is really built on the inference of pattern recognition, so if patterns change due to new data, that's really built in and is done by programming that was done before results.

This allows Back Testing to actually serve a very truthful and accurate purpose. If a machine-learned live algorithm passes the back test, it is highly likely to continue performing well in the future, not because it passed a back test, but because our hypothesis and entire model passed the back test. Unlike finding the best algorithm at the time and Back Testing for great results. With that, what we will do is take a range of data in succession and create a pattern with it. How we're going to do this is going to be with percent change. We want to have the data normalized as best we can, so it can be used no matter what the price was. We're just going to use a succession of percent change for it.

The longer the pattern, the more likely the end is to be less similar, but the actual direction of the pattern will be more similar. This can be useful, since some patterns might take more time to react than others, and we want the build-up to be most accurate, but we might actually prefer the end to be more accurate in the future, so we could do reverse percent change. We can also do a point-to-point percent change as well.

**DATA PREPARATION**

Data analysis enables an organization to handle abundant information that can affect the business. However, data analysis is challenging for various applications because of the complexity of the data that must be analyzed and the scalability of the underlying algorithms that support such processes. Data analysis has two main objectives: to understand the relationships among features and to develop effective methods of data mining that can accurately predict future observations. Various devices currently generate increasing amounts of data. Accordingly, the speed of the access and mining of both structured and unstructured data has increased over time. Thus, techniques that can analyze such large amounts of data are necessary. Available analytical techniques include data mining, visualization, statistical analysis, and machine learning.

First step is to prepare training and test data. The dataset is split into training and test data. The training data contains 80% of the total dataset. The data was sequentially sliced. Scaling the inputs can be very advantageous for neural network architectures because most common activation functions of the network’s neurons such as tanh or sigmod are defined on the [-1,1] or[0,1] interval respectively. We will scale both the inputs and targets anyway. Scaling can be easily accomplished in Python using Skelearn’s Min Max Scaler.

Data scientists are facing many challenges when dealing with Big Data. One challenge is how to collect, integrate and store, with less hardware and software requirements, tremendous data sets generated from distributed sources. Another challenge is Big Data management. It is crucial to efficiently manage Big Data in order to facilitate the extraction of reliable insight and to optimize expenses. Indeed, a good data management is the foundation for Big Data analytics. Big Data management means to clean data for reliability, to aggregate data coming from different sources and to encode data for security and privacy. It means also to ensure efficient Big Data storage and a role-based access to multiple distributed end-points. In other words, Big Data management goal is to ensure reliable data that is easily accessible, manageable, properly stored and secured.

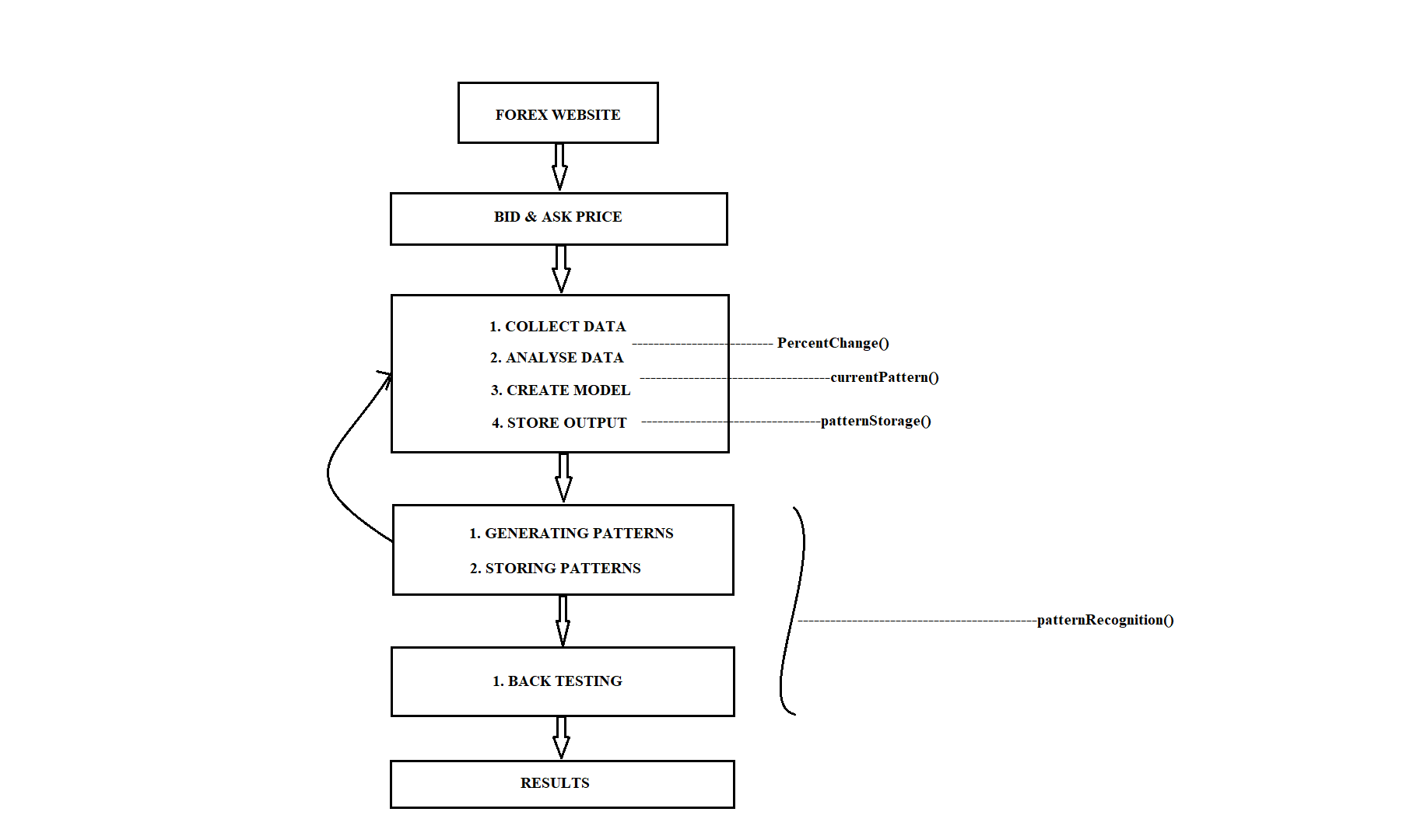
Those five steps (Cleaning, Aggregation, Encoding, Storage and Access) are not new and are known in the case of traditional data management. The challenge in Big Data is how to manage the complexity of Big Data nature (velocity, volume and variety) ([Khan et al., 2014](https://www.sciencedirect.com/science/article/pii/S1319157817300034" \l "b0220)) and process it in a distributed environment with a mix of applications. In fact, for reliable analysis results, it is essential to verify the reliability of sources and data quality before engaging resources. However, data sources may contain noises, errors or incomplete data. The challenge is how to clean such huge data sets and how to decide about which data is reliable, which data is useful.

Another challenge is to synchronize outside data sources and distributed Big Data platforms (including applications, repositories, sensors, networks, etc.) with the internal infrastructures of an organization. Most of the time, it is not sufficient to analyze the data generated inside organizations. In order to extract valuable insight and knowledge, it is important to go a step further and to aggregate internal data with external data sources. External data could include third-party sources, information about market fluctuation, weather forecasting and traffic conditions, data from social networks, customers comments and citizen feedbacks. This can help, for instance, to maximize the strength of predictive models used for analytics.

**ARCHITECTURE**

Firstly, the trading system collects price data from the system. Some algorithm trading systems may also collect data from the web for analysis. While the data is being collected, the system performs some complicated analysis on the data to look for profitable chances with the expectation of making profit. Sometimes the trading system conducts a simulation to see what the actions may result in. Finally, the system decides on the buy/sell/hold actions, the quantity of order, and the time to trade, it then generates some trading signals. The signals can be directly transmitted to the exchanges using a predefined data format, and trading orders are executed immediately through an API exposed by the exchange without any human intervention. Some investors may like to take a look at what signals the algorithm trading system have generated, and he can initiate the trading action manually or simply ignore the signals.

The stock price prediction is a difficult process owing to the irregularities in stock prices. Every trader wants to know if the pattern has been repeated in past to know what the possible output of the current situation will be. The primary objective is to propose a methodology that will use a historical dataset and provide a more accurate prediction on stock price. In this paper, we will be using machine learning pattern recognition algorithm on forex tick dataset. The learned model then will produce pattern from the given dataset and on the pattern of increasing or decreasing, the buyer will initiate a buy or sell the stock respectively.



**METHODOLOGY**

In this paper we are using pattern recognition algorithm on forex tick dataset. This algorithm is used when a group of price in a time frame is taken and then converting them to PercentChange in an effort to normalize the data. This pattern is then map into the memory, and then move forward one price point and remap the pattern, for each pattern mapping into the memory a log is maintained to where the price is that point. The percent similarity of the current pattern is compared with the entire previous patterns, and if the result has exceeded the certain threshold level, then the percent similarity gets considered. An estimated average outcome is generated from the aggregation of all their previous outcomes and if the result of the estimated average outcome is favorable then a customer might initiate a buy or else sells the shares.

One goal of this paper is to examine the role of and possibilities for big data in trading and show that it is improved data quality ('better' data) rather than merely a rise in data volumes that drives improved outcomes. Much of the increase in data quality comes from a mix of new data sources, a smart application of statistical tools and domain knowledge combined with theoretical insights. These serve also to effect better data compression, transformations and processing prior to analysis. Another goal is to examine the advent of predictive analytics in a trading context.

In this paper we have also used back testing in order to verify whether the pattern generated is accurate or not. Back testing is a term used in modelling to refer to testing a [predictive model](https://en.wikipedia.org/wiki/Predictive_modelling) on historical data. Back testing is a type of [retrodiction](https://en.wikipedia.org/wiki/Retrodiction), and a special type of [cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)) applied to previous time period. In a trading strategy, investment strategy, or risk modelling, back testing seeks to estimate the performance of a strategy or model if it had been employed during a past period. This requires simulating past conditions with sufficient detail, making one limitation of back testing the need for detailed historical data. A second limitation is the inability to model strategies that would affect historic prices. Finally, back testing, like other modelling, is limited by potential [overfitting](https://en.wikipedia.org/wiki/Overfitting). That is, it is often possible to find a strategy that would have worked well in the past but will not work well in the future. Despite these limitations, back testing provides information not available when models and strategies are tested on synthetic data.

Back testing has historically only been performed by large institutions and professional money managers due to the expense of obtaining and using detailed datasets. However, back trading is increasingly used on a wider basis, and independent web-based back testing platforms have emerged. Although the technique is widely used, it is prone to weaknesses. [Based financial regulations](https://en.wikipedia.org/wiki/Basel_Accords) require large financial institutions to back test certain risk models. For back testing to provide meaningful results, traders must develop their strategies and test them in good faith, avoiding bias as much as possible. That means the strategy should be developed without relying on the data used in back testing. That’s harder than it seems. Traders generally build strategies based on historical data. They must be strict about testing with different data sets from those they train their models on. Otherwise, the back test will produce glowing results that mean nothing.

Similarly, traders must also avoid data dredging, in which they test a wide range of hypothetical strategies against the same set of data with will also produce successes that fail in [real-time](https://www.investopedia.com/terms/r/real_time.asp) markets, because there are many invalid strategies that would beat the market over a specific time period by chance.

One way to compensate for the tendency to data dredge or cherry pick is to use a strategy that succeeds in the relevant, or in-sample, time period and back test it with data from a different, or out-of-sample, time period. If in-sample and out-of-sample back tests yield similar results, then they are likely generally valid.

In the function graphRawFX(), the dataset which we are currently using is plotted in the graph format.

In the function percentchange(), a machine learned batch is created which has millions of patterns and their results. A range of at is taken into consideration and a pattern is created in which the percent change will be performed. In this function percentchange(), at the starting forward percent change is performed. It means as long as the pattern goes, there are chances that actual end may be similar, but the actual end of the pattern may also be similar.

In the function currentpattern(), we are finding the current pattern from the tick dataset. The short term outcomes are also collected from this function, and later in the algorithm with the help of current pattern and stored pattern, a comparison will be done in order to find the similarity in the pattern.

In the function storingpattern(), the percent change values of patterns are being collectedfrom the tick data. To compare the patterns, a percent calculation to calculate the similaritybetween each percent change is done. The pattern which have exceeded the certain thresholdlevel is considered to be similar are getting stored, and average is calculated of the similarpatterns and those values are also getting stored.

In the function patternrecognition(), each pattern in pattern array that is there we have, wewill compare the similarity that with current pattern. The way this is done is to find thepercent change, we will find the percent change for each pattern that was found instorepattern() and will compare it with the current pattern that was found in currentpattern().The percent change in order to find the similarity. After finding the similarity of all the patterns, an average similarity is calculated. The average similarity is then compared with certain threshold level, so as the average whose value is greater; a graph is plotted for those patterns.

The patterns are then plotted along with their possible outcomes in the same graph, then a visually prediction data can be generated which with we can see where the predicted pattern is going. The outcome patterns are color-coded with which it is easy to identify in the graph. The green color is used in the pattern outcome when it is above the threshold level value and red color signifies that the generated pattern is below the level. The pattern in cyan colored line is the current pattern and all the other pattern line, which is color-coded, are the similar patterns from the past.

**CODE**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer

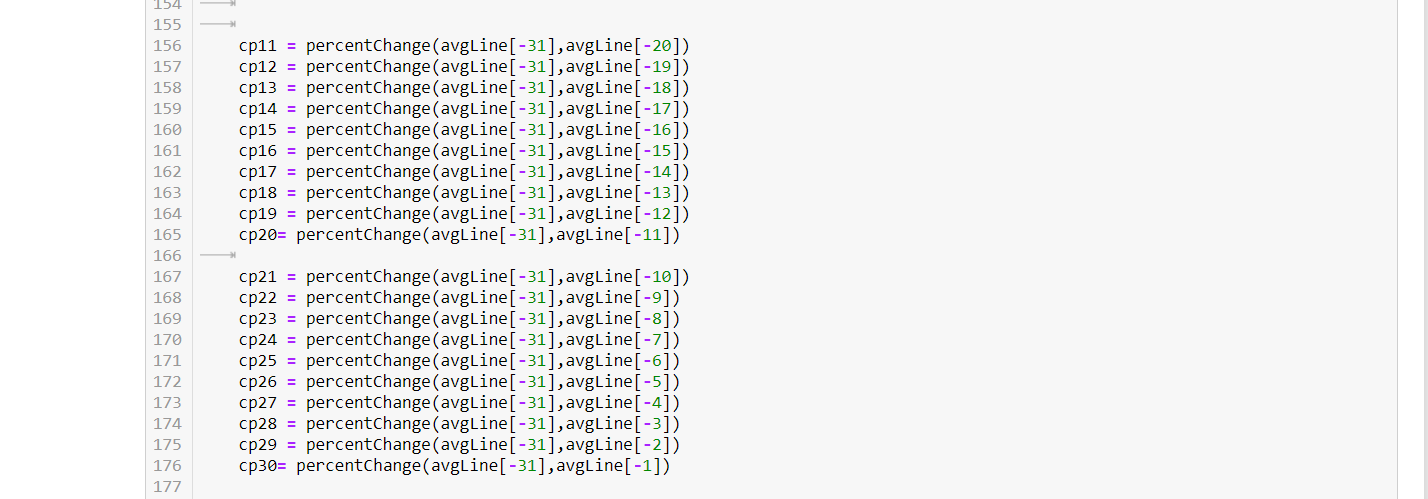
Description automatically generated**

**A close up of a logo

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

****

**A close up of a logo

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**OUTPUT**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**Back Testing Output**

**The Prediction obtained from Back testing the previous 5 actionable Trades are as**

**follows:**

data length is 62012

54939

54939

Pattern storing took: 2.4374775886535645

[-1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0]

-1.0

**drop predicted**

**0.00803561384052255**

0.006299374562092645

[100]

Entire processing took: 7.309452772140503 seconds

**Back tested Accuracy is 100.0% after 1 actionable trades**

54940

54940

Pattern storing took: 1.8911151885986328

[-1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0]

-1.0

**drop predicted**

**0.008357011532673568**

0.005785121310760738

[100, 100]

Entire processing took: 11.091313362121582 seconds

**Back tested Accuracy is 100.0% after 2 actionable trades**

54941

54941

Pattern storing took: 2.0784897804260254

[-1.0, -1.0, -1.0, -1.0, -1.0]

-1.0

**drop predicted**

**0.00835698467135923**

0.005206592445941284

[100, 100, 100]

Entire processing took: 15.073985815048218 seconds

**Back tested Accuracy is 100.0% after 3 actionable trades**

54942

54942

Pattern storing took: 1.9537765979766846

[-1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0]

-1.0

**drop predicted**

**0.008678407158707121**

0.003471005852501438

[100, 100, 100, 100]

Entire processing took: 18.857866764068604 seconds

**Back tested Accuracy is 100.0% after 4 actionable trades**

54943

54943

Pattern storing took: 1.9308865070343018

[-1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0, -1.0]

-1.0

**drop predicted**

**0.01060697617608678**

-0.0008034400087395548

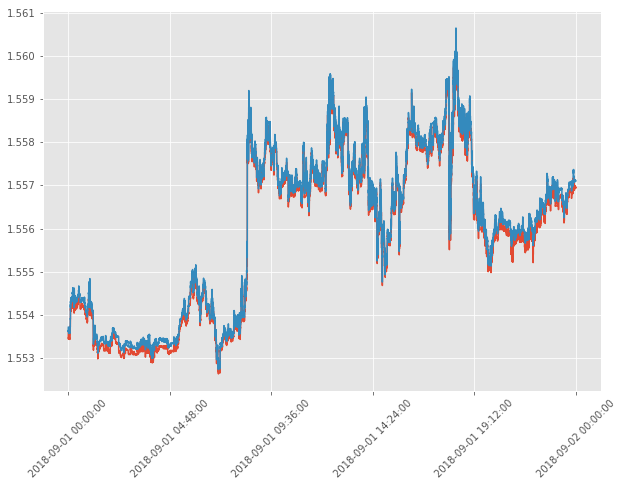
[100, 100, 100, 100, 100]

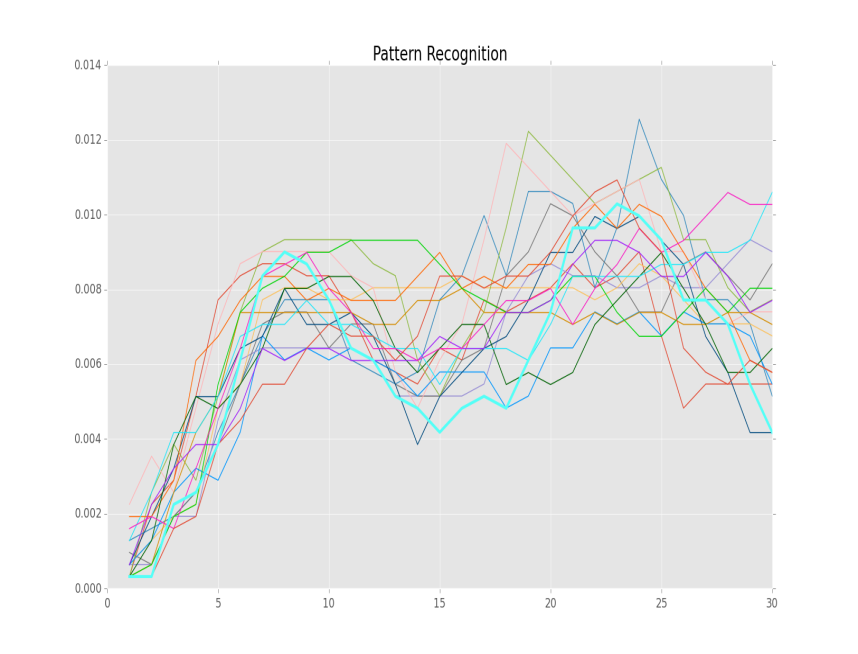
Entire processing took: 23.447073221206665 seconds

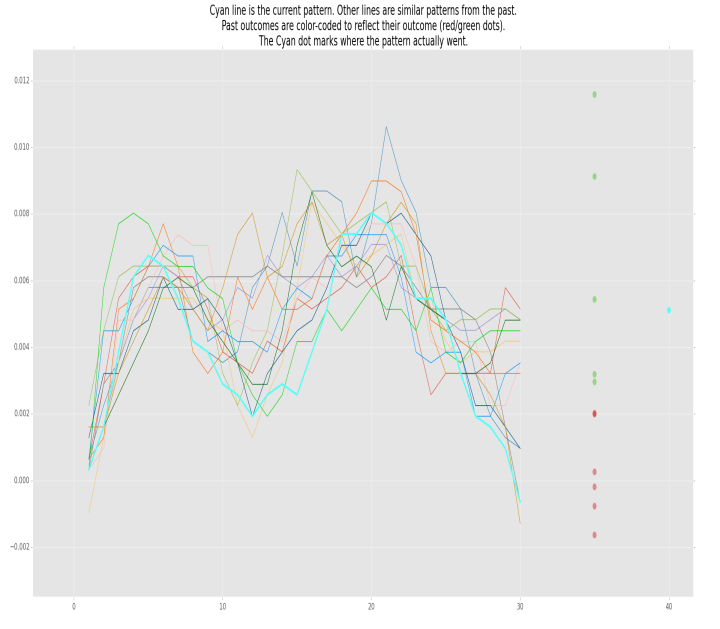
**Back tested Accuracy is 100.0% after 5 actionable trades**

**GRAPHICAL RESULTS**

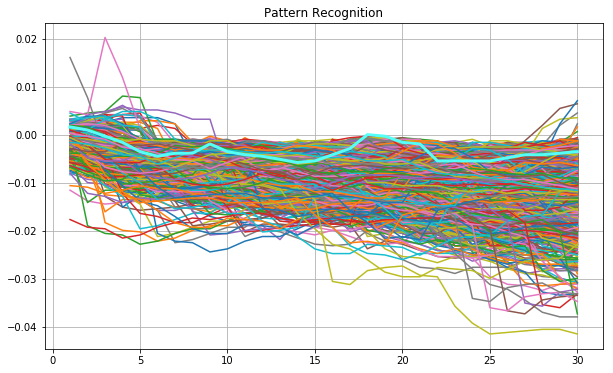
The dataset which we have used for prediction is represented in graphical format.

****



.

The average similarity is then compared with certain threshold level, so as the average whose value is greater; a graph is plotted for those patterns. The outcome patterns are color-coded with which it is easy to identify in the graph. The green color is used in the pattern outcome when it is above the threshold level value and red color signifies that the generated pattern is below the level. The pattern in cyan colored line is the current pattern.



**CONCLUSION**

In this analysis we studied how the patterns are recognized and analysis is done on them to make the future predictions about the forex. When function patternRecogntion() is completely executed we can see the visual representation of our predictions and the outcomes to those predictions. Thin lines are the comparable patterns that we found. Cyan line is the current pattern. Other lines are similar patterns from the past. Predicted outcome are the color-coded to reflect a positive or negative prediction. The Cyan dot marks where the pattern went. Only data in the past is used to generate patterns and predictions. Red and green dots represent each patterns outcome. Rise is the green dot, and if it’s a fall then a red dot. It’s not the degree at which it would rise or fall. Then the average rise or fall is calculated and plotted against the graph with dark blue dot. Then the Cyan plot shows the actual outcome of the plot. So, if the prediction is falling then consider selling, or else if it continues to rise then consider holding on to it.