IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12905

PROJECTING THE PRICE OF STOCKS USING REGRESSION MODEL

Vinodhini. D[1], Mr. Manikadan. N[2]

Student, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[1]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[2]

Abstract: Stock prices are first determined by a company's Initial Public Offering (IPO) when it first puts its shares into the market. Investment firms use a variety of metrics, along with the total number of shares being offered, to determine what the stock's price should be. Afterward, the several reasons mentioned above will cause the share price to rise and fall, driven largely by the earnings that can be expected from the company. Traders use financial metrics constantly to determine the value of the company, including its history of earnings, changes in the market, and the profit that it can reasonably be expected to bring in. Hence, stock price prediction has become an important research area. The aim is to predict machine learning based techniques for stock price prediction. The analysis of dataset by supervised machine learning technique (SMLT) using uni-variate analysis, bi-variate and multi-variate analysis. To propose a machine learning-based method to accurately predict the stock price. Proposed machine learning algorithm technique can be compared with best accuracy with precision, Recall and F1 Score.

Keywords: Machine learning, Regression, tesla

I. INTRODUCTION

Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision. However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data.In summary, Machine Learning Algorithms are widely utilized by many organizations in Stock market prediction. This article will walk through a simple implementation of analyzing and forecasting the stock prices of a Popular Worldwide in Python using various Machine Learning Algorithms.

A. OBJECTIVE:

The goal is to develop a machine learning model for Tesla Stock Price Prediction, to potentially replace the updatable Regression models by predicting results in the form of best accuracy by comparing supervised algorithm.

B. SCOPE:

Here the scope of the project is that integration of tesla stock with computer-based prediction could reduce errors and improve prediction outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of tesla stock price prediction.

II. RELATED WORK

PAPER (I) : The project aims to provide retail investors with a third-party investment mobile application to navigate through the stock market. This is achieved through the use of machine learning and mobile web technologies. Several stock price prediction approaches and models are developed including dense, feedforward neural networks, recurrent neural networks, simple linear regressions, and linear interpolations. Model architectures and hyperparameters are optimized and automatically searched by evolution algorithm. Promising results are found for trend prediction. The project serves as a foundation for democratizing machine learning technologies to the general public in the context of discovering investment opportunities. It paves the way for extending and testing out new models, and developing AutoML in the financial context in the future.

PAPER (II): Stock price prediction is a notoriously challenging problem. Typically, when trying to solve it, researchers and individuals use either technical (prices and volumes) or fundamental (text-based) data. However, it is exceedingly rare for both forms to be used. This work utilizes Tesla data with sentiment analysis performed on news titles pertaining to the company as well as technical data on its stock over a three year period in order to predict closing price movement.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12905

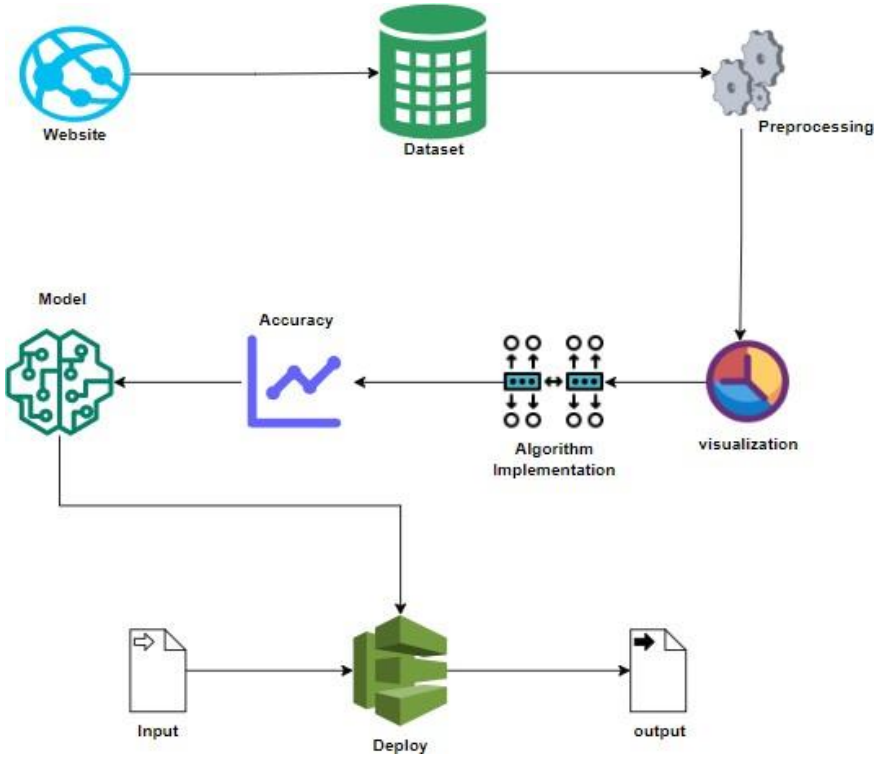
For a next day prediction, the final model architecture (a blended model) results in an accuracy of 65%, which is just under the highest accuracy observed in the literature review. Also, as time to predict goes from 1 day to 3, the GRU model does not have a large drop in accuracy, which insinuates it could be used for later predictions.

PAPER(III): Stock market is one of the most important sectors of a country's economy. Prediction of stock prices is not easy since it is not stationary in nature. The objective of this paper is to find the best possible method to predict the closing prices of stocks through a comparative study between different traditional statistical approaches and machine learning techniques. Predictions using statistical methods like Simple Moving Average, Weighted Moving Average, Exponential Smoothing, Naive approach, and machine learning methods like Linear Regression, Lasso, Ridge, K-Nearest Neighbors, Support Vector Machine, Random Forest, Single Layer Perceptron, Multi-layer Perceptron, Long Short Term Memory are performed. Moreover, a comparative study between statistical approaches and machine learning approaches has been done in terms of prediction performances and accuracy. After studying all the methods individually, the machine learning approach, especially the neural network models are found to be the most accurate for stock price prediction.

PAPER (IV): This survey starts with a general overview of the strategies for stock price change predictions based on market data and in particular Limit Order Book (LOB) data. The main discussion is devoted to the systematic analysis, comparison, and critical evaluation of the state-of-the-art studies in the research area of stock price movement predictions based on LOB data. LOB and Order Flow data are two of the most valuable information sources available to traders on the stock markets. Academic researchers are actively exploring the application of different quantitative methods and algorithms for this type of data to predict stock price movements. With the advancements in machine learning and subsequently in deep learning, the complexity and computational intensity of these models was growing, as well as the claimed predictive power. Some researchers claim accuracy of stock price movement prediction well in excess of 80%. These models are now commonly employed by automated market-making programs to set bids and ask quotes.

PAPER (V) : During the last few months, there has been increased attention in the stock market due to the Covid pandemic. The new-found leisure time has driven many people to buy and sell stocks without any knowledge on the matter at hand. The number of affiliations on investing or trading apps has increased drastically since the last year. It is natural to think that the field of predicting the stock market has increased accordingly. However, only two main approaches have been made. One focusing on day trading and using technical analysis of the markets to predict the immediate value, and the other focusing on the stocks as long-time investments and using fundamental analysis to predict the future value of the stock in the long run.

III. DESIGN ARCHITECTURE



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12905

IV. IMPLEMENTATION

A. Data Pre-processing:

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

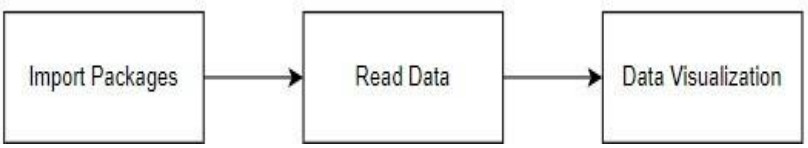
Some of these sources are just simple random mistakes. Other times, there can be a deeper reason why data is missing. It's important to understand these different types of missing data from a statistics point of view. The type of missing data will influence how to deal with filling in the missing values and to detect missing values, and do some basic imputation and detailed statistical approach for dealing with missing data. Before, joint into code, it's important to understand the sources of missing data. Here are some typical reasons why data is missing:

MODULE DIAGRAM

B. Data visualization:

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

MODULE DIAGRAM



Performance Metrics to calculate:

False Positives (FP): A person who will pay predicted as defaulter. When actual class is no and predicted class is yes. E.g. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

False Negatives (FN): A person who default predicted as payer. When actual class is yes but predicted class in no. E.g. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die.

True Positives (TP): A person who will not pay predicted as defaulter. These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

True Negatives (TN): A person who default predicted as payer. These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this passenger did not survive and predicted class tells you the same thing.

True Positive Rate(TPR) = TP / (TP + FN) False Positive rate(FPR) = FP / (FP + TN)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12905

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.
Accuracy = (TP + TN) / (TP + TN + FP + FN)
Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

Precision: The proportion of positive predictions that are actually correct. Precision = TP / (TP + FP)
Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labelled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

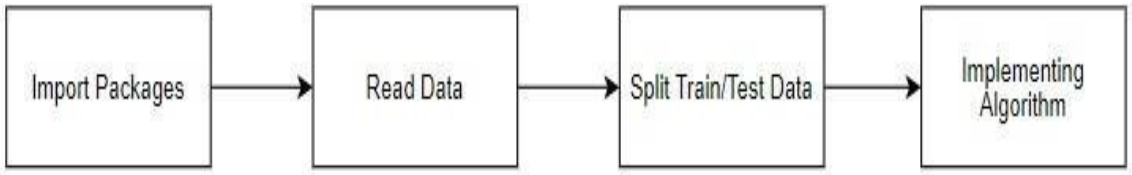
Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)
Recall = TP / (TP + FN)
Recall(Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

General Formula:
F- Measure = 2TP / (2TP + FP + FN)
F1-Score Formula:
F1 Score = 2*(Recall * Precision) / (Recall + Precision) The below 4 different algorithms are compared:

C. AdaBoost Algorithm:
AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misregressor by previous regressors. In some problems it can be less susceptible to the overfitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing, the final model can be proven to converge to a strong learner.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data
output : getting accuracy

Finding the Accuracy

```
EVS = (explained_variance_score(y_test, predictD)*100)
print("ACCURACY RESULT OF ADABOOST REGRESSOR IS :", EVS)
print("")
```

```
ACCURACY RESULT OF ADABOOST REGRESSOR IS : 99.53572767095919
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

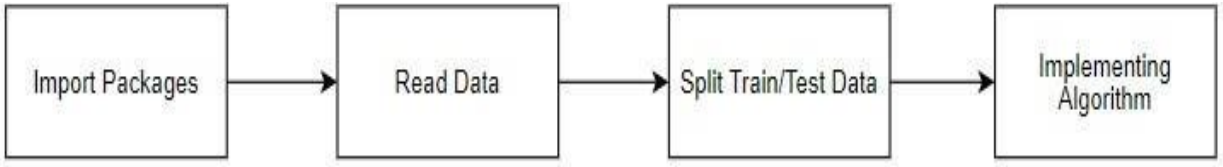
DOI: 10.17148/IJARCCE.2023.12905

D. Decision Tree Algorithm:

o Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems. It is a tree-structured regressor, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

o In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT
input : data
output: getting accuracy

Find Accuracy Score

```
R2=(r2_score(y_test,predictD)*100)
print('Accuracy result of Decision Tree regressor is :',R2)
print("")
```

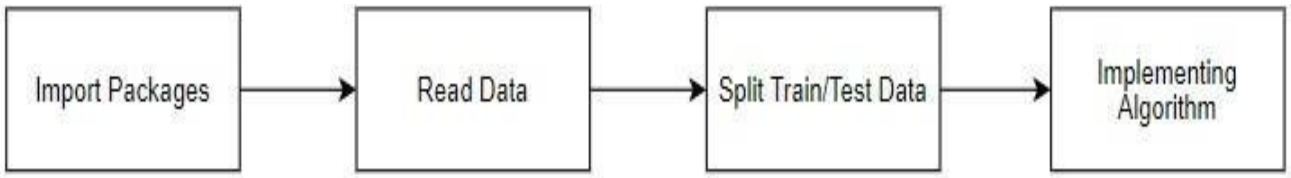
```
Accuracy result of Decision Tree regressor is : 99.06112877720524
```

E. **Ridge:**

Ridge regression is **a regularization technique, which is used to reduce the complexity of the model**. It is also called as L2 regularization. In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty.

Lasso tends to do well if there are a small number of significant parameters and the others are close to zero (ergo: when only a few predictors actually influence the response). **Ridge works well if there are many large parameters of about the same value** (ergo: when most predictors impact the response Ridge regression aims at **reducing the standard error by adding some bias in the estimates of the regression**. The reduction of the standard error in regression estimates significantly increases the reliability of the estimates

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT
input: data
output: getting accuracy

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12905

Find the R2_score

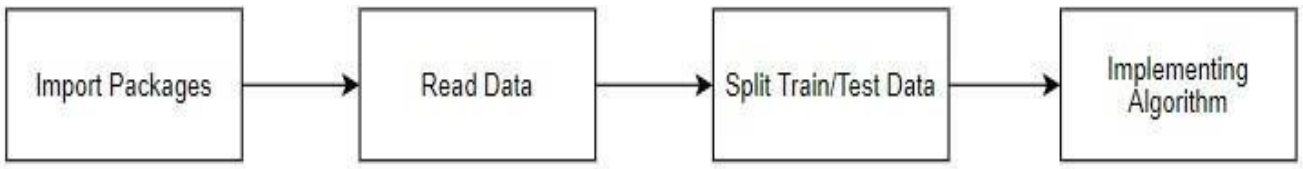
```
R2_SCORE = (r2_score(y_test, predictR)*100)
print("R2_SCORE :", R2_SCORE)
print("")
```

```
R2_SCORE : 99.93005828392613
```

F. Lasso:

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suitedfor models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination. Lasso Regression uses L1 regularization technique. It is used when we have more features because it automatically performs feature selection.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data

output : getting accuracy

Find the r2score

```
R2_SCORE = (r2_score(y_test, predictLR)*100)
print("The Accuracy of Lasso:", R2_SCORE)
print('')
```

```
The Accuracy of Lasso: 99.8798273668804
```

G. Deployment: Flask (Web FrameWork) :

Flask is a micro web framework written in Python.

It is classified as a micro-framework because it does not require particular tools or libraries.

It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

However, Flask supports extensions that can add application features as if they were implemented in Flask itself.

Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

Flask was created by Armin Ronacher of Pocoo, an international group of Python enthusiasts formed in 2004. According to Ronacher, the idea was originally an April Fool's joke that was popular enough to make into a serious application. The name is a play on the earlier Bottle framework.

Flask has become popular among Python enthusiasts. As of October 2020, it has second most stars on GitHub among Python web-development frameworks, only slightly behind Flask, and was voted the most popular web framework in the Python Developers Survey 2018.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

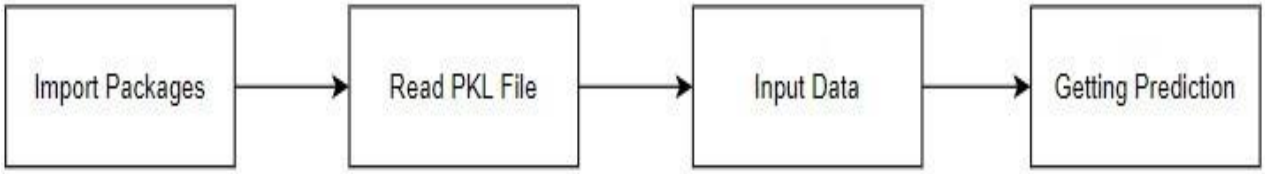
DOI: 10.17148/IJARCCE.2023.12905

Flask is based on Werkzeug, Jinja2 and inspired by Sinatra Ruby framework, available under BSD licence. It was developed at pocoo by Armin Ronacher. Although Flask is rather young compared to most Python frameworks, it holds a great promise and has already gained popularity among Python web developers. Let's take a closer look into Flask, so-called "micro" framework for Python.

Deploying the model predicting output

In this module the trained machine learning model is converted into pickle data format file (.pkl file) which is then deployed for providing better user interface and predicting the output of Human Air Pollution.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT
input : data values
output : predicting output

V. CONCLUSION AND FUTURE ENHANCEMENT

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set of higher accuracy score algorithm will be find out. The founded one is used in the application which can help to find the tesla stock price.

- ❖ Deploying the project in the cloud.
- ❖ To optimize the work to implement in the IOT system.

REFERENCES

[1]. Khodabakhsh, A., Arí, I., Bakír, M. and Ercan, A.O., 2018. Multivariate Sensor Data Analysis for Oil Refineries and Multi-mode Identification of System Behavior in Real-time. IEEE Access, 6, pp.64389- 64405
[2]. Raicharoen, T., Lursinsap, C., & Sanguanbhokai, P. (2003, May). Application of critical support vector machine to time series prediction. In Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on (Vol. 5, pp. V-V). IEEE.
[3]. Zhang, M. and Pi, D., 2017. A New Time Series Representation Model and Corresponding Similarity Measure for Fast and Accurate Similarity Detection. IEEE Access, 5, pp.24503-24519
[4]. Pati, J., Kumar, B., Manjhi, D. and Shukla, K.K., 2017. A Comparison Among ARIMA, BP- NN, and MOGA-NN for Software Clone Evolution Prediction. IEEE Access, 5, pp.11841-11851
[5]. Zhang, Q., Li, F., Long, F. and Ling, Q., 2018. Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network. IEEE Access, 6, pp.56984-56994.
[6]. Conejo, Antonio J., Miguel A. Plazas, Rosa Espinola, and Ana B. Molina. "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models." IEEE transactionson power systems 20, no. 2 (2005): 1035-1042.
[7]. Anaghi, M.F. and Norouzi, Y., 2012, December. A model for stock price forecasting basedon ARMA systems. In Advances in Computational Tools for Engineering Applications (ACTEA), 2012 2nd International Conference on (pp. 265-268). IEEE.
[8]. A. Akbar, G. Kousiouris G, H. Pervaiz, J. Sancho J, P. Ta-Shma, F. Carrez, and K. Moessner,"Real-Time Probabilistic Data Fusion for Large-Scale IoT Applications," IEEE Access, 6:10015-27, 2018.
[9]. E. Olmezogullari and I. Ari, "Online association rule mining over fast data," In Big Data (BigData Congress), IEEE International Congress on, pp. 110–117, 2013, IEEE.
[10]. E. Lughofer, M. Pratama, and I. Skrjanc, "Incremental Rule Splitting in Generalized Evolving Fuzzy Systems for Autonomous Drift Compensation," IEEE Transactions on Fuzzy Systems, 2017.