An Efficient Novel Approach on Machine Learning Paradigmsfor Food Delivery Company through Demand Forecastıng in societal community

Subbarayudu Yerragudipadu [1*], *Vijendar Reddy* Gurram[1] ,*Navya Sri* Rayapudi[1], *Bhavana* Bingi[4], *Likhitha* Gollapalli[1] *and Ukritha* peddapatlolla[1]

[1]Department of Information and Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

Abstract. A food delivery business must be able to accurately forecast demand on a daily and weekly basis since it deals with a lot of perishable raw components. A warehouse that keeps too much inventory runs the danger of wasting items, whereas a warehouse that maintains too little inventory runs the risk of running out of stock, which might lead consumers to switch to your competitors. Planning for purchasing is essential because most raw materials are perishable and delivered on a weekly basis. For this issue to be resolved, demand forecasting is crucial. With the aid of historical data-driven predictive research, demand forecasting determines and forecasts future consumer demand for a good or service. By predicting future sales and revenues, demand forecasting assists the organisation in making more educated supply decisions. Regression methods like linear regression, decision trees, and Xgboost are used to overcome this issue.

1 Introduction

The success of a restaurant now not solely relies upon on taste, atmosphere however also on service. The most essential phase amongst the offerings is serving sparkling food. In order to furnish this, the eating places want to put together meals daily, this requires shopping for some of clean self existence meals merchandise each day. The fundamental venture that one would face in this will be predicting the volume of merchandise to be offered and prepared. It is very challenging to predict the wide variety of orders in a given restaurant on a given day. An incorrect prediction might also give up buying and making ready muchless quantity of meals which will motive scarcity or buying and making ready greater which will lead to wastage of food. So, predicting the genuine demand is an

* Corresponding Author: subbu.griet@gmail.com

assignment due to the fact of uncertainty and fluctuations in purchaser demand. These editions advert fluctuations in demand might also be due to the fact of rate change, promotions, exchange in customer's preferences and climate changes. All these elements suggest that some dishes are bought usually at some point of restricted length of time. Although we comprehend that some normal seasonal sample is expected, the aspects that predict these seasons are nolonger immediately observed. Thus, drops and rises in orders due to the fact of these seasonal adjustments are challenging to predict. This study focuses on identifying and solving an issue that involves forecasting the quantity of orders for food based on previously provided customer data.

2 Literature Survey

Wasted food is described as food that has been partially consumed or that a retailer has discarded due to its colour or appearance. Certain deliveries to the store are never sold due to damage, the label's expiration date, or how long the item has been on the shelf. Numerous traditional data science forecasting techniques, such as multiple regression, exponential smoothing, ARIMA, supervised regression and classification models, random forest, gradient boosting, and stochastic optimisation, are often used in the forecasting of food demand[1].

One of the primary issues in supply chains, demand forecasting, is improved in this research through the use of a novel methodology. To do this, a support vector regression algorithm (SVR), a demand forecasting model based on the DL technique, and nine different time series approaches are created. Using a novel integration strategy that is comparable to the boosting ensemble strategy, nine various time series approaches, the SVR algorithm, and the DL model are merged to produce the final decision of these models for the proposed system.The other novel feature of this work is the use of the boosting method to the demand forecasting model. [2].

For developing and overseeing supply chain management plans. The chance of faulty demand estimates is reduced when demand forecasting is supported with accurate initial data, preventing mistakes like the "bullwhip effect" brought on by information asymmetry. Systematic demand forecasting has previously used multiple linear regression models and ARMA (Autoregressive Moving Average) statistical models for supply chain demand prediction.

The forecasts produced by these linear models have the flaw of being unreliable. Greater precision is attainable thanks to the amount of data and advanced computation. [3]. In this study, we investigate whether long-short-term memory neural networks (LSTMs) are appropriate for demand forecasting in the e-grocery retail sector. Univariate and multivariate LSTM-based models were developed and assessed for 100 fast-moving consumer goods as part of a master's thesis. In terms of results for food products, the developed models performed better than comparable models from the statistical and machine learning families.

Only in the beverage industry did linear regression and random forest generate somewhat more favourable results. This finding suggests that LSTMs can be applied to forecast product-level demand[4]. Most research studies either use advanced marketing strategies or concentrate on a specific kind of product. This paper suggests the application of machine learning techniques. Information about consumer interest in novel products was acquired from the Ozon online store. The price, name, category, and text description of the product are only a few of the characteristics that make up the algorithm's input data. Different gradient boosting method implementations, like as XGBoost, Light GBM, and Cat Boost, were used to tackle the regression problem.

The accuracy of the forecast is 4. The suggested system can be used independently or in conjunction with a more intricate system. [5]. In the first half of this essay, the factors that significantly influence food costs have been identified by a review of the pertinent literature. Then, historical information on the variables was acquired. Multivariate time series data make up these data. The proposed model is then provided in the next stage using a combination of optimised neural networks. The suggested model, denoted as CNN-LSTM-GA, combines genetic algorithms, convolutional neural networks (CNN), and long short-term memories (LSTM) (GA). This network is capable of deriving complicated features from various variables.

The LSTM layer is best for modelling time information from erratic trends in time series components, whereas the CNN layer may extract features between different variables affecting food prices. [6]. In spite of the fact that the antiquated ARIMA method has been extensively investigated for restaurant analysis, in this study we want to widen the scope of the current research by analysing a range of machine learning (ML) algorithms on a curated, real-world dataset. We employ a strategy that is both simple and thorough to compare a wide range of models.

When executing a prediction assignment, an ideal ML model will capture minute information like vacations while maintaining performance when the forecast window is increased from one day to one week. It will receive training with the ideal quantity of features. We describe how a set of forecasting models with low performance can be transformed from a raw dataset [7].By including data on factors relating to product features, advertising, the environment, local economies, and internet social media, this study addresses the problem of demand forecasting.

To manage the massive volume of data, an Apache Spark-based big data framework is developed. This framework allows for efficient data processing, modelling, and analysis in order to generate projections. There are three types of time-series data-based demand forecasting methods: statistical techniques, machine learning techniques, and hybrid techniques. Time-series forecasting methods use the auto-regressive integrated moving average (ARIMA) and the exponential smoothing approach as decomposition models. Numerous multivariate time-series techniques, including ARIMAX, have also been used. The specifics of these techniques can be found in Hyndman & Athansopoulos [8].

The objective is to improve demand forecasting, which will improve the system for production planning used in Ecuador's textile sector. These businesses make it difficult to provide a reliable projection of future demand because of recent changes in the Ecuadorian backdrop. The demand for textile products has been influenced by a number of factors, including sales pricing, manufacturing costs, manufacturing gross domestic product, and the unemployment rate. These elements, which also result in uncertainty scenarios, have a significant impact on the forecast's accuracy and quality.To compare a range of conventional techniques, such as ARIMA, STL Decomposition, Holt-Winters, artificial neural networks, Bayesian networks, random forests, and support vector machines, is the aim of this study[9].

On a chosen test dataset, we assess the models' propensity to forecast time intervals of one day and one week. The greatest outcomes in one-day forecasting come from linear models with a sMAPE of only 19.6%. With errors around 20%, ensemble models and the two RNN models LSTM and TFT also did well. When forecasting for one week, non-RNN models performed poorly, with forecasts that were more than 20% off. RNN models extended more successfully, producing a high result of 19.5% and excellent sMAPE scores. RNN models typically performed badly on datasets with trend and seasonality removed, although many simpler ML models performed well when linearly separating each training instance[10].

Demand forecasting may have a substantial impact on a company's ability to control expenses, increase productivity, and remain competitive. However, accurate demand forecasting is essential for this to occur[11]. On this issue, the current study uses a framework and machine learning techniques to investigate time series data pertaining to the demand for a certain style of luxury handbag for women.

The Group Model of Data Handling (GMDH), Discrete Wavelet Transform-Neural Networks (DWTNN), Radial Basis Function Neural Network (RBFNN), and Adaptive Neuro-Fuzzy Inference System (ANFIS) were used as the five machine learning models for this[12].

This study's goal is to evaluate several Machine Learning Algorithms (MLAs) for estimating consumer demand for in-flight meals. Four MLAs—a Multilayer Perceptron Neural Network, Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and Linear Regression (LR)-were selected after a review of previous articles (MLP)[13].

The paper explores the best MLA for the problem at hand and what factors are most crucial for forecasting consumer demand for in-flight meals. The key areas of interest are finding appropriate MLAs and analysing, contrasting, and modifying the characteristics of the MLAs to better optimise the selected models[14] .

3 System Architecture

The distribution of valuable correspondences is what a system architecture design is all about. These are formal pieces that encapsulate standards and data. Between elements, among characteristics, and among the surrounding elements, architecture specifies the members of the family. Creating an architectural sketch is no longer simple.

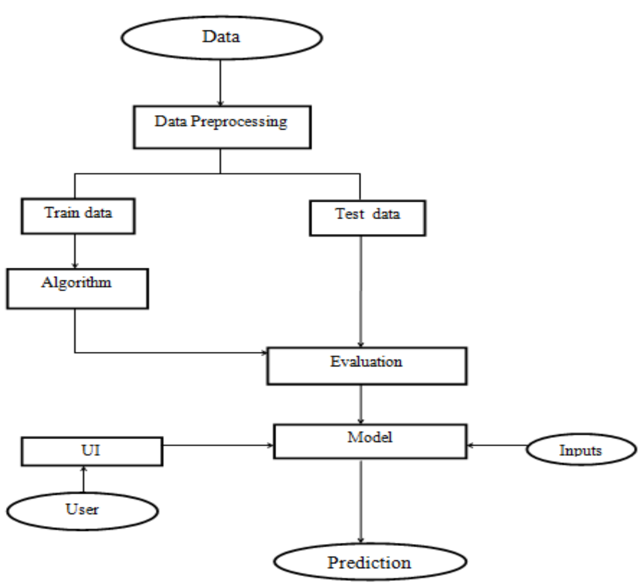

Fig. 1. System Architecture

4 Methodologies

4.1 Dataset

Four csv files that contain test data, train data, and other pertinent information make up our initial collection of data.Information such as ID, week, centre, meal, checkout price, base pricing, emailer for promotion, homepage featured, and number of purchases are among the data in the train.the CSV file. During the testing of training.data, this file is utilised.The csv file contains the following columns: ID, week, centre, food, checkout, base, emailer for promotion, and highlighted webpage. Meal_info.csv: This file, which is used for testing, provides information on each meal that is provided. The file fulfillment_center_info.csv contains details about each fulfilment centre

4.2 Algorithm

4.2.1. Xgboost Regression

Decision trees are used in the gradient boosting technique known as XGBoost. In comparison to other algorithms or frameworks, artificial neural networks performed better in prediction tests utilising unstructured input (images, text, etc.). On the other hand, decision tree-based algorithms now maintain the top spot for small to medium-sized structured/tabular data. The use of XGBoost facilitates execution speed and model execution.The XGBoost programme is used to create gradient boosting decision tree approaches. When we have a lot of observations, it may be utilised. Additionally, when the data contains both category and numerical information.

4.2.2 Decision Tree

Decision Tree Regressor can be used for two purposes that is for Classification and Regression. It chops down a dataset into smaller and smaller chunks in the background while building a related selection tree.The obtained final tree consists of leaf nodes and choices. A selection node can contain two or more branches, each representing a value for the property being evaluated. A leaf node represents a numerical goal selection. Decision bushes can handle both explicit and numerical inputs.

4.2.3 Linear Regression

One value can be predicted based on the other value, in general this is called linear regression. The attempt is to predict the dependent variable. Input variables are the input factors that are used to predict the other value. By creating a straight line, linear regression reduces the differences between expected and actual output values. For paired data, Simple Linear Regression use the "least squares" method to find the best-fit line.

4.2.4 Method

The roc curve that is receiver operating characteristic curve is a graph that represents how good a classification model performs at every level of categorization (ROC curve). Two valuess are shown on this curve.

1.True Positive Rate

2.False Positive Rate
Recall is also known as true positive rate (Tpr), which can be calculated as:

$$\boldsymbol{Tpr=Div(Tp,Tp+Fn)} \qquad (1)$$

False Positive Rate (Fpr) can be calculated as:

$$\boldsymbol{Fpr=Div(Fp,Fp+Tn)} \qquad (2)$$

Workflow and Algorithm
This case study will help the restaurants to make predictions regarding number of orders for next ten weeks. To put into practise a system with a web-based graphic user interface and a machine learning algorithm, the following will be done.
1. Firstly, the user gives input through a website which is a user interface.
2. The input comprises of various fields such as Cuisine, Category, Region ID,Area etc. Then this data is pre-processed into machine understandable language. Some of the pre-processing techniques which can convert the input data into machine understandable language are encoding, and label encoder.
3. Visualize and analyse the data in order to make good decisions.
4. Split the dataset into Training and Test set.
5. Then model is trained by using various algorithms like XGBoost, Linear Regressor and Random Forest Regressor upon this pre-processed data.
6. After this the model undergoes testing upon the test data. By using test data we can obtain the result.
7. The results of the prediction will be made available to the food delivery companies.

5 Experimental Results And Discussions

When applied different regression algorithms like Decission Tree Regressor, Linear Regression and XGBoost Regressor , XGBoost Regressor achieved greater accuracy of more than 80%.1.When a given data collection is presented in a graphical format, it is called data visualisation. It aids in the discovery of trends, correlatioons, and patterns that can go unnoticed in text-based data.2. Understanding your data and the relaationships within it is just as crucial to machine learninng model training as any algorithm. In fact, poorly displayed and comprehended data will cause even the most advanced machine learning algorithms to underperform.

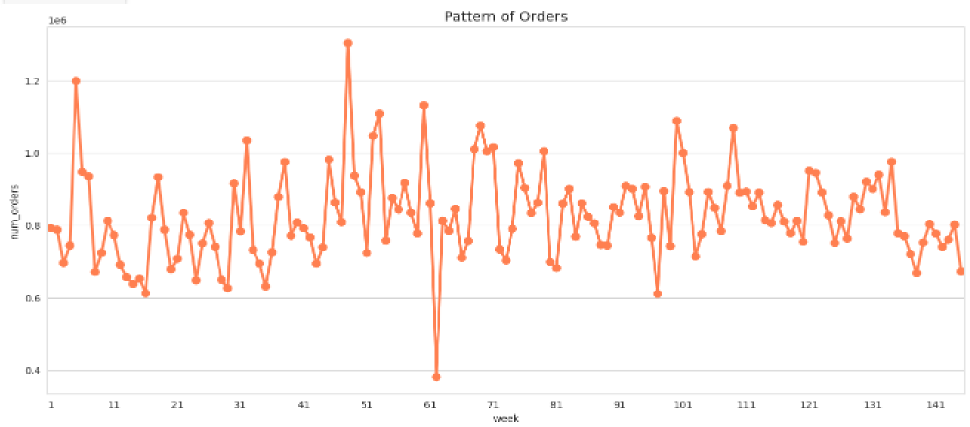

Fig. 2. When we looked at the overall pattern of orders placed across the weeks, we could see that week 48 had the most orders and week 62 had the fewest.

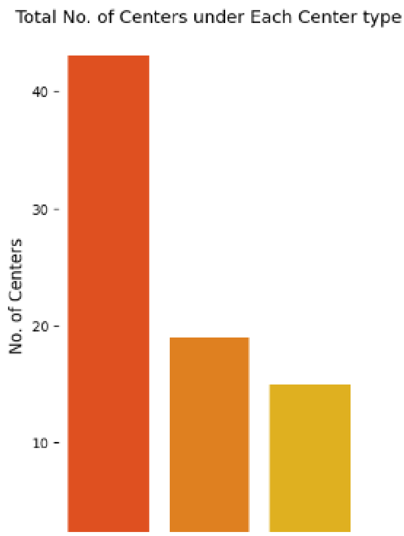

Fig. 3. Type_A centres put the most orders, whereas Type_C centres place the fewest orders.

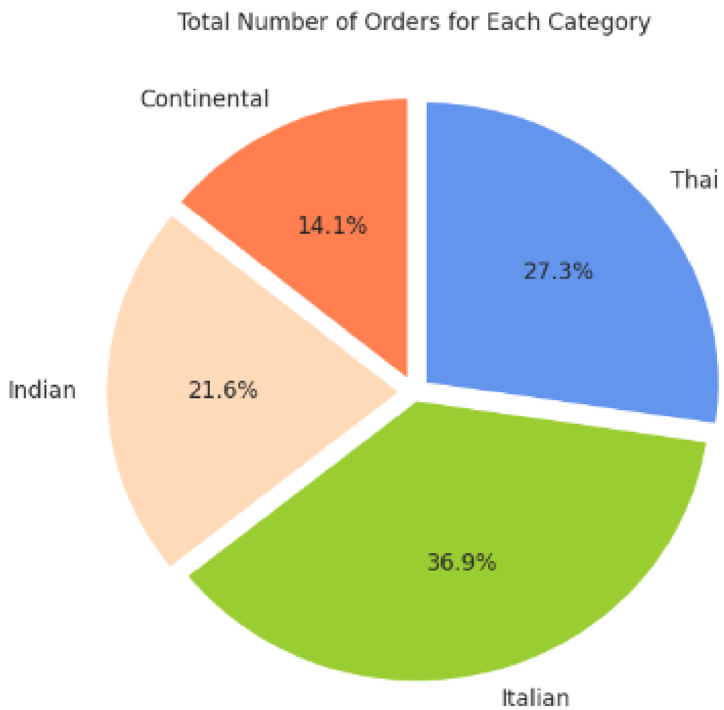

Fig. 4. The most orders were for Italian food, while the least were for continental cuisine.

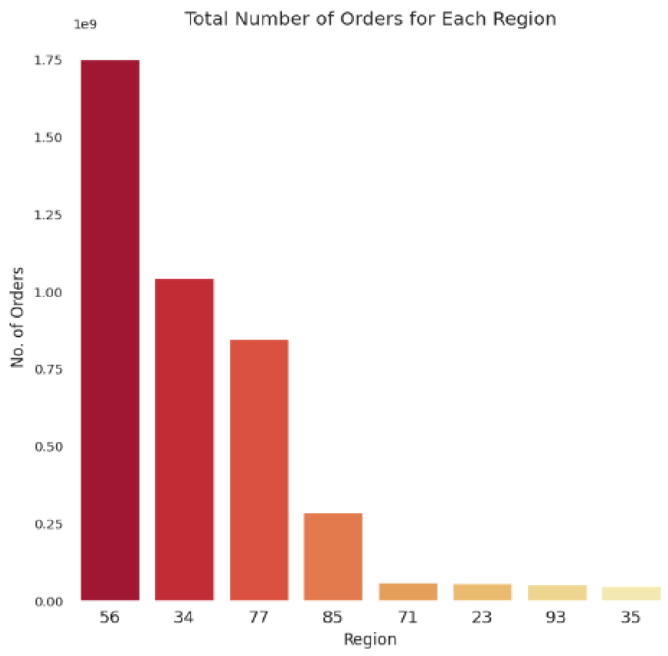

Fig. 5. When we looked at the amount of orders by region, we could see that Region 56 acquired the most orders, 60.5M orders, which is about 35M orders more than the second-highest region, Region 34, with 24M orders.

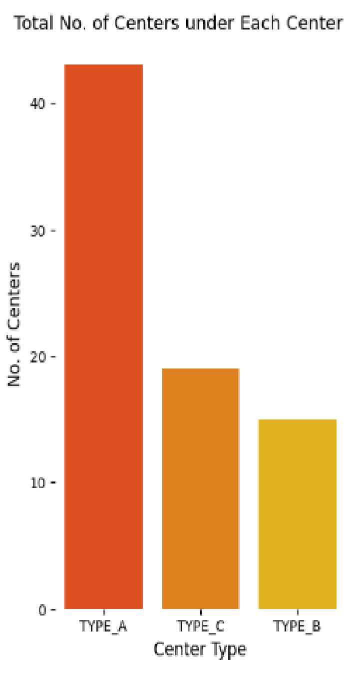

Fig. 6. The reason TYPE_A gets the most orders is because it has the most centers—43 centres.

Fig. 7. In an effort to ascertain whether there is any correlation between the volume of orders and the discount, or the difference between the base price and checkout price, a new feature called Discount was created. Unexpectedly, there is no connection between the discount and the number of orders.

Table 1. Results

Method	Accuracy	Rmse Value
Xgboost Regressor	80.1	51.1
Decision Tree Regressor	74.8	68.2
Linear Regression	70.3	62.6

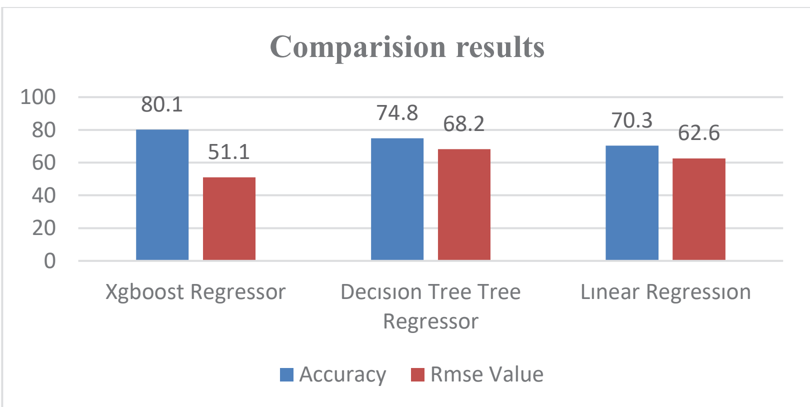

Fig 8. Comparision of XgBoost Regressor, Decision Tree Regressor and Linear Regression

8 Conclusion And Future Enhancements

Food is one of a person's most fundamental needs, hence data science is highly demanded and required in the food industry. Food producers and delivery services have a challenge and a clear path to profitability by reducing food waste. The production of raw materials is frequently influenced by price, discount, location, and a wide range of other factors. It is possible to anticipate the volume of orders for a particular week using machine learning. The present study illustrated a potential use for forecasting output order numbers based on meteorological input parameters. Any meal delivery service may utilise the created webpage, which is user-friendly and has a prediction accuracy of more than 80%, showing superior forecast accuracy. Anyone can utilise this user-friendly website by providing the necessary information. As a food demand approach in this research, we employed ensemble techniques such XGBOOST, Decision Tree Regressor, and Linear Regression strategy. The accuracy rate continues increasing as we use a proprietary system for prediction. A decision-based boosting method that raises accuracy rates is the XGBoost algorithm. The dataset also includes internal and external data for the forecast, including information about the week, the region, and other things. When taking into account a variety of additional factors, customer reach, feedback, cultural practises, religious holidays, consumer preferences, as well as the use of neural networks, would all contribute to a deeper understanding and a greater degree of accuracy in stock predictions and customer reach. By automating meal ordering and predicting staff needs based on forecasting results, this method can be used in the future. The strategy will encourage the business to provide meals as effectively as feasible.

References

1. K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk and A. Kozina, "Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network," in IEEE Access, vol. **9**, pp. 146123-146136, (2021)
2. Kilimci, Zeynep Hilal, A. Okay Akyuz, Mitat Uysal, Selim Akyokus, M. Ozan Uysal, Berna Atak Bulbul, and Mehmet Ali Ekmis. “An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain.” Complexity (2019)
3. Saha, Priyam & Gudheniya, Nitesh & Mitra, Rony & Das, Dyutimoy & Narayana, Sushmita & Tiwari, Manoj”Demand Forecasting of a Multinational Retail Company using Deep Learning Frameworks". IFAC-PapersOnLine. **55**. 395-399. 10.1016/j.ifacol.2022.09.425.
4. Golabek, Marta & Senge, Robin & Neumann, Rainer “Demand Forecasting using Long Short-Term Memory Neural Networks”, (2020)..
5. P S Smirnov[1] and V A Sudakov[1,]2, “Forecasting new product demand using machine learning” ,(2021)
6. Mobina Mousapour Mamoudan, Zahra Mohammadnazari, Ali Ostadi & Ali Esfahbodi “Food products pricing theory with application of machine learning and game theory approach”, International Journal of Production Research(2022)
7. P. Meulstee and M. Pechenizkiy, "Food Sales Prediction: "If Only It Knew What We Know",IEEE International Conference on Data Mining Workshops, Pisa, Italy, 2008, pp. 134-143, (2008)
8. Sushil Punia, Sonali Shankar,Predictive analytics for demand forecasting: A deep learning-based decision support system,Knowledge-Based Systems,Volume **258**,109956,ISSN 0950-7051, (2022).

9. Lorente-Leyva, LL.; Alemany Díaz, MDM.; Peluffo-Ordóñez, DH.; Herrera-Granda, ID”A Comparison of Machine Learning and Classical Demand Forecasting Methods: A Case Study of Ecuadorian Textile Industry”. Lecture Notes in Computer Science. 131-142, (2021).
10. Schmidt, A.; Kabir, M.W.U.; Hoque, M.T. Machine Learning Based Restaurant Sales Forecasting. Mach. Learn. Knowl. *Extr. 4*, 105-130,(2022).
11. Akbar Abbaspour Ghadim Bonab. "A comparative study of demand forecasting based on machine learning methods with time series approach*"*. Journal of Applied Research on Industrial Engineering*,* 9, 3, 331-353,(2022).
12. Hast, Matteus. “Evaluation of machine learning algorithms for customer demand prediction of in-flight meals.” (2019).
13. Yerragudipadu subbarayudu , alladi Sureshbabu “Distributed Multimodal Aspective on Topic Model Using Sentiment Analysis for Recognition of Public Health Surveillance” Expert Clouds and Applications, Singapore Print ISBN 978-981-16-2125-3 Online ISBN 978-981-16-2126-0(2021)
14. Syed Umar, Yerragudipadu Subbarayudu,K. Kiran Kumar, N. Bashwanth, “Designing of Dynamic Re-clustering Leach Protocol for Calculating Total Residual Time and Performance”,International Journal of Electrical and Computer Engineering (IJECE)Vol.**7**, No.3, pp. 1286~1292ISSN: 2088-8708, (2017)