Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 25,January 2023

A Machine Learning System for Demand Forecasting using Light Gradient Boosting MachineAlgorithm

Srinithi.K , Gomathi.K , Sujitha.T
Computer Science and Engineering
Saranathan College of Engineering
Trichy,India

***Abstract*—Improved sales forecasts for individual products in grocery stores can have a beneficial effect both environmentally and economically. Historically these forecasts have been done through a combination of statistical measurements and experience. However, with the increased computational power available in modern computers, there has been an interest in applying machine learning to this problem. Inventory forecasting aims to predict the demand for a particular product in the future and reserve the number of products based on the forecasting results. An accurate and reliable inventory prediction can avoid product overstocking and greatly reduce maintenance costs. To produce more precise predictions and analysis, we use the Gradient Boosting algorithm which is an outstanding method for its prediction speed and accuracy, peculiarly with large and complex datasets. For all machine learning solutions for business, this algorithm has produced the best results. LightGBM is one of those which is a framework that uses a tree-based learning algorithm. It grows leaf-wise while another algorithm grows level-wise which reduces more loss. This method had the lowest MAE (Mean Absolute Error) deviance, RMSE (Root Mean Squared Error) deviance, and MAPE (Mean Absolute Percentage Error). This concludes that the Gradient Boosting model performed better than existing algorithms like S-ARIMA, exponential smoothing, and neural networks.**

***Keywords*—** Gradient boosting; Light GBM; S-ARIMA;

I. INTRODUCTION

Inventory Demand Forecasting is to predict the demands of a grocery store by using a suitable Machine learning algorithm. A dataset of a grocery store with previous months' sales shall be taken for analysis, feature engineering, development of the model, and evaluation performance.[1] The dataset obtained should be used in such a way that the project can predict sales inthe next few months, according to which the products can be purchased or manufactured in the future. Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time. Time-series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time-series data can be used for forecasting—predicting future data based on historical data. Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms.

Models of time series analysis include:

- **Classification:** Identifies and assigns categories to the data.
- **Curve fitting:** Plots the data along a curve to studying the relationships of variables within the data.
- **Descriptive analysis:** Identifies patterns in time series data, like trends, cycles, or seasonal variation.
- **Explanative analysis:** Attempts to understand the data and the relationships within it, as well as cause and effect.
- **Exploratory analysis:** Highlights the main characteristics of the time series data, usually in a visual format.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 25,January 2023

- **Forecasting:** Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along with future plot points.
- **Intervention analysis:** Studies how an event can change the data.
- **Segmentation:** Splits the data into segments to show the underlying properties of the source information.

The forecasting execution process consists of data preparation, diagnosis, model selection, fit or forecast, and automated forecast quality evaluation. The user can define relevant performance metrics to evaluate the generated forecast.

II. Related works

The concept of inventory management includes a vast set of areas to be dealt with. With the rapid expansion of the e-commerce industry, the demand for efficient inventory management is the need of the hour. More research effort is required to improvise the existing inventory management techniques. E-commerce giants define an end-to-end machine learning system using probabilistic demand forecasting models that are built on Apache Spark. Such e-commerce giants contain large datasets.

2.1 Algorithm Used In Previous Work:

To forecast sales, several types of algorithms have been proposed with neural networks and auto-regression being the most prominent. This is expected as the problem was, in essence, a time series problem. ARIMA and Long Short-Term Memory (LSTM) have yielded much discussion and promising results in the academic literature, However, regression models such as Lasso, support vector regression (SVR), and Random Forest have shown promising results as well, indicating that this approach could yield prominent results. To build the model, the regression analysis and neural network were used as the prediction models. The learning rate in this model is set to be adaptive by using the Adadelta method. The 5-fold cross-validation was utilized to test the validity and performance of the result. The performance indexes are conventional prediction error measures such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) between the predicted sales amount and actual sales amount. All simulations are conducted in the R programming language.

2.2Autoregressive Integrated Moving Average:

This model has many uses in many industries. It is widely used in Demand Forecasting, such as in determining future demand in food manufacturing. That is because the model provides managers with reliable guidelines in making decisions related to supply chains. ARIMA models can also be used to predict the future price of your stocks based on past prices.

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

2.3Artificial Neural Network:

Artificial Neural Networks (ANNs) are one of the popular machine learning methods for information processing and pattern identification. ANNs are inspired by biological neural systems, especially the human brain system, and have been used in many areas. The objective of this method is to develop mathematical algorithms that will enable ANNs to learn by mimicking information processing and knowledge transferring of the human brain. An ANN is a network of many simple computing elements called neurons which are highly interconnected and organized in layers. ANNs have been successfully implemented in many fields especially in business and economics because of their ability to solve many complex real-world problems. For instance, ANNs can be used to predict future trends based on large historical data. Among different types of ANNs, one of the most used in forecasting is FFNN, with a backpropagation algorithm, called Back Propagation Neural Network (BPNN).

2.4Support Vector Regression:

It tries to find a line or hyperplane which is in multidimensional space that separates these two classes. Then it classifies the new point depending on whether it lies on the positive or negative side of the hyperplanedepends on the classes to predict.

2.5Lasso Regression:

The word “LASSO” stands for Least Absolute Shrinkage and Selection Operator. It is the statistical formula for the regularisation of data models and feature selection. Lasso regression is a regularisation 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. This particular type of regression is well-

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 1, Paper 25,January 2023

suited for models showing high levels of multi-collinearity or when you want to automate certain parts of model selection, like variable selection or parameter elimination.

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

2.6 Random Forest:

Random Forest can also be used for time series forecasting, although it requires that the time series dataset be transformed into a supervised learning problem first. It also requires the use of a specialized technique for evaluating the model called walk-forward validation, as evaluating the model using k-fold cross validation would result in optimistically biased results.

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \qquad \sigma = \sqrt{\frac{\sum_{b=1}^{B}(f_b(x') - \hat{f})^2}{B-1}}.$$

III. PROPOSED SYSTEM

Machine learning not only increases the accuracy of demand forecasts, but it also automates large amounts of planner work and can process enormous data sets far more than any human planner would be capable of. To generate an accurate demand forecast, a system must be able to process an enormous amount of data on the wide range of variables that can potentially impact demand.With advancements in large-scale data processing and in-memory computing, modern demand planning systems can make millions of forecast calculations within a minute, taking into consideration more variables than ever before possible.Consider the three broad areas of variability that continuously impact demand: recurring variations in baseline demand patterns, your own internal business decisions, and external factors such as weather or local events.

To predict the impact of business decisions, you must leverage machine learning algorithms that can process large amounts of retail data and integrate them into the baseline demand forecast to be accounted for.

3.1 Problem Definition:

To predict the sales of multiple grocery stores from historical data and forecast the future demands using the Gradient Boosting algorithm.

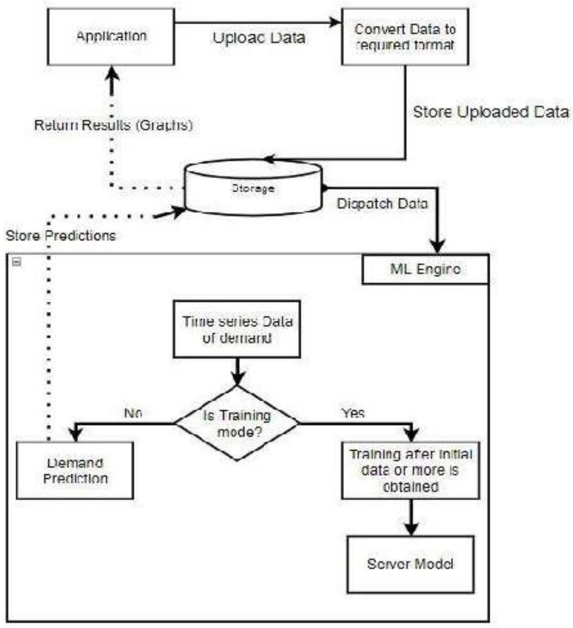
3.2 Methodologies of Problem Solving:

1. Analysis of the chosen data which consists of three-year sales data.
2. Applying a machine-learning algorithm to build a model capable of predicting demands and evaluating the accuracy of the model built.
3. Allowing the user to choose the suitable model according to their requirements.

3.3 System Architecture:

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 25,January 2023



3.4 Objectives of Demand Forecasting:
Demand forecasting is an important concept required for making important business decisions.
a) Formulating Production Policy
b) Formulating Price Policy
c) Controlling Sales
d) Arranging Finance
e) Deciding the Production Capacity
3.5 Factors Affecting Demand Forecasting:
Demand forecasting helps in determining the place, time, and quantities of particular products. Several factors affect demand forecasting.
a) Types of Goods
b) Competition Level
c) Price of Goods
d) Level of Technology
e) Economic Viewpoint
3.6 Process of Demand Forecasting:
A systematic and scientific approach can lead us to an effective, fruitful, and required Demand Forecasting process for an organization.
a) Setting the Objective
b) Determining Period
c) Selecting a Method for Demand Forecasting
d) Collecting Data.
3.7 Proposed Methodologies:
3.7.1 Boosting:
In machine learning, boosting is an ensemble meta-algorithm for primarily reducing bias and variance in supervised learning. The principle behind the boosting algorithm is first we built a model on the training dataset, then the second model is built to rectify the errors present in the first model.
3.7.2 Gradient Boosting Algorithm:

This algorithm is mainly used to build models sequentially where these subsequent models try to reduce the errors of the previously built models. This can be done by building the new model on the errors of the previous model. When the target column is continuous, we use a Gradient boosting regressor.

3.7.3 Light Gradient Boosting Machine:

LightGBM is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm.LightGBM extends the gradient boosting algorithm by adding a type of automatic feature selection as well as focusing on boosting examples with larger gradients. This can result in a dramatic speedup of training and improved predictive performance.LightGBM has become an algorithm for machine learning competitions when working with tabular data for regression and classification predictive modeling tasks. As such, it owns a share of the blame for the increased popularity and wider adoption of gradient boosting methods in general, along with Extreme Gradient Boosting.LightGBM, short for Light Gradient Boosting Machine, is a free and open-source distributed gradient boosting framework for machine learning originally developed by Microsoft. Its initial release was in the year 2016 since then it has been proved to be a successful model for forecasting in data science competitions and recent literature.LightGBM has many boost advantages, including sparse optimization, parallel training, multiple loss functions, regularization, bagging, and early stopping. It has been actively used in online communities with promising results. Gradient boosting refers to a class of ensemble machine learning algorithms that can be used for classification or regression predictive modeling problems.

Ensembles are constructed from decision tree models. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. This is a type of ensemble machine learning model referred to as boosting.

Models are fit using any arbitrary differentiable loss function and gradient descent optimization algorithm. This gives the technique its name, “gradient boosting” as the loss gradient is minimized as the model is fit, much like a neural network.Light Gradient Boosted Machine, or LightGBM for short is an open-source implementation of gradient boosting designed to be efficient and perhaps more effective than other implementations. LGBM is based on decision tree algorithms and is used for ranking, classifications, and other machine learning tasks. The development focus is on performance and scalability.

3.7.4 Advantages of Proposed Methodologies:

Faster training speed and higher efficiency: Light GBM usesa histogram-based algorithm. It buckets continuous feature values into discrete bins which fasten the training procedure.

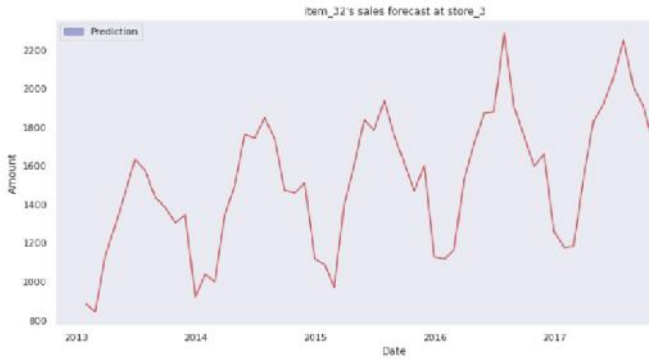
Lower memory usage: Replaces continuous values with discrete bins which results in lower memory usage.Better accuracy than any other boosting algorithm: It produces much more complex trees by following leaf wise split approach rather than a level-wise approach which is the main factor in achieving higher accuracy. However, it can sometimes lead to fitting which can be avoided by setting the max_depthparameter. Compatibility with Large Datasets: It is capable of performing equally well with large datasets with a significant reduction in training time as compared to other algorithms. Support of Parallel, Distributed, and GPU learning

4. Results and Discussion

To evaluate and compare the different models fairly, the choice of evaluation metrics was important as each metric has different characteristics. It was also important to include several metrics since different metrics could display different flaws or benefits in the models. Root mean squared error (RMSE), mean absolute error (MAE), and Mean Absolute Percentage Error (MAPE) have been used extensively in the academic literature and could, therefore, be deemed to be the most useful. In comparison with these, when analyzing the online communities, data science competitions, and sources outside the academic literature, it was clear that Symmetric Mean Absolute Percentage Error (SMAPE) can be beneficial when comparing the models. By utilizing multiple performance metrics, with different characteristics, as specified above, the chances of locating the best algorithm for a specific outcome are increased.

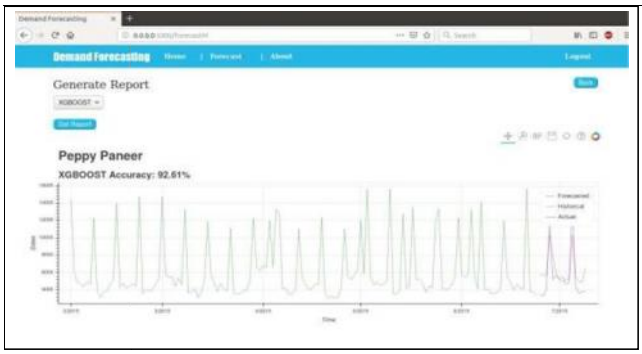
Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 25,January 2023



- LightGBM is considered to be a really fast algorithm and the most used algorithm in machine learning when it comes to getting fast and high accuracy results. Parameter Tuning is an important part that is usually done to achieve good accuracy, fast result, and to deal with overfitting.Use small max_bin
- Use small num_leaves
- Usemin_data_in_leafandmin_sum_hessian_in_leaf
- Use bagging by set bagging_fraction and bagging_freq
- Use feature sub-sampling by set feature_fraction
- Use bigger training data
- Try lambda_l1, lambda_l2 and min_gain_to_split for regularization
- Try max_depth to avoid growing deep tree

4.2 Accuracy of XGBoost Algorithm



Volume 6- Issue 1, Paper 25,January 2023

5. Conclusion

Thesystemcanbeusedsuccessfullyindifferentcorporatesectorsfordemandforecasting,inventoryoptimization,andbudgetoptimization for desired products and regions of operation. Any machine learning algorithm cannot be deemed perfect for makingpredictions.Light GBM usesa histogram-based algorithm i.e it buckets continuous feature values into discrete bins which fasten the training procedure. Replaces continuous values with discrete bins which results in lower memory usage. It produces much more complex trees by following leaf wise split approach rather than a level-wise approach which is the main factor in achieving higher accuracy. However, it can sometimes lead to overfitting which can be avoided by setting the max_depth parameter. It is capable of performing equally well with large datasets with a significant reduction in training time.

References

[1].MehdiKhashei and Mehdi Bijari, A novel hybridization of artificial neural networks and ARIMA models for time seriesforecasting,AppliedSoftComputing,Volume 11, Issue 2,2011

[2] L. Wu, J. Y. Yan, and Y. J. Fan, ”Data Mining Algorithms and Statistical Analysis for Sales Data Forecast,” 2012 FifthInternationalJointConferenceonComputationalSciencesandOptimization,Harbin, 2012

[3] T. Choi, C. Hui, and Y. Yu, ”Intelligent time series fast forecasting for fashion sales: A research agenda,” 2011 InternationalConferenceonMachine LearningandCybernetics, Guilin, 2011,pp.1010-1014.

[4] A.D.Lacasandile,J.D.Niguidula,andJ.M.Caballero,”Miningthepasttodeterminethefuturemarket:SalesforecastingusingTSDMframework,” TENCON2017-2017IEEE Region10Conference, Penang,2017,pp.461-466.

[5] P. Sobreiro, D. Martinho and A. Pratas, ”Sales forecast in an IT company using time series,” 2018 13th Iberian Conference onInformationSystemsandTechnologies(CISTI), Caceres,2018,pp.1-5.DOI:10.23919/CISTI.2018.8399191

[6] G. Nunnari and V. Nunnari, ”Forecasting Monthly Sales Retail Time Series: A Case Study,” 2017 IEEE 19th Conference onBusinessInformatics(CBI),Thessaloniki, 2017, pp. 1-6.

[7] B. Siregar, E. B. Nababan, A. Yap, U. Andayani and Fahmi, ”Forecasting of raw material needed for plastic products based inincomedatausingARIMAmethod,”20175thInternationalConferenceonElectrical,ElectronicsandInformationEngineering(IEEE), Malang, 2017,pp.135-139.

[8] F.M.Thiesing, U.Middelberg,andO.Vornberger,”Shorttermpredictionofsalesinsupermarkets,”ProceedingsofICNN’95-International ConferenceonNeuralNetworks,Perth,WA,Australia,1995,pp.10281031vol.2.

[9] C. Chatfield, ”Time series forecasting with neural networks,” Neural Networks for Signal Processing VIII. Proceedings of the1998IEEE Signal Processing SocietyWorkshop(Cat.No.98TH8378), Cambridge,1998,pp. 419-427.

[10] D. H. F. Yip, E. L. Hines, and W. W. H. Yu, ”Application of artificial neural networks in sales forecasting,” Proceedings ofInternationalDemandForecastingFor InventoryOptimization

[11]M.Gurnani,Y.Kirke,P.Shah,S.Udmale,V.Sambhe,andS.Bhirud,”Forecastingofsalesbyusinga fusionofmachinelearningtechniques,” 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), Pune, 2017, pp. 93-101.

[12]AiyunZheng,WeiminLiu,andFanggengZhao,”DoubletrendstimeseriesforecastingusingacombinedARIMAandGMDHmodel,”2010ChineseControland DecisionConference, Xuzhou,2010, pp.18201824

[13]https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/

[13] Jaiswal R. C. and Danish khan, “Arduino based Weather Monitoring and Forecasting System using SARIMA Time-SeriesForecasting”,JournalofEmergingTechnologiesandInnovativeResearch(JETIR),OpenAccess,PeerReviewedandrefereed Journal,ISSN-2349-5162, ImpactFactor:5.87,Volume 7, Issue11, pp.1149-1154, November 2020

[14] Jaiswal R.C. and Lokhande S.D, “A Novel Approach for Real Time Internet Traffic Classification”, ICTACT Journal onCommunicationTechnology,September2015,volume:06,issue:03,pp.1160-1166.(Print:ISSN:0976-0091,OnlineISSN:2229-6948(ImpactFactor:0.789in2015).

[15] JaiswalR.C.andLokhandeS.D,“StatisticalFeaturesProcessingBasedRealTimeInternetTrafficRecognitionandComparativeStudyofSixMachineLearningTechniques”,IMCIP-InternationalMultiConferenceonInformationProcessing-(ICCN-

InternationalConferenceonCommunicationNetworks-2014,PP-120-129,ISBN:9789351072515,UniversityVisvesvarayaCollegeofEngg.Departmentof Computer Scienceand Engineering Bangalore University, Bangalore.

[16]AiyunZheng,WeiminLiuandFanggengZhao,"DoubletrendstimeseriesforecastingusingacombinedARIMAandGMDHmodel,"2010ChineseControland DecisionConference, Xuzhou,2010, pp.18201824