PREDICTIVE ANALYSIS OF FOOD SALES FORECASTING

[1]Kavitha T, [2]Pooja Shree C, [3]Tamilsanthiya K V
[4]Nivetha V, [5]Shunmathi N R
[1]AP, [1,2,3,4]Dept.of. CSE
Vivekanandha College Of Technology For Women. Tiruchengode, India
[1] kavitha@vcew.ac.in, [2]cpooja2102@gmail.com, [3]tsanthiya036@gmail.com, [4]nivethav603@gmail.com, [5]Shunmathi09052002@gmail.com

ABSTRACT

Time series forecasting is required for food sales prediction. This can be done with traditional statistical method like Autoregressive Integrated Moving Average (ARIMA). The estimation of future sales by companies in the food industry, such as supermarkets, grocery stores, restaurants, bakeries, and patisseries, is the focus of food sales prediction. Companies can reduce the number of stocked and expired products in their stores and avoid missing sales with accurate short-term sales prediction. In addition, it looks at machine learning algorithms that have been used to predict food sales and how to measure how accurate they are. Neural networks for prediction have attracted more interest in recent years. To forecast furniture sales, this study looks into a public dataset that includes a retail store's sales history. Several forecasting models are used to achieve this goal. First, classic time-series forecasting methods like Triple Exponential Smoothing and Seasonal Autoregressive Integrated Moving Average (SARIMA) are used. After that, more advanced techniques like SVM and Prophet are used. Different accuracy measurement techniques, such as Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), are used to compare the models' performance. The results demonstrate that the Stacked SVM approach is superior to others.

KEYWORDS: SARIMA, ARIMA, SVM, MAPE

INTRODUCTION

FORECASTING

Based on historical data and an examination of the current situation, forecasting is the process of predicting future events or trends. Individuals and organizations alike can use it to make well-informed decisions about their strategies and plans for the future. Businesses can make adjustments to their operations and remain competitive by using forecasting to anticipate changes in consumer demand, economic conditions, and industry trends. Sports, politics, finance, the weather, and forecasting are just a few of the many areas where forecasting can be applied. Forecasting, on the other hand, is used by meteorologists to forecast weather patterns and issue warnings about potential natural disasters. For instance, financial analysts use forecasting to predict stock prices and make decisions regarding investments. The quality and quantity of data that is available, as well as the forecasting methods and techniques that are utilized, all have an impact on a forecast's accuracy. Trend analysis, regression analysis, time series analysis, and simulation modeling are some of the various forecasting techniques. In general, forecasting is a potent instrument that can assist both individuals and businesses in making well-informed decisions regarding the future. Forecasting allows us to anticipate potential outcomes and plan accordingly by analyzing historical data and current conditions.

FOOD

One of life's most fundamental necessities is food. It gives us the energy and nutrients our bodies need to grow, develop, and stay healthy. Not only is food necessary for our survival, but it is also central to our culture and social lives.

Whether it's a family get-together, a special occasion, or a meal shared with friends, it brings people together. There are many different things that go into the food we eat, such as animals, plants, and processed foods. Our health, well-being, and quality of life can be affected by the quality of our food and our access to it. Maintaining a healthy weight, lowering the risk of chronic diseases like diabetes and heart disease, and improving one's overall health are all benefits of eating well. Environmental impact from food production and consumption is also significant. Greenhouse gas emissions, water pollution, and land degradation can all be exacerbated by agriculture, livestock production, and food processing. So, it's important to think about how our food choices affect the environment and how sustainable they are. In general, food has an impact on our health, culture, and environment and is an essential part of our lives. In order to improve people's health and the well-being of their communities, it is essential to comprehend the significance of access to high-quality food, sustainable food production, and good nutrition.

MACHINE LEARNING

The development of algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed is machine learning, a subfield of artificial intelligence (AI). To put it another way, machine learning enables computers to learn from experience and grow, much in the same way that humans do. supervised learning, unsupervised learning, and reinforcement learning are just a few of the many methods and algorithms that make up machine learning. A computer algorithm is trained through supervised learning on a labeled dataset, with each data point labeled with an associated output or response. Training an algorithm on an unlabeled dataset for unsupervised learning requires the algorithm to independently identify data patterns and relationships. In reinforcement learning, an algorithm is trained through trial and error and given rewards or penalties for its actions. Predictive analytics, image and speech recognition, natural language processing, recommendation systems, fraud detection, and more all make use of machine learning. It has changed healthcare, finance, and transportation, making it possible to make self-driving cars, improve financial modeling, and make diagnoses that are more accurate.

SALES PREDICTIONNEED INTRODUCTION

Predicting how much a product or service will sell in the future is an important part of business planning. It aids businesses in making well-informed choices regarding marketing, inventory management, production, and financial planning. A company must be able to meet demand, cut costs, and make a profit by accurately predicting sales. Historical sales data analysis, market research, trend analysis, and statistical models are some of the tools used to predict sales. Forecasting sales based on past trends, consumer behavior, market conditions, and other relevant factors is made possible by these methods, which rely on data analysis and statistical techniques. Businesses of all sizes and in all sectors rely heavily on accurate sales prediction. It enables businesses to effectively plan and allocate resources, identify potential obstacles, and make informed decisions. In today's dynamic and constantly changing market, businesses can optimize their operations and remain competitive with accurate sales prediction.

2. LITERATURE REVIEW

THE POWER OF SIMPLICITY: PROCESSING FLUENCY AND THE EFFECTS OF FACTORY CUES ON RETAIL SALES

This paper proposed by Andreas that, despite the fact that ambient scents in retail stores have been shown to influence customers, there are few real-world examples of scent effects and few theoretical explanations for observed effects. The theoretical framework of processing fluency is used in the

current study to address these unanswered questions. The complexity of a scent influences consumer responses to olfactory cues, supporting a processing fluency explanation, according to four studies. In contrast to a scent that was more complex, a scent that was simple (i.e., easier to process) resulted in no change in actual spending or ease of cognitive processing. The results have implications for retail theory and practice. New York University, 2012. All rights reserved, Elsevier Inc. publisher.

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.18

Olfactory cues can clearly influence consumers' perceptions and (sometimes) behaviors in retail settings, as previous research has demonstrated. There has been little, if any, research into the impact of scent on actual behavior and the identification of theoretical underpinnings for observed effects, despite the obvious commercial interest in these findings. Instead of pushing for a more in-depth theoretical explanation, a lot of previous work relied on or assumed the rather simplistic stimulus-organism-response model of environmental psychology.

MACHINE LEARNING BASED FOOD SALES PREDICTION USING RANDOM FOREST REGRESSION

This paper proposed by Hruthvik Naik;Kakumanu Yashwanth;Suraj P;N. Jayapandian in that the Sales forecasting is crucial in the food industry, which experiences high levels of food sales and demand. The industry has concentrated on a well-known and established statistical model. Due to modern technologies, it has gained tremendous appeal in improving market operations and productivity. The main objective is to find the most accurate algorithms to predict food sales and which algorithm is most suitable for sales forecasting. This research work has mentioned and discussed about several research articles that revolve around the techniques usedfor sales prediction as well as finding out the advantages and disadvantages of the said techniques. Various techniques were discussed as to predicting the sales but mainly Incline Increasing Regression and Accidental Forestry Lapse is used for attention. The manufacturing has concentrated on a well-known and established statistical model. Although algorithms like Modest Direct Regression, Incline Increasing Lapse, Provision Course Lapse, Accidental Forest Lapse, Gradient Boosting Regression, and Random Forest Regression are well familiar for outdoing others, it has remained decisively established that Random Forest Regression is the most appropriate technique when associated to the others. After doing the whole examination, the Random Forest Regression technique fared well when compared to other algorithms. The feature importance is generated for the selected dataset using Python and Random Forest Regression and the nose position chart is also explained in detail. The proposed model is compared three major parameters that are accuracy score, mean absolute error and max error. The proposed random forest regression accuracy score is improved nearly 1.83% and absolute error rate is reduced 4.66%.

TREND MOMENT METHOD ON IDENTIFICATION OF FOOD PRODUCT SALES

This paper proposed by Rida Utami; Fitri Pranita Nasution. They state that Shakila Mart is a retail company that has been established since 1989. The system that runs Shakila Mart is still semi-computerized using Microsoft Excel in the process of recording and determining the prediction of the production of all products and product sales in the following year. The production department has problems in making reports of production needs every month, so in the submission of reports to the leadership requires a long time, so that resulting in late ordering of goods that result in a vacancy of stock. The method used in this study is the Trend Moment which is a method that is widelyused in terms of prediction/forecasting. In this study, we are using Shakila Mart sales data as many as 76 items of goods, where the data taken is the total sales of items per month in 2018. The results of this study resulted in a Forecast Sales Forecasting application and had been tested with 76 existing items of data, then the prediction results compared with sales in January 2019, obtained an accuracy of 98.23%.

THE MATHEMATICS OF DECISION TREES, RANDOM FOREST AND FEATURE IMPORTANCE IN SCIKIT-LEARN AND SPARK

This post attempts by X yang and co to consolidate information on tree algorithms and their implementations in Scikit-learn and Spark, as proposed in this paper. It was written, in particular, to explain how feature importance is calculated. This post is not intended to be a substitute for the numerous excellent online resources that discuss the creation of decision trees and random forests. Even though it includes brief definitions for context, it assumes that the reader is familiar with these ideas and wants to know how the algorithms in Scikit-learn and Spark are implemented. To predict the target value, decision trees learn how to divide

the dataset into smaller and smaller subsets. The possible outcomes are represented by the "branches" (edges) and the condition, or test, by the "leaf" (node). This splitting process continues until either a predetermined condition is met, such as the tree's maximum depth, or no more gains can be made. Even without hyperparameter tuning, random forests, a supervised machine learning algorithm, produce excellent results in most regression and classification problems. Because of its simplicity, it is probably the algorithm that is used the most. If the problem is a classification one, it creates a number of decision trees from various samples and votes with the majority. I assume you have already read about Decision Trees; if not, don't worry; we'll start from the beginning. We will learn how to use the Random Forest algorithm, how it works, and the math underlying this straightforward algorithm in this article.

AN OVERVIEW OF STATISTICAL LEARNING THEORY

This paper proposed by Yang and co that statistical learning theory was developed toward the end of the 1960s. It was a purely theoretical investigation of the issue of function estimation from a particular set of data up until the 1990s. Support vector machines—new types of learning algorithms based on the developed theory—were proposed in the middle of the 1990s. As a result, statistical learning theory became a tool for practical algorithms for estimating multidimensional functions in addition to a tool for theoretical analysis. The theoretical and algorithmic aspects of statistical learning theory are covered in this article's broad overview. This overview aims to show how new algorithmic approaches to function estimation problems were inspired by the abstract learning theory's establishment of conditions for generalization that are more general than those discussed in traditional statistical paradigms. A broad overview of statistical learning theory is given in this article. It demonstrates how a general model of generalization can be discovered through abstract analysis. This model asserts that, in addition to the dimensionality of the space or the number of free parameters in the loss function—these concepts serve as the foundation for the traditional paradigm of generalization—capacity concepts are what determine learning machines' capacity to generalize. The new understanding of the mechanisms that underlie generalization not only alters the algorithmic approaches to function estimation problems but also the theoretical foundation of generalization (for instance, from the new perspective, the Occam's razor principle is not always true).

3. EXISTING SYSTEM

Moving averages of various lag or simple regression models are frequently utilized in the existing algorithm's NB and LR basic approaches. In this setting, managers frequently use their intuition and expertise to override baseline predictions. When demand is flat, moving average predictions may be effective. However, the moving average's response is too sluggish when demand follows trends or seasonal patterns. In seasonal peak periods, managers often try to improve performance by prudently increasing stock. Another common strategy is to have a number of reminders that should hint at upcoming holidays, weather, and other demand triggers. However, erroneous predictions and poor decision-making may be caused by human factors like a lack of experience, forgetfulness, or an excessive amount of information. Costs. Moving averages and exponential smoothing are two examples of traditional statistical techniques used in retail businesses' sales forecasting systems. To forecast sales in the future, these strategies make use of historical data. However, for products with intricate sales patterns or businesses experiencing rapid shifts in sales trends, these methods may not be sufficient. Also, these methods don't take into account external factors like changes in consumer behavior or the economy that could affect sales. Forecasting tools like Microsoft Excel that are based on spreadsheets are another system that is already in place. Even though these tools can do basic forecasting, they might not be strong enough for companies with a lot of data or complicated sales patterns. Additionally, it's possible that these tools won't be able to handle large datasets effectively and will require a lot of manual input.

4. PROPOSED SYSTEM

Retail stores can use the proposed system to accurately forecast sales of seasonal furniture

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.18

items. To forecast furniture item sales trends in the future, it makes use of a variety of traditional and cutting-edge time-series forecasting techniques, including SVM with negative and positive binominals. The system's goal is to find the best method for accurate sales forecasting by comparing their results. In the retail industry, where accurate sales predictions are necessary for inventory management and cost control, the system was developed to address the difficulty of forecasting seasonal items. The system is able to identify patterns and trends in the sales of furniture items by analyzing historical sales data, taking into account seasonal fluctuations in demand. The utilization of cutting edge guaging strategies, for example, SVM with Negative and positive binominal gives more exact expectations contrasted with conventional determining techniques. For a variety of businesses involved in manufacturing, logistics, marketing, wholesaling, and retailing, the issue of sales prediction is a significant one. Predicting sales of products with a short shelf life and seasonal shifts in demand are more important to food companies. Sales prediction is typically performed arbitrarily by managers, so we propose a SVM with negative and positive binominals. However, skilled managers are hard to come by and may not be available at all (for example, due to illness or leave). Therefore, computer systems should be used to support sales prediction. These systems should be able to act as a competent manager when the manager is not present and/or assist the manager in making the right decision by providing estimates of future sales. Companies in the manufacturing, wholesale, or retail of products can benefit greatly from the timely and accurate estimation of future sales, also known as sales prediction or sales forecasting. Long-term predictions can assist in business development decision making, while short-term predictions are primarily useful for production planning and stock management.

LOAD SUPERSTORE SALES DATASET:

This module involves loading a retail store's sales dataset, which includes information for three distinct categories from 2014 to the end of 2017: furniture, technology, and office supplies the datasets are provided in the CSV format, and they will be used as input. The data sets are taken from the UCI or Kaggle repository. Collect the Big Mart sales data from the stores and centralize it in a database. By involving this as information we load the informational index for the machine which will prepare and give the outcome utilizing the dataset which contains the subtleties of the huge store deals.

LABEL ENCODING

Label encoding is used in this module to turn the dataset's categorical variables into numerical data. The process of turning labels into numbers so that machines can read them is called label encoding. Machine learning methods can then be used to better understand how these labels work. In supervised learning, it is a crucial pre-processing step for the structured dataset. In machine learning, label encoding is a method for converting categorical data into numerical data so that algorithms can process and analyze it. It involves giving each category in a column or features a unique numerical value. For instance, assume we have a section in a dataset addressing the shades of vehicles: red, blue, and green. Each color would have a numerical value in label encoding, like red = 1, blue = 2, and green = 3. The data can be processed and interpreted by machine learning algorithms thanks to this conversion. It's important to remember that label encoding doesn't mean the categories are in a certain order or have a certain relationship with each other; As unique identifiers, the numerical values are simply assigned. As a result, nominal rather than ordinal categorical data,where the categories are naturally arranged, typically employ label encoding. Using if- else statements or libraries like Scikit-learn, you can manually encode labels. Notwithstanding, it is crucial for note that mark encoding can once in a while prompt issues like giving mathematical qualities an accidental importance or weight, and it could be smarter to utilize one-hot encoding or other encoding methods all things considered.

SPLIT DATASET INTO TRAINING AND TESTING (72-25%):

The dataset is divided into training and testing sets in this module, with 25% of the data going to training. The dataset's training and testing sets were going to be separated. To avoid over fitting, two distinct datasets are not imported for training and testing. As a result, the train and test sets of the same dataset are separated. The

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.18

terms "training dataset" and "testing dataset" are used interchangeably to refer to the datasets used to train and evaluate our model.

SEASONAL ITEMS SALES FORECASTING USING SVM:

For seasonal item sales forecasting, the SVM (Support Vector Machine) method is used in this module. Date, sales volume, and any other pertinent information, such as discounts or promotions offered at that time, should be included in the data. After the SVM model has been trained and tested, it can be used to predict how much the seasonal item will sell for in the future. Businesses can use this information to make better decisions about marketing, sales, and inventory management. For many years statistical methods such as ARIMA and Exponential Smoothing have been used to this aim. However the statistical methods could fail if high irregularity of sales are present, as happens for instance in case of promotions, because they are not well suited to model the nonlinear behaviors of the sales process. In recent years new methods based on machine learning are being employed for forecasting applications. A preliminary investigation indicates that methods based on the support vector machine (SVM) are more promising than other machine learning methods for the case considered. The paper assesses the application of SVM to sales forecasting under promotion impacts, compares SVM with other statistical methods, and tackles two real case studies.

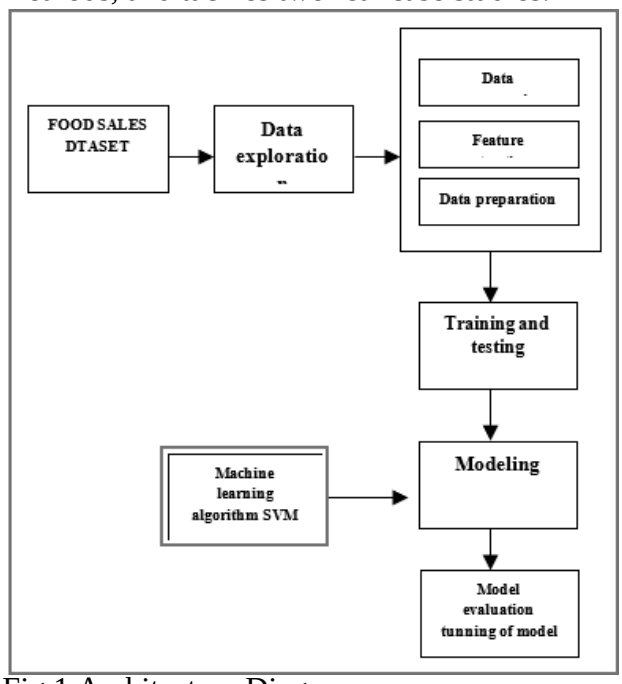

Fig.1 Architecture Diagram

5. RESULT AND DISCUSSION

The accuracy of the classification algorithms is evaluated using a confusion matrix, which provides the number of predictions for each class that can be compared to the occurrences of each class, as well as accuracy within each class.

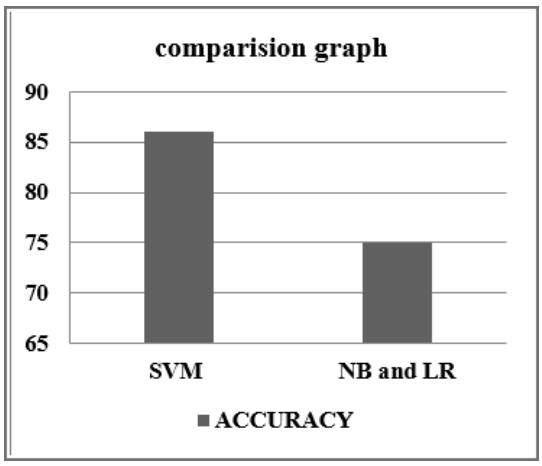

Fig. 2 Graph

The Error Rate is calculated using the Root Mean Square Error, Mean Square Error, and Absolute Error, respectively. Finally, predict food sales using the SVM model and examine the outcomes. To determine the model's accuracy, compare the predicted sales to the actual sales data.

algorithm	ACCURACY
SVM	86
NB and LR	75

Table 1

To gain insight into the patterns and trends in the sales data, you can also visualize the results using graphs and charts. If the GBT implementation is further enhanced with the use of a large data set, the accuracy rate will rise.

6. CONCLUSION & FUTURE SCOPE

On a set of retail sales data, we compare and contrast the various algorithms' results and determine which one is the most effective. The outcomes of comparing various approaches are presented. In addition, we discovered that our model outperforms the competition because it has the lowest RMSE. Data cleaning and

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.18

analysis can be improved, and additional machine learning techniques can be used to improve the accuracy of the model. A larger dataset can lead to predictions that are more accurate. If we want to get better data or a better result, we need to add the kinds of attributes that are already there to the dataset. To get a better result, we need to use a dataset that is well balanced and has different values in each field. To get a better result, we will need a dataset that is well balanced and has different values in each field. The only difference is that the dataset needs to be updated, but the same procedures can be used. Large datasets are suggested for the best results. Based on a performance evaluation, a best-fit prediction model for anticipating trends in sales is provided. The findings discuss the effectiveness and accuracy of the prediction and forecasting methods used. Furthermore, in future more refined prediction can be done based on many other factors like cultural habits, religious holiday, consumer preferences etc. In future, this method can be used for predicting work force requirement, automated food ordering based on forecasting results. Additionally, we want to extend our approach such that it can also predict on hourly basis. Accurate sales predictions on an hourly basis will allow for an optimization of workforce planning.

8. REFERENCES

1. “Sales-forecasting of Retail Stores Using Machine Learning Techniques,” Akshay Krishna, Akhilesh V, Animikh Aich, and Chetana Hegde (2019), IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, IEEE Xplore, 160-166. [CrossRef]

2. Shaniba Ibrahim, Saju Mohanan, and Sunitha Cheriyan; “Intelligent Sales Prediction Using Machine Learning Techniques,” by Susan Treesa, IEEE Xplore, 53- 58, 2019, International Conference on Computing, Electronics, and Communications Engineering (ICCECE).

1. “A Multi-Task Prediction Framework for Sales Prediction,” Chenhui Lu, Shuo Feng, Jiahao Huang, and Xiaojun Ye, International Conference on Computing, Electronics, and Communications Engineering (ICCECE), IEEE Xplore, 2021, pp. 194-19

2. “Sales forecasting by combining clustering and machine- learning techniques for computer retailing,” by I-Fei Chen and Chi-Jie Lu in 2019, is published in Neural Comput & Applications, Vol. 18, pp. 105–112.

3. Machine Learning Model for Sales Forecasting by Using XGBoost, IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE 2021), IEEE, 480-483 [CrossRef], Xie dairu1, and Zhang Shilong1 (2021).

4. Lasek, A.; Cercone, N.; Saunders, J. Restaurant Sales and Customer Demand Forecasting: Literature Survey and Categorization of Methods, in Smart City 360°; Springer International Publishing: Cham, Switzerland, 2016; pp. 479–491.

5. Green, Y.N.J.; Weaver, P.A. Approaches, techniques, and information technology systems in the restaurants and foodservice industry: A qualitative study in sales forecasting. Int. J. Hosp. Tour. Adm. 2008, 9, 164–191.

6. Holmberg, M.; Halldén, P. Machine Learning for Restauraunt Sales Forecast; Department of Information Technology, UPPSALA University: Uppsala, Sweden, 2018; p. 80.

7. Rao, G.S.; Shastry, K.A.; Sathyashree, S.R.; Sahu, S. Machine Learning based Restaurant Revenue Prediction. Lect. Notes Data Eng. Commun. Technol. 2021, 53, 363–371.

8. Schmidt, A. Machine Learning based Restaurant Sales Forecasting. In Computer Science; University of New Orleans: New Orleans, LA, USA, 2021.

9. S.N Gunjal, S K Yadav, D B Kshirsagar, “A Distributed Item Based Similarity Approach For Collaborative Filtering On Hadoop Framework”, Advances in Parallel Computing (IOS Press),March 2020.pp.407-415.[CrossRef]

10. S. N. Gunjal, S. K. Yadav and D. B. Kshirsagar, “A hybrid scalable collaborative filteringbased recommendation system using ontology and incremental SVD algorithm”; 2020 International Conference on Smart Innovations in Design, Environment, Management,Planning and Computing (ICSIDEMPC), Aurangabad, India, 2020, pp. 39-45. [CrossRef]

11. A. S. Weigend and N. A. Gershenfeld, “Time series prediction: Forecasting the future and understanding thepast”, Addison- Wesley, 2019

12. Zhang B, Tan R, Lin C J. Forecasting of e- commerce transaction volume using a hybrid of extreme learning machine and improved mothflame optimization algorithm[J]. Applied Intelligence, 2020: 1-14.

13. Dovžan, D., Logar, V., & Škrjanc, I. (2019, June). Solving the salesprediction problem with fuzzy evolving methods. In 2012 IEEEInternational Conference on Fuzzy Systems (pp. 1-8). IEEE. [CrossRef]