Journal of CUDIMAC (J-CUDIMAC)
ISSN 0794-4764 (Print) ISSN 2651-6063 (Online)
Vol. 10, No.1, September, 2022

https://cudimac.unn.edu.ng/volume-10/

Pattern Recognition and Prediction of Univariate Time Series

[1]Nwajiobi, Emmanuel N. and [2]Nwangwu, Emmanuel C.

[1]*Department of Computer Science, Nwafor Orizu College of Education, Nsugbe, Anambra State, Nigeria*
[2]*Department of Computer and Robotics Education, University of Nigeria, Nsukka, Enugu State, Nigeria*

Abstract

These days, there is massively evolving data which are represented as observations made sequentially in time from businesses, economic activities and scientific researches. Data that are calculated, measured or observed sequentially on a regular basis about business or economic activity over a period of time is called time series. There is therefore need, to find and extract recurring patterns within the time series temporal databases and also predict future patterns. In view of the forgoing, this research was carried out to: (1) find an efficient method to represent univariate time series (2) detect and extract recurring patterns and, (3) finally predict future patterns in a univariate time series. To achieve these research interests, many representation algorithms were reviewed and finally a novel datamining-based time series pattern recognition model was developed. The model was tested and validated with historical time series data obtained from the Nigerian Stock Exchange (NSE) and Yahoo finance websites. Our model uniquely represented patterns with three symbols (U, D, F for Up, Down and Flat respectively). Out of the fifteen (15) datasets used in the experiment, the model predicted twelve (12) correctly (which is 80% accuracy). Two methodologies were employed in the design and implementation of the system, namely: prototyping and object-oriented analysis and design methodology (OOADM); while the programming tools used consisted of Wampserver64 (virtual host server), MYSQL (for database implementation), PHP (for backend production), HTML, JavaScript and CSS for front end development.

Keywords: Data Mining, Pattern Recognition, Pattern, Time Series, Pattern Extraction, Time Series Representation

Introduction

At present, majority of activities in businesses, companies and organisations generate large amounts of data which are typically saved in databases. This is made possible by the availability of large storage systems, fast computer systems and efficient information systems. However, the question of what to do with such huge amounts of data is not always easily obvious to the owners of such large databases. This therefore calls for efficient computational algorithms to analyse the data to find patterns and relationships buried in data. Massive data sets are rarely profitable; their real worth lies in the possibility to extract useful information for making decisions or for understanding the phenomena that generated such data.

Consequent upon the foregoing, information retrieval is no longer enough anymore for decision-making. Therefore, the availability of these huge collections of data have now created new needs that will help us make better and informed decisions, including making predictions about the future. These new needs include but not limited to: automatic summarization of data, extraction of information buried in stored data, discovery of patterns, and prediction of future patterns. Data mining techniques can be used to discover patterns from large datasets, which are mostly in form of time series.

Time series is a collection of observations made sequentially in time (Abdullah, 2016). It is an ordered sequence of values (real numbers) of a variable or variables measured, observed or calculated at regular time intervals over a period of time. According to Pohl and Bouchachia (2012), the following activities can be performed on a time series data: detecting motifs, recognizing and extracting patterns, finding correlation between time series or finding similar time series. Similarly, analysis of a time series can be said to comprise three processing steps, namely: (a) Abstraction (or representation), (b) Mining and Discovery of trends and patterns, and (c) Prediction (Pohl and Bouchachia, 2012).

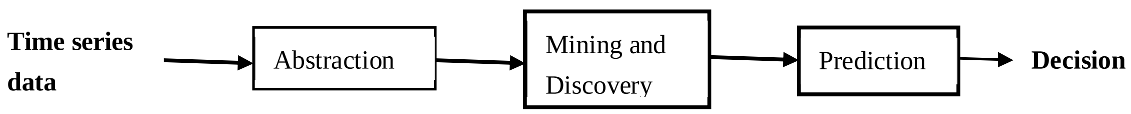

Figure. 1: Processing stages in time series analysis
Source: Pohl and Bouchachia (2012)

Any information of the sequential nature can be processed by pattern recognition algorithms to make the sequences comprehensible and enable its practical use. The term pattern recognition connotes automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities actions such as classifying the data into different categories can be taken (Bishop, 2006). These regularities in data are referred to as patterns in this paper.

In view of the above background information, this research work was able to develop a model for representation, detection and prediction of univariate time series. The type of time series data considered in this work were mostly those that can generate forecasts, like stock closing price. Time series datasets collected via yahoo finance website from different sources were used to test and validate the model.

Statement of the Problem

The goals of any time series analysis or data mining tasks on time series are usually to identify the nature of the phenomenon represented by the sequence of observations and possibly predict the future values of the time series variable. Incidentally, both goals require that the pattern of the observed time series be identified. Some studies on time series domain point to the fact that patterns in time series can repeat themselves. Therefore, detection of patterns similar to those that have occurred in the past can readily provide useful information about the future of time series evolution. It is against this backdrop that this study sought to develop a model that can detect, extract and predict specific patterns of interest from large discrete univariate time series datasets. Specific

Objectives of the Study

1. Provision of a new symbolic representation scheme for time series.
2. Provision an algorithm to "mine" time series to discover patterns
3. Extracting specific patterns of interest and storing same in a database
4. Use extracted patterns in the database to predict future time series patterns

Literature Review

Time series according to Nguyen and Duong (2007) is a sequence of real numbers, each number representing a value at a time point. It is a collection of observations made sequentially in time (Abdullah, 2016). A time series is an ordered sequence of values of a variable (univariate) or many variables (multivariate) measured, observed or calculated at equally spaced time intervals over a period of time. It consists of a sequence of values and their corresponding timestamps (i.e. the time at which the values were observed or measured).

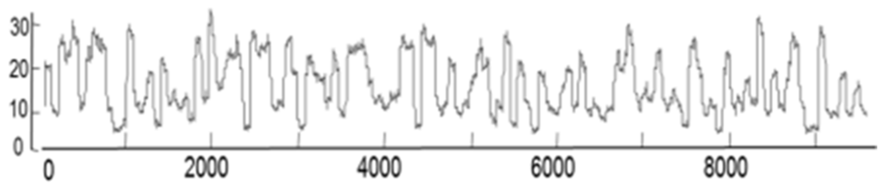

Figure 2: A typical illustration of a time series plot. On the Y-axis are the values, and along the X-axis are time stamps
Source: Lin et al, (2002)

Time series representation is concerned with finding a suitable way to represent it for further computational analysis. The process creates a more compact representation of the data, while at the same time preserving the information content of the original time series. Good representation fosters further time series analysis towards discovering patterns and making informed decisions.

Kimoto, Asakawa, Yoda and Takeoka (1990) used several learning algorithms and prediction methods to predict the Tokyo Stock Exchange Prices Index (TOPIX). The proposed system used neural network that learned the relationships between the various factors. The output of the system was the best time to buy and sell stocks. They executed simulation of buy and sell stocks to evaluate the system. In their study, vector curve, turnover ratio, foreign exchange rate and interest rate were used as input variables. Trading profit using the system proved better than using ordinary buy and hold strategy.

Trippi and Desieno (1992) in their work performed daily prediction of up and down direction of S&P 500 Index Futures using artificial neural network (ANN). Input variables in the study were technical variables for a two-week period to the trading day: open, high, low, close price, and the price fifteen minutes after the market opening of the current trading day. The output variable was a long or short recommendation. They performed composite rule generation procedure to generate rules for combining outputs of networks. They reported prediction accuracy of 45.3% to 52.8%. Singh (2000) addressed the problem of time series representation by creating an algorithm called binary representation, in which "1" was used to represent increase and "0" was used to represent decrease. It partially solved the problem of time series representation by transforming it into strings of ones and zeros for further processing. It did not address the issue of patterns in time series. What about a situation where there exist consecutive increases or decreases? The model was silent on that. Nguyen and Duong, (2007) proposed the use of Piecewise Linear Approximation (PLA) to segment time series, as a preprocessing approach necessary for further analysis. The approach represents a time series with straight lines. PLA refers to the approximation of a time series T, of length n with k straight lines (where $k < n$) (Nguyen and Duong, 2007). The PLA is composed of a series of segments representing the trend (up and down) of the raw data. Thus, PLA approximates a time series into a representation of linear segments that is efficient to manipulate and faster to process than the raw data.

Robert et al (2009), established a financial time series forecasting model by clustering stocks in Taiwan Stock Exchange Corporation (TSEC). The forecasting model integrated a data clustering technique, a fuzzy decision tree (FDT), and a genetic algorithm (GA) to construct a decision-making system based on historical data and technical indexes. The set of historical data was divided into k sub-clusters by adopting K-means algorithm. GA was then applied to evolve the number of fuzzy terms for each input index in Fuzzy Decision Tree so that the forecasting accuracy of the model can be further improved. Different forecasting models were generated for each sub-cluster. According to their study, the proposed Genetic Algorithm Fuzzy Decision Tree (GAFDT) model had the best performance when compared with other approaches on various stocks in TSEC.

Álvarez (2010) proposed a clustering approach to find patterns in electricity time series. He applied K-means, Expectation Maximization (EM) and Fuzzy C-Means (FCM) clustering techniques to find patterns in stock market data and electricity pricing data. The model proposed can be used to forecast a stock market and electricity pricing time series as recorded in the study. The approach did not delve into the use of large historical data to find patterns necessary for pattern extraction and prediction. No definite means of extracting patterns from a historical database.

Lee, Lin, Kao and Chen (2010) proposed an effective approach to stock market prediction. The method they proposed converted each financial report to feature vector and used hierarchical agglomerative clustering to divide the feature vector into clusters and then applied K-means for each sub-cluster so that most feature vectors in each sub-cluster belonged to the same class. Then, for each sub-cluster, a centroid was chosen as the representative feature vector and finally this feature vector was employed to predict the stock price movements.

Jiangling et al, (2011) developed a novel time series segmentation method that was based on turning points to extract trends from the maximum or minimum points of the time series. It was a very solid and useful idea for detecting patterns in a time series. It segmented time series into up and down structure that minimized destruction of the original underlying trends in the dataset. It did not address the issue of how to symbolically (or otherwise) represent the time series or the discovered trends.

Babu, Geethanjali and Satyanarayana (2012) proposed the use of an effective clustering method, HRK (Hierarchical agglomerative and Recursive K-means clustering) to predict the short-term stock price movements. They used the proposed framework to classify stock time series based on similarity in their price trends. Result of their model HRK outperform support vector machine (SVM) in terms of accuracy and average profit, even as their work used financial report as features. Senthamarai; Sailapathi; Mohamed and Arumugam (2012) proposed techniques which were able to predict whether future closing stock price will increase or decrease. They combined five methods of analyzing stocks to predict if the day's closing price of a stock would increase or decrease. The methods are Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Average (MA). The results of their technique showed that the algorithm was able to predict if the following day's closing price would increase or decrease. The algorithm performed well on half of the stocks and not so well on the other half of the stocks since it was able to generate both increase and decrease predictions. Thus, the algorithm could perhaps be used as a buying or selling signal, or be used to give confidence to a trader's prediction of stock prices.

Prasanna, S. and Ezhilmaran, D. (2013). Performed analysis of past and present financial data to generate patterns and decision-making algorithms using artificial intelligence and data mining techniques. The study was able to establish that data mining can be applied in evaluating past stock prices and acquire valuable information. The weakness of the study was its inability to define the type of patterns that can be generated and how to represent them. Badhiye, et al (2015) addressed time series representation to facilitate data mining of large time series databases. The method used symbolic piecewise trend approximation to represent the original dataset. It achieved dimensional reduction, and was able to symbolically represent time series dataset. The shortcoming of the approach was classification of trend into two: up and down only. It ignored the existence of flat trend, and lacked the ability to predict future trend.

Keogh Eamonn and Jessica Lin in 2002 invented SAX. SAX stands for Symbolic Aggregate Approximation. It was the first, and a novel symbolic representation for a time series. SAX is a symbolization method that involves placing a symbol for each segment obtained by using PAA, since it is based on the Piecewise Aggregate Approximation (PAA) representation. The PAA representation is merely an intermediate step required to obtain SAX. SAX uses alphabet symbols (a – z) to represent segments obtained through PAA. In order to place the symbols, it is essential to specify the number of symbols to be used and the intervals (or breakpoints) of the values for each symbol. To this end, Burcu, et al (2011) stated that the number of symbols to be used is generally determined by an expert having knowledge about the application domain under study. However, to help solve the problem of specifying the intervals (breakpoints) for each symbol, Burcu, et al (2011) suggested the use of histograms of the data values, see figure 3 below.

Figure 3: Histogram of segment values to help determine breakpoints

Another way to surmount the problem of determining breakpoints is to make use of the predefined statistical table. Table 2.1 below shows a typical predefined lookup statistical table for 3 to 10 alphabets.

β_i \ a	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Table 1: Lookup table from a pre-defined statistical table that contains the breakpoints (β1 .. β9) for alphabet size a = 3 to 10 that divides a Gaussian distribution into an arbitrary number.
Source: Lin et al. (2003).

SAX normalizes data in order to transform the series into a Gaussian distribution so that the breakpoints can be determined from the curve in accordance with the required alphabet size. SAX has also the potential for dimensionality reduction. Kuo-Ping, W; Yung-Piao, W and Hahn-Ming, L. (2014) presented a model to predict the stock trend based on a combination of sequential chart pattern, K-means and AprioriAll algorithm. The stock price sequence was cut short into charts by sliding window. The resulting charts were clustered by K-means algorithm to form chart patterns. Thus, the chart sequences were now successfully converted into chart pattern sequences, such that the frequent patterns in the sequences can be extracted by AprioriAll algorithm. The existence of frequent patterns implies that some specific market behaviors often appear, therefore, the corresponding trend can be predicted. Experimental results showed that the proposed system can produce better index return with fewer trades. As a result, the proposed method can make profits on the real market, even in a long-term usage.

Shunrong, Haomiao and Tongda (2015) proposed the use of data collected from different global financial markets as the input features to a machine learning algorithm such as support vector machine (SVM) to predict the stock market index movement. Various machine learning based models were proposed for predicting daily trend of US stocks, and numerical results obtained suggested high accuracy. In addition, a practical trading model was built upon their trained predictor and the model generated higher profit compared to selected benchmarks. Hence, they were convinced that index value of stock markets and commodity prices can provide useful information in the prediction process.

Methodology

This study employed two major types of used in the development of computer programs. They are: (1) Prototyping (2) Object-Oriented Analysis and Design Method (OOADM). They were used in the development of the pattern recognition model for times series representation, extraction and prediction of patterns. Software tools applied in the development of the pattern recognition model included WampServer, MySQL, HTML 5, CSS 3, Notepad plus, Star UML and Microsoft Excel.

For ease of identification of patterns, extraction, representation and prediction, we predefined patterns as either Up (U), Down (D) or Flat (F) and represented them using UDF symbols. Each pattern represents a segment, and can be drawn as shown in figure 4 below. The algorithm begins with a historical time series dataset which it receives as input. Prior to that, the dataset should have been preprocessed by removing blank cells of data and transforming (normalizing) the dataset into the range [0,1], such that the highest value is 1 and the least value in the series is 0. After this normalization process, the pattern recognition algorithm can be applied to the resulting dataset to detect patterns of interest and thus represent them with symbols. All patterns identified are symbolized and stored in a database for future uses and manipulation.

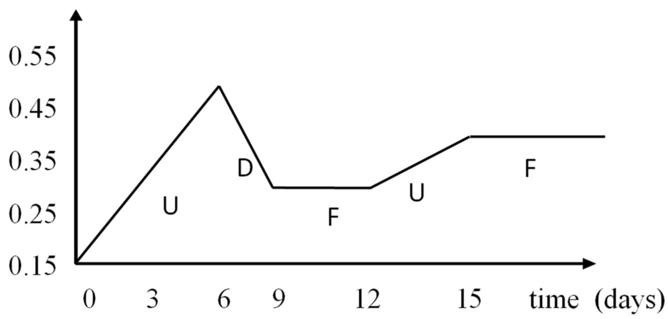

Figure 4: Visualisation of patterns of a time series, showing the up, down and flat patterns. (**Courtesy of** Nwajiobi E.N, 2019.), where value = average value for segments: Up, Down or Flat pattern

From figure 4 above, the symbolic representation of the time series is UDFUF. So, a time series of length 25 (data points) has been reduced to a string of UDFUF (which is five characters).

The algorithm for pattern detection, extraction and symbolic representation:

Input: *S, Segment_size*
Output: *Pattern string symbols (for Up, Down and Flat patterns)*
Repeat
Initialize tup = tdn = 0;
For (i = 0; i ≤ Segment_size - 1, i ++) {
df = (i + 1) – i;
if df is positive, tup++ //augment increasing pattern variable
if df is negative, tdn++ //augment decreasing pattern variable
}
If tup = Segment_size - 1 or tdn = Segment_size - 1 then, pattern is Up or Down respectively
Calculate segment average; //Call SegmentAvg ( ) function
Store segment MinDate, MinValue MaxDate, MaxValue, Segment_Symbol (U,D,F), SegmentAvg
Until end of S is reached.

The Prediction algorithm

1. *Start*
2. *Enter new set of data*
3. *Detect pattern of the new set of data (pat)*
4. *Calculate its average (avg)*
5. *Open the patterns turning point database table*
6. *Find match for average (avg) and pattern (pat) in the database*
7. *If match is found, get the next pattern ('np') in the database*
8. *If match is not found, then increment or decrement average (avg) until a match is found*
9. *Display predicted future pattern*
10. *return*

Results

The model was tested with real-valued discrete univariate time series data, mostly stock market data, obtained online from the Nigerian Stock Exchange (NSE) and Yahoo finance website. The model achieved 100% success in detecting and extracting the three pre-defined patterns and representing the patterns with symbols (UDF). Out of the 15 different datasets used for experimental prediction, the system was able to predict 12 correctly and missed 3. Thus, it achieved 80% success, which is an impressive and acceptable outcome. Thus, the model achieved all the stated objectives, which included time series representation using symbols, detection, extraction and prediction of patterns. The results of this study showed also that patterns not only exist, but can also can repeat over time. Thus, a time series can be represented by a string of UUDFFDUDFF instead of the numerical values. By converting time series into patterns and storing same symbolically using the three alphabets, the

Journal of CUDIMAC (J-CUDIMAC)
ISSN 0794-4764 (Print) ISSN 2651-6063 (Online)
Vol. 10, No.1, September, 2022

system thus achieved dimensionality reduction of the size of time series dataset. Below are some of the outputs from the system.

Record No	Date	Value	Normalised Value	Year
1	2000-01-03	11357.5097656250	0.8104355931	2000
2	2000-01-04	10997.9296875000	0.6239265800	2000
3	2000-01-05	11122.6503906250	0.6886174083	2000
4	2000-01-06	11253.2597656250	0.7563626170	2000
5	2000-01-07	11522.5595703125	0.8960445523	2000
6	2000-01-10	11572.2001953125	0.9217924476	2000
7	2000-01-11	11511.0800781250	0.8900903463	2000
8	2000-01-12	11551.0996093750	0.9108479023	2000
9	2000-01-13	11582.4296875000	0.9270983338	2000
10	2000-01-14	11722.9804687500	1.0000000000	2000
11	2000-01-18	11560.7197265625	0.9158377051	2000
12	2000-01-19	11489.3603515625	0.8788245916	2000
13	2000-01-20	[illegible]	[illegible]	2000

Table 1: Sample raw and normalized time series data. It has 5158 records (data points).

FUFUFDDUDUFDFDUDFDFUUDUFFFUFDFDFFFFDUFUDFDFFFUFFDF
FFFUFDFUFFFFFUDUUUDFDFFUUUFUFFFUUFFUFDFDDFFDFUFDDF
FFFUFUUDFDDFFDFUFDUDUUDFFUUFFDFFFFFFFUFFDFDFDFFDUU
UFFFFDFUFFFFUFUUDFUFDFDFFUUDFFUUFDFDFUDDFFDFUFFFDU
UDFFFFFDFFFFDFUDDDDUFFUFDUFFFFUFDUUFFUFFFFFFUDDFUU
UFUUFUDDDFFFUFUDDFFUDFUFFFFUFUFDDUDDUDFFFFDFDUFFFF
FFUDUDDFFDFFFFDFFDFFDDDFDFFFFDUUDUFFFFDDFUFDFDFFFF
FDDFUUFUFFFFUUDFUFFFFFDDFUDFDFDFUFDUUDDFDUFFDFDFFF
FFFFFFDUUUUFDDUFFFFFFUDUDFFFFFDFUUFUUFUFUDFFFFFUDU
UDFFFFDFFUUFUUFDFUUFFFFFFFDFFUUFFFDFFFUUFFDFDDFUUU
UFFFFFUUUFFFFFFFUFDFFFFUFFFDDFFFFFDFFFDDFUFUFFDUDU
UFFDUDDUFFDFFFUFUFFDFUDFFDFUFDFFDUUFFDFFUFFUFFFFDU
UFUFDFUFFFFDFFDDUUFUUDUFUFUFFFDUFUFUUFDDDFFFUDDUDU

Figure 5: Showing the symbolic representation of the extracted patterns

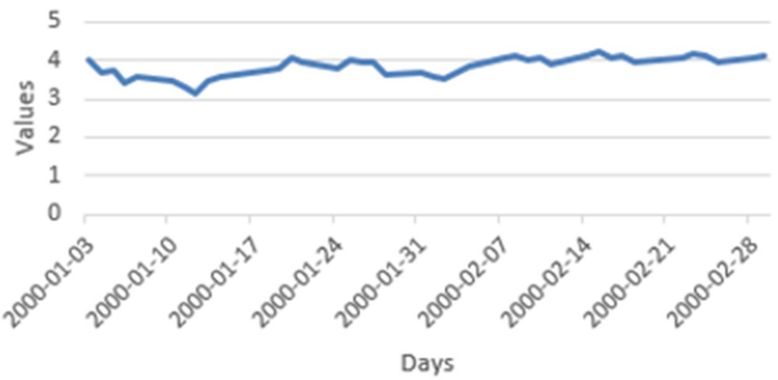

Figure 6: This shows the raw data plot without patterns extracted for two (2) months

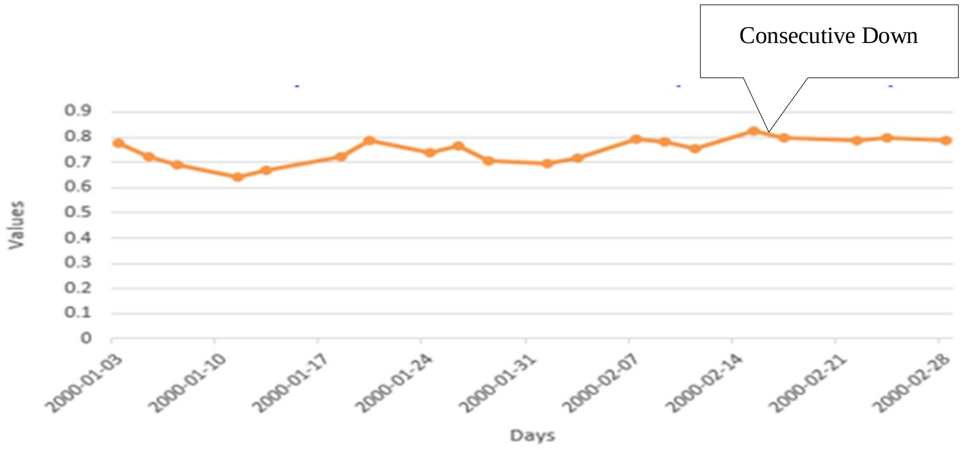

Figure 7: This shows plot of the extracted patterns for the same 2 months.

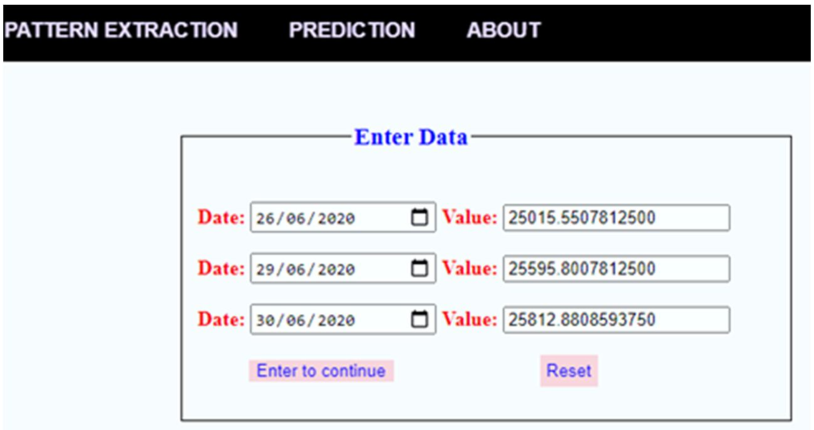

Figure 8: This is the pattern prediction window.

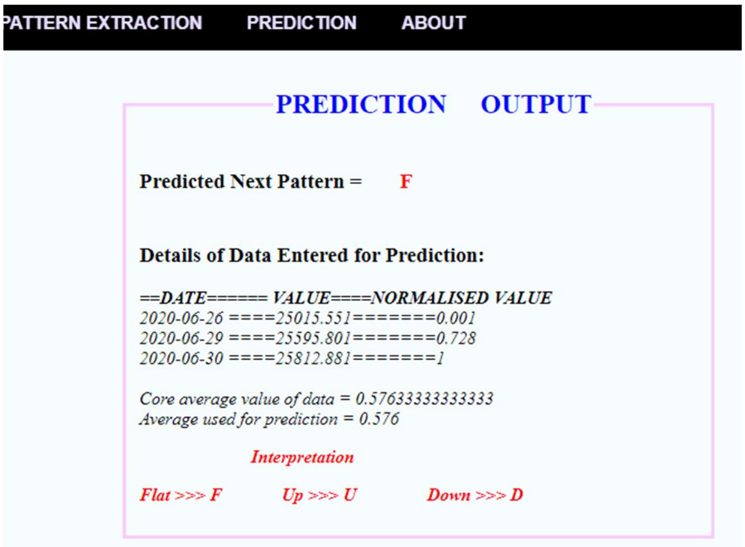

Figure 9: The prediction output from the system.

Conclusion

This research has shown that through the process of data mining, that patterns: (a) can be identified in a univariate discrete time series (b) can be predicted (c) repeats in historical datasets. The aim and all the stated objectives of the research were met. Thus, with the new model developed, time series data miners and the general time series community have gained knowledge that were once not obvious. Researchers in time series analysis from different perspectives like statistics, economics and data mining will benefit immensely from the contributions of this work. Further research can be carried out to explore the applicability of the algorithms to other time series domains aside stock market data. In addition, consideration of multivariate time series datasets can also be explored.

References

Abdullah, M; Suman, N and Jie, L (2016). *Similarity Search on TS Data: Past, Present and Future*. CIKM2016 Tutorial, obtained from http://www.cs.unm.edu/~mueen/Tutorial/CIKM2016Tutorial.pdf on 17/4/2017.

Álvarez, F.M. (2010) *Pattern sequence analysis to forecast time series*. Unpublished thesis work. Universidad Pablo De Olavide De Sevilla

Babu, M.S; Geethanjali, N and Satyanarayana, B (2012). Clustering Approach to Stock Market Prediction. *International Journal of Advanced Networking and Applications*, 03, (04), 1281-1291

Badhiye, S.S; Hatwar, K, S. and Chatur, P.N (2015). Trend based Approach for Time Series Representation. *International Journal of Computer Applications*, 113, (16), (0975 – 8887).

Bishop, C.M (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer Science+Business Media, LLC.

Burcu, K; Serhan, O and Bora K (2011). *Application of Symbolic Piecewise Aggregate Approximation (PAA) Analysis to ECG Signals*. Artificial Intelligence & Design Lab, Mechanical Engineering Department, Computer Engineering Department, Izmir Institute of Technology, 35430 Izmir, Turkey. A pdf file obtained 15/4/2017

Jiangling, Y; Yain-Whar, Si and Zhiguo, G (2011). *Financial Time Series Segmentation Based on Turning Point*s. Proceedings of 2011 International Conference on System Science and Engineering, Macau, China - June 2011.

Jingpei, D; Weiren, S; Fangyan, D and Kaoru, H (2013). Piecewise Trend Approximation: A Ratio-Based Time Series Representation. *Journal of Abstract and Applied Analysis*, 2013. http://dx.doi.org/10.1155/2013/603629

Kimoto, T.; Asakawa, K.; Yoda, M. and Takeoka, M., (1990). *Stock market prediction system with modular neural networks*. Proceedings of the IEEE International Joint Conference on Neural Networks, 1990, pp. 11-16.

Kuo-Ping, W; Yung-Piao, W and Hahn-Ming, L. (2014). Stock Trend Prediction by Using K-Means and AprioriAll Algorithm for Sequential Chart Pattern Mining. *Journal Of Information Science and Engineering,* 30, 653-667.

Lee A. J. T; Lin M.-C.; Kao R.-T.; and Chen K.-T. (2010). *An Effective Clustering Approach to Stock Market Prediction*. PACIS 2010 Proceedings. 54.

Nguyen, Q.V.H and Duong, T.A (2007). *Combining SAX and Piecewise Linear Approximation to Improve Similarity Search on Financial Time Series*. International Symposium on Information Technology Convergence, IEEE Computer Society.

Pohl, D., Bouchachia, A. (2013). *Financial Time Series Processing: A Roadmap of Online and Offline Methods*. In: Rausch, P., Sheta, A., Ayesh, A. (eds) Business Intelligence and Performance Management. Advanced Information and Knowledge Processing. Springer, London. https://doi.org/10.1007/978-1-4471-4866-1_10.

Prasanna, S. and Ezhilmaran, D. (2013). An analysis on Stock Market Prediction using Data Mining Techniques. *International Journal of Computer Science & Engineering Technology* (IJCSET), 4 (02); 49-51. https://www.ijcset.com/docs/IJCSET13-04-02-004.pdf

Robert K.L; Chin-Yuan F.; Wei-Hsiu H and Pei-Chann C (2009). Evolving and clustering fuzzy decision tree for financial time series data forecasting. *An International Journal of Expert Systems with Applications*, 36, (2), 3761-3773. https://www.sciencedirect.com/science/article/abs/pii/S0957417408001474

Senthamarai, K.K.; Sailapathi, P.S; Mohamed, M.S and Arumugam, P. (2012). Financial Stock Market Forecast using Data Mining Techniques. *IJCSNS International Journal of Computer Science and Network Security*, 7(12), 555-559. https://www.iaeng.org/publication/IMECS2010/IMECS2010_pp555-559.pdf

Sameer Singh (2000) Pattern modelling in time-series forecasting. *Cybernetics and Systems*, 31:1, 49-65, DOI: 10.1080/019697200124919

Shunrong Shen, H.J. and Tongda, Z. (2015) Stock Market Forecasting Using Machine Learning Algorithms. *Journal of Computers & Operations Research*, 32, 2513–2522. https://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf

Trippi, R. and Desieno, D. (1992). Trading equity index futures with a neural network. *The Journal of Portfolio Management*, 5 (3). DOI:10.3905/jpm.1992.409432