Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022
Vol.4, Issue.1

Weka for Stock Market Analysis and Visualization

Gretta Anna Abraham
Department of Computer Applications
Amal Jyothi College Of Engineering
Kanjirapally,Kottayam
grettaabraham72@gmail.com

Gloriya Mathew
Assistant Professor
Amal Jyothi College Of Engineering
Kanjirapally,Kottayam
gloriyamathew@amaljyothi.ac.in

Abstract—Knowledge and mindfulness influence stock investment attitudes. including literal, the stock request includes a large amount of data that changes over time. Stock prices are influenced by a variety of factors, ranging from the company's performance to the state of the economy in general. As a result, in order to manage investment portfolios, stock request data must be reviewed on a regular basis in order to detect implicit relationships between colored stocks and, as a result, to change investment based on movements in connected companies. The graphical depiction of numerical data is aided by data visualization.

I. INTRODUCTION

The stock market has a major impact on the economy of any country. Stocks are frequently utilised as an investment by people who wish to make money from them. The risk-to-reward ratio of stock market investing is quite high, and there is a lot of data that changes over time in the stock market. A number of things have an impact on stock prices, ranging from the company's prosperity to the economy's overall state. A single poor decision can lead to large losses, whilst a single good decision can lead to significant profits. We must select the suitable stock to ensure that our investment decisions are not disastrous. Options for displaying data such as closing prices, highs, and lows have been incorporated.

Time series analysis is a method for modelling and understanding a mathematical approach for calculating a time-dependent series of data points the approach of utilizing a model to incorporate predictions (forecasts) for future events that are based on known historical events is known as time series forecasting. Unlike standard data mining and machine learning applications, When each data point represents a separate representation of the concept to be learnt, and the data is presented in a logical order point within a data collection is irrelevant, time series data has a natural temporal ordering. Capacity planning, inventory restocking, revenue forecasting, and future workforce levels are all examples of time series applications. Weka now has a specialized stochastic analysis environment (>= 3.7.3) for developing, evaluating, and visualizing prediction models. The "Explorer" user interface is installed in this environment. The package manager can be used to install Weka's graphical "Explorer" user interface. The Weka Time Series Framework employs machine learning and data mining to transform data into meaningful information. A format that can be used by classic propositional learning approaches to model time series. This is accomplished by removing the temporal ordering of individual put samples and recording a separate input field for the time dependency "LAGGED" variables are another name for these variables. In order for the algorithms to model trends and seasonality, many more fields are automatically generated Any of the After the data has been converted, Weka regression algorithms can be used to learn a model. Multiple linear regression is a natural choice. However, any method that can forecast a continuous goal can be used.

II. LINEAR REGRESSION

By fitting a simple equation to observable data, linear regression seeks to model the connection between two variables.

A scatterplot can help you figure out how strong a correlation between two variables is If it appears that the hypothesized explanatory and dependent variables are unconnected (i.e., the scatterplot shows no increasing or decreasing patterns). It's also improbable that analysing the data with a linear regression model will give a meaningful model. The correlation coefficient, which can vary from -1 to 1, indicates how closely the two variables' observed data are related . A helpful numerical measure of the link between two variables is the correlation coefficient.

$Y = a + bx$ is the equation for a linear regression line, where X is the explanatory variable and Y is the dependent variable. The line has a slope of b and an intercept of c (y at x = 0).

DOI: 10.5281/zenodo.6369862
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering KanjiraKottayamppally,

III. IMPLEMENTATION

- Download Weka as well as the time series analysis plugin and install them.
- Once installation completed launch Weka explorer
- The interface is shown in Fig. 1.
- Open a file which you want to proceed.
- Here we use the basic configuration.(Fig. 2)
- Choose the field(s) in the data that they want to target. forecast on the forecast tab.
- Set the number of time units to forecast
- Now is the moment to put your plan into action after you've filled all of the relevant fields. Look for a start button on the left bottom to continue.
- The training evaluation which includes graphs that show anticipated values beyond the training data's end textual projected values, and a textual description of the learnt model are all instances of output generated by basic configuration panel settings

Table I. Dataset

	A	B	C	D	E	F	
1	date	open	high	low	close	volume	Nam
2	01-02-2018	72.54	72.59	71.465	71.77	2135564	ZTS
3	01-03-2018	71.77	72.18	71.34	72.1	2328506	ZTS
4	01-04-2018	72.75	72.9128	72.26	72.53	2534010	ZTS
5	01-05-2018	72.83	73.92	72.57	73.36	2166065	ZTS
6	01-08-2018	73.43	74.42	73.1607	74.24	3631552	ZTS
7	01-09-2018	74.7	75.475	74.465	75.11	2721946	ZTS
8	01-10-2018	74.58	74.8	73.28	73.91	2257464	ZTS
9	01-11-2018	74.05	74.68	73.72	74.59	1629279	ZTS
10	01-12-2018	74.91	75.69	74.77	75.39	1915669	ZTS
11	1/16/2018	76.06	76.4	75.21	75.54	2599094	ZTS
12	1/17/2018	75.81	77.03	75.39	76.77	2769587	ZTS
13	1/18/2018	76.65	76.65	75.85	76.33	2588995	ZTS
14	1/19/2018	76.69	76.91	76.22	76.62	4829602	ZTS
15	1/22/2018	76.67	77.58	76.62	77.48	4195103	ZTS
16	1/23/2018	77.3	78.32	76.97	77.59	2264132	ZTS
17	1/24/2018	78.16	78.51	77.6	78.33	2585326	ZTS
18	1/25/2018	78.47	79.38	78.345	79.25	2327262	ZTS
19	1/26/2018	79.49	80.13	79.38	80.09	2532808	ZTS
20	1/29/2018	79.81	79.95	79.11	79.18	2662383	ZTS

Fig. 1: Weka explorer interface

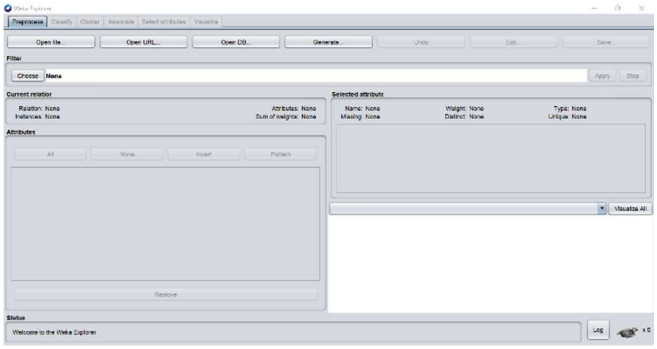


Fig. 2: Basic configuration panel

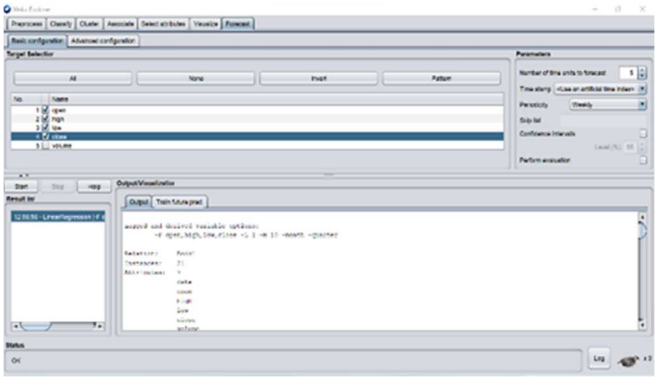
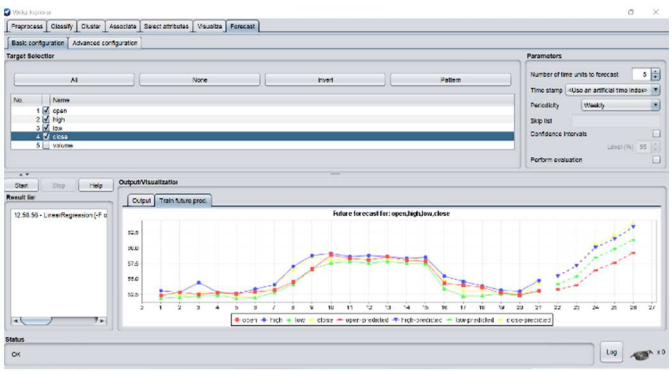


Fig. 3: Graphical representation of stock market changes



IV. CONCLUSION

Forecasting is the process of creating future forecasts based on historical and current data as well as the most common trend analysis. Time series forecasting comes in handy when we need to make future judgments or perform analysis. A time series is a collection of data points collected at regular intervals over a period of time (seconds, minutes, hours, etc.). hours, days, months etc.) Time series data mining, despite the fact that it can provide useful information for long-term business decisions, is underutilized in most companies

V. ACKNOWLEDGMENT

DOI: 10.5281/zenodo.6369862
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering KanjiraKottayamppally,

The Amal Jyothi College of Engineering's Department of Computer Applications supported this study. Without the continuous support of my supervisor, Ms. Gloriya Mathew, this report and the research that went into it would not have been possible. Her intelligence and encouragement assist me in finishing my work.

VI. RFERENCES

[1] Wikipedia article on time series: https://en.wikipedia.org/wiki/Time series
[2] Time-series-analysis https://pestleanalysis.com/time-series-analysis/
[3] Data forecasting with WEKA's Time Series Datamining https://www.researchgate.net/publication/328203079 BY Yash Data forecasting with WEKA's Time Series Data Mining
[4] PentahoCommunity https://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka