Turkish Journal of Computer and Mathematics Education *Vol.12 No.6 (2021), 3132-3139*
Research Article

A Directionality & Novel Function for Stock Price Modelling Prediction Dynamics based on Multiple Learning

Sarthak Gupta [1*], Aditi Poddar[2], Rahul. Rajesh[3] and Caroline El Fiorenza.J[4]

1 Student, Depart of Computer Science, SRM Institute of Science and Technology, Ramapuram
2 Student, Depart of Computer Science, SRM Institute of Science and Technology, Ramapuram
3 Student, Depart of Computer Science, SRM Institute of Science and Technology, Ramapuram
4 Professor, Depart of Computer Science, SRM Institute of Science and Technology, Ramapuram

Abstract: Several methods have been used to forecast stock market movements. However, owing to the complexities of the stock market, the findings are not entirely acceptable. Normally, to deduce a time-scale problem standardized method are taken into consideration. Other techniques followed are mainly lacking behind in terms of distinctive time attributes in the stock dataset as well as the simple and rational concept of the drift. In the study of this paper, we introduce a method to improve prediction accuracy by using the features of the time of the given stock dataset. Firstly, we never treat the data unsystematically, we use the weighted time function accurately to allocate loads to data based on their random proximity to the dataset that is expected. Secondly, concepts of stock patterns were conventionally introduced by citing the monetary thesis and speculations. Time series forecasting can be difficult since there are several methods to choose from, each with its own set of hyperparameters. The Prophet library is a free and open-source library for forecasting univariate time-series datasets. It is simple to use and is intended to automatically find a good collection of hyperparameters for the model in order to make accurate forecasts for data that has patterns and seasonal structure by default. We will use the stock prices of Santander Group, a multinational Spanish financial firm headquartered in Madrid in Spain whose founder is Banco Santander. Furthermore, as the world's 16th largest banking institution, Banco maintains a name in all major money-related centres.

Keywords: Prophet Library, Univariate time series, Weighted time function

1. Introduction

During the stock market exploration, the first thing one should know is that the movement of prices always follows a particular trend. As the evaluation continues. we know that the market moves non-linearly, we may expect zigzag lines in the graph. Nonetheless, the concept of trend is widely accepted, the exposition differs. We can see that the ebbs and flows always come in pairs. The market tends to move in the opposite direction before it follows its original direction, rise or fall. We propose a model that focuses on utilizing the essential nature of time in time-scale datasets of stock. Primarily, three aspects are covered during the exploration. Firstly, one should understand that all the data should be treated distinctively as they every now and then. The following subsection shows some time-weighted functions and to attain this we allocate dissimilar weights to datasets at a distinct time. For business flourishment, understanding of time-based patterns based on metric points dataset is a must which is possible using the time series analysis. This time-based analysis helps one answer many business-related questions. For instance, like how much production material would be required, how much crowd one can anticipate in their online stores to will there be an increase in sales for the following month or not- these vital issues. can be easily solved. The main purpose of this analysis is to help us find the solution to the problems arising and to forecast things associated with time.

Prophet adjusts time functions – linear and nonlinear as elements considering time as an independent variable. Instead of looking peculiarly at the observations which are timely dependent on each other, Prophet associates the prediction of problems as a curve adjustment practice.

2. Literature Study

[1] The deep neural network architecture is used in this paper for analysis, as well as for applying the set of instructions to acquire important conjecturing notions and for exploring hierarchies over the evoked factors. It also assesses the merged visualization framework by going through different studies in stock price prediction using social media tweets of companies and financial news that revolve online. This method employs the LRP algorithm, which produces a list of terms with a nonzero relevance score on each day, which is referred to as keywords in this context. These keywords, together with related bigrams and news/tweet names, comprise the

possible driving factors for stock price shift. This method supports the Frequency Trading Patterns Modelling Approach and discovering the dynamic evolutionary but fails to analyse multivariate nonlinear time series and does not take dependencies between time segments into account, resulting in unreliable prediction model results and non-stationary time is not included in this system.

[2] The proposed method in this paper determines the correlation matrix of all macroeconomic variables considered. Validation methods such as the Kaiser-Meyer-Olkin and Bartlett tests are used to demonstrate the need for dimensionality reduction and the presence of multicollinearity. The Principal Component Analysis method is used to minimize dimensionality to seven variables, and then PCA with the varimax rotation method is used to locate factors with the greatest variance. The impact of these seven variables on the NSE Nifty and BSE SENSEX indices is also examined using regression. With the aid of macroeconomic indicators, an Artificial Neural Network is used to forecast stock market movement. This aids this method in forecasting market reliability supports the outcomes of an unknown event even if the findings are inadequate to draw conclusive conclusions about the selection does not produce substantial improvements in prediction power and does not support both stationary and non-stationary conditions.

[3] This paper examines expert selection in these markets in order to enhance their dependability. Instead of using prices to aggregate decisions from a specific community of experts, we describe a market deconstruction based on player portfolios. Through the evolution of their portfolio, this decision technology makes the attitudes of experts against their decisions visible. The key contribution of this paper is the discovery of two Persistent Homological Invariants that can classify experts into groups based on the backgrounds of their portfolios. Surprisingly, this translates into the concept of two dominant classes. Following the Microeconomic jargon, a simulation of the Prediction Market with artificial agents helps one to view these two groups as rational and irrational players. This assists the system in bridging a major gap in real-world implementation and in this prediction model the primary components learned are visually interpreted. and extract predictive factors that are important. in contrast, does not support a hierarchical visualization interface and is unable to extract textual factors related to regular prediction results. It has not been implemented and does not correlate algorithm-extracted textual factors.

[4] In this paper, a new approach for simplifying noisy-filled financial temporal series using sequence reconstruction and to acquire the non-linear formation of time series we use the convolutional neural network. Our volunteered method is efficient in feature learning, outperforms conventional process approaches, and deals with trading related to frequency models is evident in the final results by 4%-7%. For stock time-series pattern prediction, this paper proposed a method that combines motif-based sequence reconstruction with CNN. This assists the method in determining the correlation matrix of all macroeconomic variables considered. Also, minimize the dimensionality. Predict stock market movement using macroeconomic metrics, while It is impossible to obtain nontrivial information, and the number of factors is unknown, even when extraneous variables are not taken into account.

3. Architecture Diagram

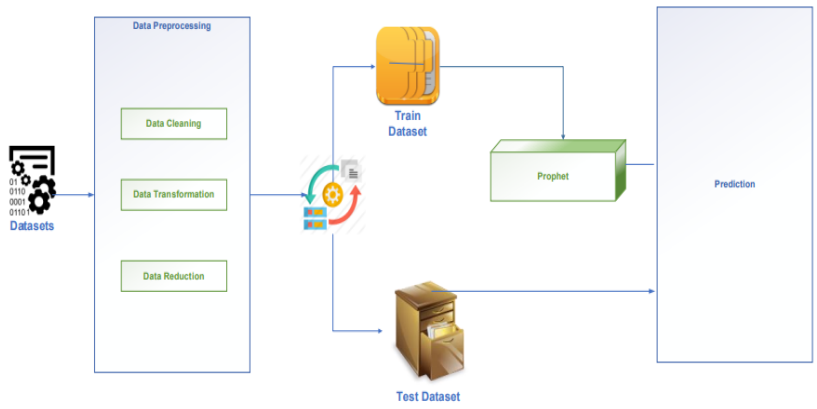

Figure 1. Architecture Diagram

4. Modules

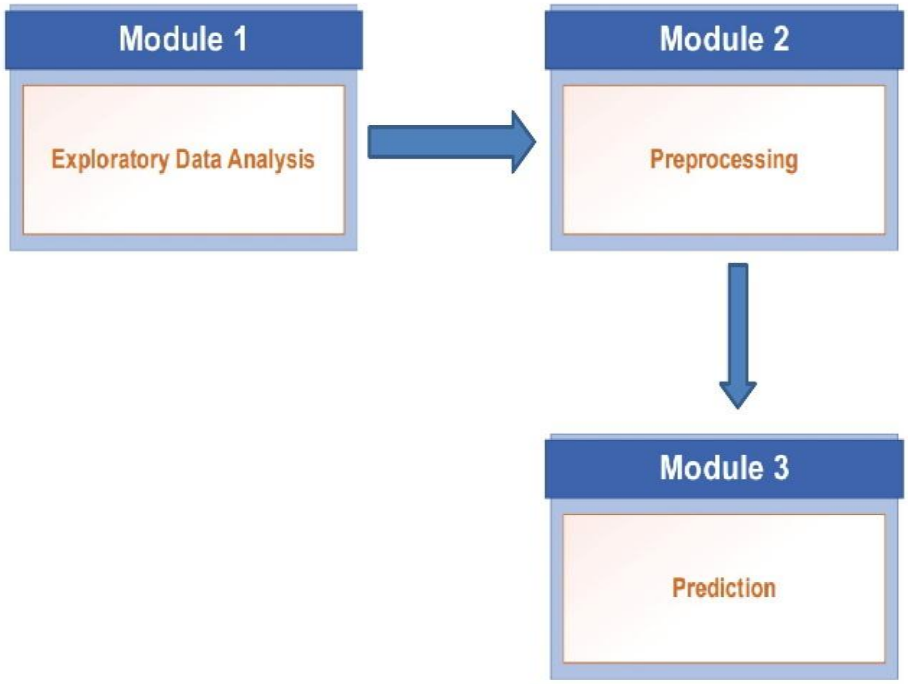

Figure 2. Modules

4.1 Exploratory Data Analysis

It is one of the most important steps in the prediction of stock prices. In this step, an initial or first investigation is done on the extracted data which leads to information such as Anomaly detection. This gives information about the normal and unusual behavior of the data; glitches and a conclusion are derived with the help of stats and graphical representation.

For Example:

";" is used to separate data in the information given. To understand the data better we make use of the pandas library functions like ". head ()" and ". tail ()". The former returns the first five observations whereas the latter returns the last five observations of the given data set.

We should also be aware of the columns and the data residing in the same and make sure that they do not hold null values. It is mandatory that the data used contain integer and decimal values. Null value columns should be avoided.

In this system, previous data sets from companies are used and analysed to identify patterns and spot anomalies, providing us with valuable data such as

- Opening Price
- Highest Price
- Lowest price
- The volume of stocks sold

4.2 Pre-Processing

The process of transformation of data or data that is encoded so that it can be easily decoded by the machine is called data pre-processing. On a whole, the attributes of the user data can be easily understood by the algorithm. A collection of different types of data is known as the dataset. The types of data could be classified as records, instances, entities, patterns, samples, or observations. Data objects are characterized by a set of features that

capture the fundamental characteristics of an object, for instance, the mass of a physical object or the timing of the occurred event, among other things. characteristics, Variables, attributes, fields, and measurements are all words for functions.

Attributes like:

Definitive: Values derived from a predefined set. For instance, seasons of the year: Spring, summer, autumn, winter are all categories as their values are all drawn from this collection. The Boolean set is another example: True or False.

Numerical: Consists of numbers or constant values. It is all about numbers and their majority of assets are numbers only. For instance, the number of times one sneezes in a day or the number of steps you take daily.

4.3 Prediction

Prophet is a method for forecasting time series-based data that uses an incremental approach to match non-linear patterns with annual, weekly, and regular seasonality, as well as holiday impacts. Prophet fits well for time series that have distinct seasonal influences and historical evidence from several seasons. Prophet is immune to lost data and trend updates, and it usually handles outliers well.

Firstly, a time series is modelled with the defined criterion by the analysts, followed by generating and assessing the forecasts. Human analysts are also taken into consideration during a performance problem or when a manual involvement arises, human tends to be helpful in prediction and can possibly help in changing the model if necessary, based on the input. If the time series is much more than two cycles long, Prophet will match weekly and yearly seasonality by design. For a sub-daily time, sequence, it will also match daily seasonality. If necessary, you can add additional seasonality (monthly, quarterly, or hourly). Prophet will generate instability intervals for pattern variables by deliberately designed to simulate future trend shifts in your time series. If you'd like to forecast future seasonal variation or festival effect, run a few hundred HMC simulations (which only takes a few minutes) and your forecasts will also include seasonal instability in predictions. Prophet forecasts using a linear model by default, but a logistic model can be used by transferring it as an attribute.

Prophet algorithm employs a model which is time decomposable and consists of three vital model elements: pattern, cyclical, and holidays.

When brought together the equation looks like:

m(t)= n(t)+o(t)+p(t)+et

n(t): represents the continuous-time undeviating or logistic growth curve for modelling aperiodic changes.

o(t): represents repeated changes (for instance: cyclical, seasonality)

p(t): represents occasional or irregular schedules (holidays)

et: depicts error term account for unwanted changes which is not accustomed by the model.

In this system, data is analysed and pre-processed before being fed into the prophet library, where predictions are made using time series data. Which provides details such as

- Predicted price
- Predicted lower price
- Predicted the highest price

5. Technology Used

Techniques used in the development of our system's backend are:

- NumPy
- Sci-learn
- TensorFlow

- Jupyter Notebook

5.1 NumPy

NumPy is a popular Python data analysis program. NumPy allows you to ease up your workflow and interact with the other Python ecosystem modules such as sci-kit-learn, which use NumPy underneath.

It is a free and publicly accessible Python-based numerical toolkit library. Often referred to as multidimensional arrays and matrices, it has data types such as arrays and matrix structures. The calculator will work for all kinds of arrays, including statistical, trigonometric, and algebraic array functions. As a result, the library includes a plethora of mathematical and translation functions. NumPy is a Numeric and Numerary extension. Random number generators are also used in NumPy. NumPy is a Python extension for a C-based library. Pandas' artifacts are highly reliant on NumPy objects. Pandas is an extension of NumPy.

5.2 Sci-Learn

It is an open-source library for Python with strong data processing and mining methods used for analysing a dataset. It is a simple and effective tool for predictive data analysis It is distributed under BSD license and is based on the machine learning libraries mentioned below: NumPy is a Python library that allows you to handle multidimensional arrays and matrices. Which also includes a large set of mathematical functions to perform different calculations. SciPy is an ecosystem of libraries for performing technical programming activities. Matplotlib is a library for creating maps and graphs. Scikit-learn has a large number of built-in algorithms which help data science initiatives succeed.To predict the price of a stock or Predicting a continuous-valued feature of an entity Regression algorithm is used which predicts successful value outputs.

5.3 TensorFlow

It is an open-source Python-friendly library for numerical computing that speeds up and simplifies machine learning. TensorFlow will train and run deep neural networks for handwriting digit detection, recurrent neural networks, word embeddings, image recognition sequence-to-sequence, natural language processing (NLP), models for machine translation, and PDE-based simulations. Most of all, TensorFlow will forecast performance at scale using the same models that were used for testing.

5.4 Jupyter Notebook

The Jupyter Notebook is an open-source and free web application that lets you build and exchange documents with live coding, calculations, visualizations, and narrative text. numerical simulation, mathematical modelling, data visualization, data cleaning and transformation, artificial learning, and several other applications are possible.To import the dataset into Jupyter Notebook, first, download the required dataset, then import it and then open the dataset in Jupyter Notebook. When you type jupyter notebook into Terminal, a Jupyter Notebook would open automatically.

6. Results and Discussion

Performed analysis on Santander Dataset to predict future stock prices. In Order to predict the Stock Price Data set of Santander Group is fed and analysis is performed using NumPy and Pandas.

	Date	Close	High	Low	Open	Volume
2	#########	1.656	1.656	1.634	1.64	5460000
3	#########	1.659	1.672	1.65	1.656	7370000
4	#########	1.665	1.687	1.659	1.678	5180000
5	#########	1.653	1.665	1.64	1.653	1830000
6	#########	1.646	1.65	1.624	1.628	5610000
7	#########	1.659	1.687	1.656	1.681	6940000
8	#########	1.681	1.681	1.634	1.659	3180000
9	#########	1.697	1.697	1.684	1.684	5670000
10	#########	1.687	1.697	1.678	1.697	4830000
11	#########	1.697	1.7	1.678	1.7	3890000
12	#########	1.675	1.691	1.665	1.684	2380000
13	#########	1.672	1.681	1.662	1.669	4240000
14	#########	1.694	1.694	1.684	1.691	4840000
15	#########	1.678	1.697	1.665	1.687	2500000
16	#########	1.681	1.684	1.672	1.672	3090000
17	#########	1.687	1.691	1.675	1.678	3310000
18	#########	1.691	1.694	1.681	1.687	3730000
19	#########	1.694	1.7	1.691	1.694	2770000
20	#########	1.691	1.694	1.681	1.691	4600000

Figure 3. Santander Dataset

Results for the Santander Dataset After performing analysis on it

Date	Open	High	Low	Close	Volume
2015-01-02	6.696	6.725	6.559	6.645	42200000
2015-01-05	6.600	6.687	6.327	6.341	58030000
2015-01-06	6.370	6.400	6.228	6.231	47480000
2015-01-07	6.285	6.404	6.208	6.270	51880000
2015-01-08	6.407	6.500	6.349	6.479	22910000

Figure 4. Analysis Result

The above data when fed into the Prophet Library yields predicted prices data.

	ds	yhat	yhat_lower	yhat_upper
1493	2020-06-02	3.015981	1.428165	4.592060
1494	2020-06-03	3.011780	1.415372	4.570897
1495	2020-06-04	3.009097	1.348551	4.625449
1496	2020-06-05	2.991608	1.391893	4.622340
1497	2020-06-06	3.355567	1.689577	4.985972

Figure 5. Prophet library Result

Output: Predicted Stock Price Graph plotted from the information derived during analysis and Prophet library Results.

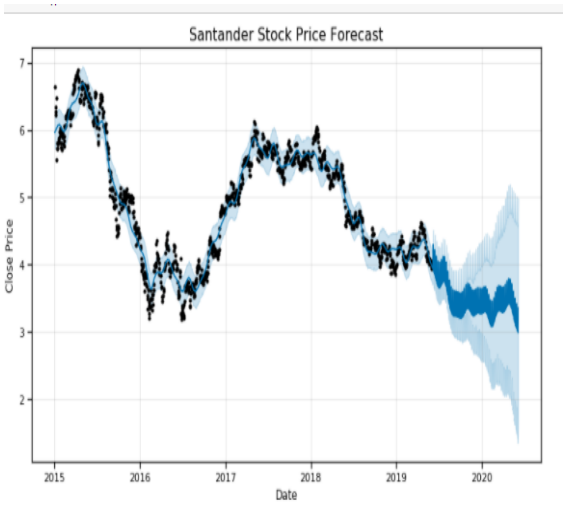

Figure 6. Predicted Price Graph

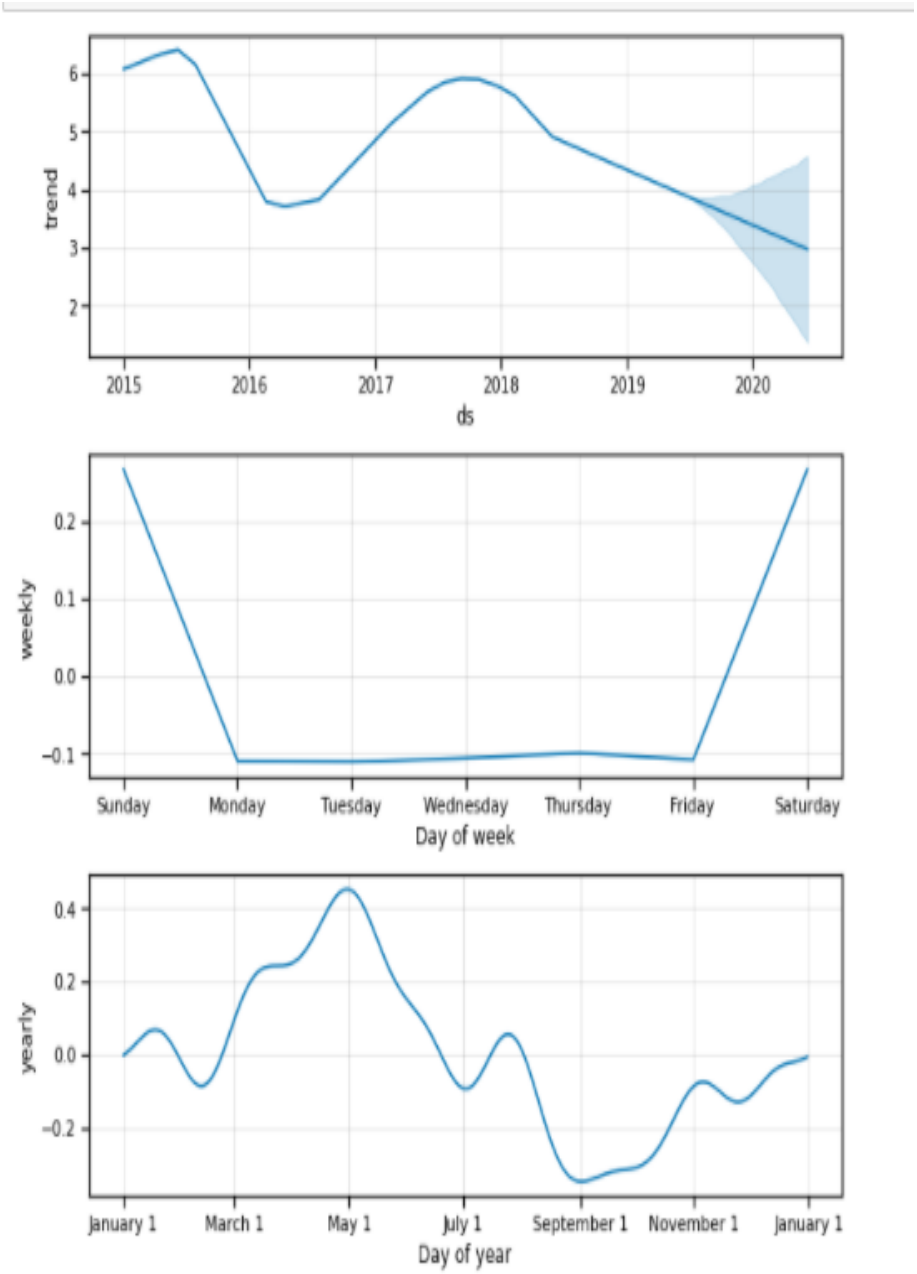

Fig 4. Graph of predicted prices

7. Future Work

In the future, we will investigate the different values of time in much more detail that in turn will aid the time-scale modelling problems. Trading techniques can also be paired with our forecast model to build a full trading framework. Where we want to work on the accuracy and the reliability of the prediction of stock. This will further enhance the system and will give more confidence to use the system.

8. Conclusion

Our paper proposes a model for market pattern prediction that takes advantage of the Prophet library based on time-series, time-varying significance functions for details, and insights from classic financial theories. The type of pattern was formally described in conjunction with Time Series Theory. The theory that closer evidence has a stronger impact on estimation was then confirmed in this experiment. The findings proposed that there may be a quasilinear association between the value of data and its time sequence.

9. References

[1] S. Zohren, Z. Zhang, and S. Roberts "Deep Convolutional Neural Networks for Limit Order Books" IEEE Access, vol. 67, do 10.1109/TSP.2019.2907260,08.10.2018
[2] S. N. Shah and, F. E. T. Burton "Efficient market hypothesis," CMT Level I 2017: An Introduction to Technical Analysis, 2017.
[3] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," Expert Systems with Applications, vol. 42, no. 20, pp. 7046–7056, 2015.
[4] J. Agrawal, V. Chourasia, and A. Mittra, "State-of-the-art in stock prediction techniques," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, no. 4, pp. 1360–1366, 2013.
[5] J. S. Armstrong, ''Forecasting with econometric methods: Folklore versus fact,'' J. Bus., vol. 51, no. 4, pp. 549–564, Oct. 1978.
[6] X. Zhang, X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, ''Stock market prediction via multi-source multiple instance learning,'' IEEE Access, vol. 6, pp. 50720–50728, 2018
[7] X. Li, H. Xie, R. Y. K. Lau, T. Wong, and F. Wang, ''Stock prediction via sentimental transfer learning,'' IEEE Access, vol. 6, pp. 73110–73118, 2018.