IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

# Sales Prediction Using Machine Learning Algorithms

**K R SABEENA #[1], Mrs. Dr.G.ANITHA [#2]**

Master of Computer Application, Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India

Associate Professor, Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India

**ABSTRACT:** Machine Learning is remodeling each and every stroll of existence and has end up a fundamental contributor in actual world scenarios. The modern functions of Machine Learning can be considered in each and every subject inclusive of education, healthcare, engineering, sales, entertainment, transport and various more; the listing is in no way ending. The standard strategy of income and advertising desires no longer assist the companies, to cope up with the tempo of aggressive market, as they are carried out with no insights to customers' buying patterns. Major transformations can be considered in the area of income and advertising as a end result of Machine Learning advancements. Owing to such advancements, a variety of essential components such as consumers' buy patterns, goal audience, and predicting income for the latest years to come can be without difficulty determined, for this reason assisting the income group in formulating plans for a improve in their business. The goal of this paper is to recommend a dimension for predicting the future income of Big Mart Companies maintaining in view the income of preceding years. A complete find out about of income prediction is completed the usage of Machine Learning fashions such as Linear Regression, K-Neighbors Regressor, XGBoost Regressor and Random Forest Regressor. The prediction consists of records parameters such as object weight, object fats content, object visibility, object type, object MRP, outlet institution year, outlet measurement and outlet vicinity type.

**KEYWORDS:** Standard Scaler, Label Encoder, Linear Regression, K-Neighbors Regressor, XGBoost Regressor, Random Forest Regressor**.**

## I. INTRODUCTION

Sales is a existence blood of each business enterprise and income forecasting performs a necessary function in conducting any business. Good forecasting helps to boost and enhance enterprise techniques by way of growing the information about the marketplace. A general income forecast appears deeply into the conditions or the prerequisites that beforehand came about and then, applies inference concerning purchaser acquisition, identifies inadequacy and strengths earlier than putting a price range as nicely as advertising plans for the upcoming year.

In different words, income forecasting is income prediction that is based totally on the handy assets from the past. An in-depth understanding of the previous sources lets in to put together for the upcoming wishes of the enterprise and will increase the possibility to be successful irrespective of exterior circumstances. Businesses that deal with income forecasting as the foremost step, have a tendency to function higher than these facts mining predictive methods by way of stacking is viewed a two-level statistical approach. It is named as two-level due to the fact stacking is carried out on two layers in which backside layer consists of one or extra than one gaining knowledge of algorithms and pinnacle layer consists of one studying algorithm.

Stacking is additionally regarded as Stacked Generalization. It essentially includes the education of the mastering algorithm current in the pinnacle layer to mix the predictions made with the aid of the algorithms current in the backside layer. In the first step, all the studying algorithms are skilled the usage of the departmental shop dataset and in the 2d step, Stacking performs higher than any single mannequin due to the fact a stacking includes greater facts for prediction.

In this venture the method has been achieved underneath six stages. In first stage, statistics is accrued from dataset. In 2d stage, troubles are analysed from the records collection. In 0.33 stage, specialty of the records is explored. In fourth

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

stage, information cleaning is completed to realize and right the dataset. In fifth stage, information modeling methods is used to predict the data. In sixth stage, the characteristic engineering is used to import the information from the desktop studying algorithm. Sales prediction is completed precisely with the aid of the usage of computing device gaining knowledge of algorithms.

## II. RELATED STUDY

[1] 'Walmart's Sales Data Analysis - A Big Data Analytics
Perspective' In this study, inspection of the statistics amassed from a retail store and prediction of the future techniques associated to the store administration is executed. Effect of quite a number sequence of events such as the climatic conditions, vacation trips etc. can actually alter the kingdom of specific departments so it also studies these outcomes and examines its affect on sales.
[2] 'Applying computing device gaining knowledge of algorithms in sales prediction'This is a thesis in which countless wonderful tactics of machine getting to know algorithms are utilized to get better, optimal results, which are in addition examined for prediction task. It has made use of 4 algorithms, an ensemble technique etc. Feature choice has additionally been implemented using distinctive tactics.
[3] 'Sales Prediction System Using Machine Learning'In this paper, the goal is to get desirable effects for predicting the future income or needs of a company with the aid of applying
techniques like Clustering Models and measures for sales
predictions. The plausible of the algorithmic techniques are
estimated and for this reason used in similarly research.
[4] 'Intelligent Sales Prediction Using Machine Learning
Techniques' This lookup affords the exploration of the choices to be made from the experimental records and from the insights obtained from the visualization of data. It has used data mining techniques. Gradient Boost algorithm has been shown to showcase most accuracy in picturizing the future
transactions.
[5] 'Retail income prediction and object pointers using customer demographics at keep level'This paper outlines a income prediction gadget alongside with the product suggestion system, which was once used for the benefit of the crew of retail stores. Consumer demographic details have been used for exactly designing the income of each individual.
[6] 'Utilization of synthetic neural networks and GAs for
constructing an wise income prediction system' In the study, utilization of deep neural community methods is to know about their income method involving electronic components beforehand in time. Some optimization algorithms are also used to maximize the effectivity of the system: like Genetic Algorithm.
[7]'Bayesian gaining knowledge of for income charge prediction for thousands of retailers' In this paper it is proven that from the prediction of the single one's charge of transactions, many providers would benefit from it, that capability the statistics received ought to be beneficial for the building of a set-up that would estimate massive range of outputs. The prediction makes use of neural network approach. Here they have practiced Bayesian learning to obtain insights.
[8] 'Combining Data Mining and Machine Learning for
Effective User Profiling' This lookup describes the way of detecting suspicious behavior with the aid of using an automated prototype. Several machine mastering methodologies have been made in use for concluding this splendid prototype. Here statistics mining and constructive induction methods are merged to pull out the discrepancy observed in the conducts of the proprietors of cell phones.

.

## III. SYSTEM METHODOLOGIES

### *A. EXISTING SYSTEM*

In early days' income prediction and forecasting is no longer accomplished the use of any analytics. Sales prediction equipment and fashions had been no longer used to predict the income of a product. The evaluation of income does no longer have any patterns to advise the future forecasting of a product. The prediction is executed manually with the aid of accumulating the dataset of a product.

**DRAWBACKS**
Some of the drawbacks are:
- Manually gathering information consumes greater time.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

- Numerous facts used to be accumulated to deal with a product for forecasting.
- It depends on historic statistics to predict future forecasting.

### *B. PROPOSED SYSTEM*

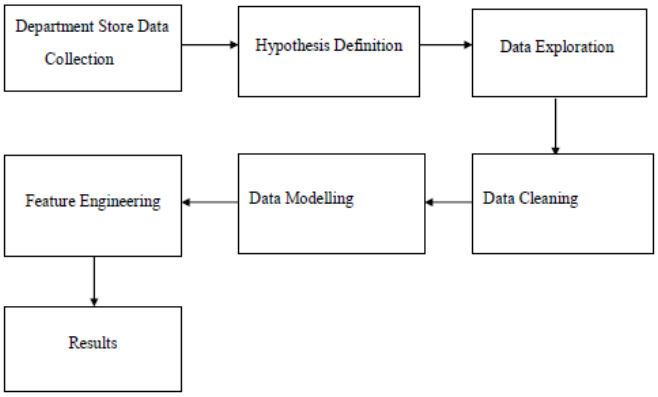
Predictive analytical algorithms and statistical fashions to analyze massive datasets to check the probability of a set of attainable outcomes. These fashions draw upon current, contextual, and historic information to predict the likelihood of future events. As new records is made available, the gadget accommodates greater records into the statistical mannequin and updates its predictions accordingly. Throughout this method of laptop mastering (ML), the mannequin receives "smarter" and predictions grow to be more and more accurate.

**FEATURES**

Some of the blessings are:

- Better alignment of income teams.
- Increased affectivity and productiveness of the income cycle.
- More correct income forecasts and predictions of future revenue.

### *C. FLOW DIAGRAM*




## IV. DESCRIPTION OF MODULES

- DATASET COLLECTION
- HYPOTHESIS DEFINITION
- DATA EXPLORATION
- DATA CLEANING
- DATA MODELLING
- FEATURE ENGINEERING

**DATASET COLLECTION**

A records set is a series of data. Departmental save facts has been used as the dataset for the proposed work. Sales information has Item Identifier, Item Fat, Item Visibility, Item Type, Outlet Type, Item MRP, Outlet Identifier, Item Weight, Outlet Size, Outlet Establishment Year, Outlet Location Type, and Item Outlet Sales.

**HYPOTHESIS DEFINITION**

This is a very vital step to analyse any problem. The first and most important step is to apprehend the hassle statement. The thought is to discover out the elements of a product that creates an influence on the income of a product. A null speculation is a kind of speculation used in facts that proposes that no statistical magnitude exists in a set of given observations.An choice speculation is one that states there is a statistically great relationship between two variables.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

**DATA EXPLORATION**

Data exploration is an informative search used by using records customers to structure real evaluation from the data gathered. Data exploration is used to analyse the records and records from the facts to shape genuine analysis. After having a seem to be at the dataset, positive records about the facts used to be explored. Here the dataset is no longer special whilst accumulating the dataset. In this module, the forte of the dataset can be created.

**DATA CLEANING**

In statistics cleansing module, is used to observe and right the inaccurate dataset. It is used to take away the duplication of attributes. Data cleansing is used to right the soiled statistics which consists of incomplete or out of date data, and the flawed parsing of report fields from disparate systems. It performs a sizeable section in constructing a model.

DATA MODELLING

In information modelling module, the desktop mastering algorithms have been used to predict the Wave Direction. Linear regression and K-means algorithm had been used to predict a number sorts of waves. The person gives the ML algorithm with a dataset that consists of preferred inputs and outputs, and the algorithm finds a approach to decide how to arrive at these results.

Linear regression algorithm is a supervised studying algorithm. It implements a statistical mannequin when relationships between the impartial variables and the based variable are nearly linear, suggests most appropriate results. This algorithm is used to exhibit the path of waves and its peak prediction with expanded accuracy rate.

K-means algorithm is an unsupervised studying algorithm. It offers with the correlations and relationships through analysing on hand data. This algorithm clusters the facts and predict the price of the dataset point. The educate dataset is taken and are clustered the usage of the algorithm. The visualization of the clusters is plotted in the graph.

**FEATURE ENGINEERING**

In the function engineering module, the system of the usage of the import information into computer mastering algorithms to predict the correct directions. A characteristic is an attribute or property shared with the aid of all the unbiased merchandise on which the prediction is to be done. Any attribute ought to be a feature, it is beneficial to the model.

## V. ALGORITHMS USED

**1. LOGISTIC REGRESSION**

Linear Regression Linear Regression is the most commonly and widely used algorithm Machine Learning algorithm. It is used for establishing a linear relation between the target or dependent variable and the response or independent variables. The linear regression model is based upon the following equation:

**2. K-NEIGHBORS REGRESSOR**

KNN algorithm for Regression is a supervised learning approach. It predicts the target based on the similarity with other available cases. The similarity is calculated using the distance measure, with Euclidian distance being the most common approach. Predictions are made by finding the K most similar instances i.e., the neighbors, of the testing point, from the entire dataset. KNN algorithm calculates the distance between mathematical values of these points using the Euclidean distance.

**3. XGBOOST REGRESSOR**

XGBoost also known as Extreme Gradient Boosting has been used in order to get an efficient model with high computational speed and efficacy. The formula makes predictions using the ensemble method that models the anticipated errors of some decision trees to optimize last predictions. Production of this model also reports the value of each feature's effects in determining the last building performance score prediction. This feature value indicates that outcome in absolute measures – each characteristic has on predicting school performance. XGBoost supports parallelization by creating decision trees in a parallel fashion. Distributed computing is another major property held by this algorithm as it can evaluate any large and complex model. It is an out-core-computation as it analyses huge and varied datasets. Handling of utilization of resources is done quite well by this calculative model. An extra model needs to be implemented at each step in order to reduce the error.

**4.RANDOM FOREST REGRESSOR**

Random Forest is defined as the collection of decision trees which helps to give correct output by making use of bagging mechanism. Bagging along with boosting are two of the most common ensemble techniques which intend to

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

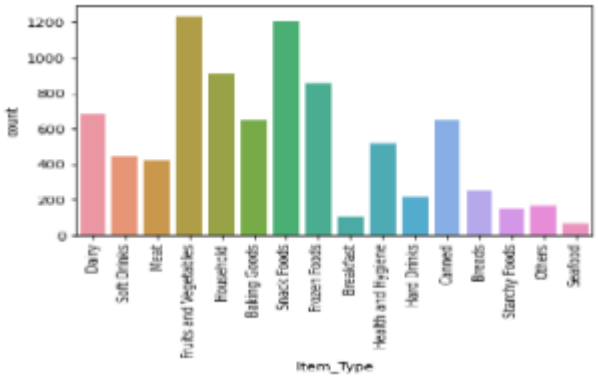
|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

tackle higher variability and higher prejudice. In bagging, we have multiple base learners, or we can say base models, which in turn takes various random samples of records from the training dataset. In case of Random Forest Regressor decision trees are the base learners, and they are trained on the data collected by them. Decision trees are itself not accurate learners as, when it is implemented up to its full depth, mostly there are chances of overfitting with high training accuracy, but low real accuracy. So, we give out the samples of the main data file by utilizing row sampling and feature sampling with replacement technique to each of the decision trees and this method is referred to as bootstrap. The result is that every model has been trained on all of these data files and then whenever we feed a test data to each of the trained one out there, the predictions estimated by each of them are combined in a way such that the final output is the mean of all of the results generated. The process of combining the individual results here is known as aggregation. The hyperparameter that we need to regulate in this algorithm is the no of decision trees to be considered to create a random forest.

## VI. RESULTS

Machine Learning algorithms such as Linear Regression, KNearest Neighbors algorithm, XGBoost algorithm and Random Forest algorithm have been used to predict the sales of various outlets of the Big Mart. Various parameters such as Root Mean Squared Error (RMSE), Variance Score, Training and Testing Accuracies which determine the precision of results are tabulated for each of the four algorithms. Random Forest Algorithm is found to be the most suitable of all with an accuracy of 93.53**%.**




## VII. CONCLUSION

Every business enterprise needs to comprehend the demand of the consumer in any season formerly to keep away from the scarcity of products. As time passes by, the demand of the save to be greater correct about the predictions will expand exponentially. So, massive lookup is going on in this area to make correct predictions of sales. Better predictions are at once proportional to the earnings made by using the departmental store. The reason of measuring accuracy used to be to validate our prediction with the authentic result. In this project, an effort has been made to predict income of the product from an outlet precisely via the use of a two-level statistical mannequin that reduces the imply absolute error value. The two-level statistical model carried out than the different single mannequin predictive strategies and contributed higher predictions to the departmental save dataset.

## VIII. FUTURE WORK

Further enlargement of the machine additionally can be performed in future if needed. The software can be greater in the future with the desires of the departmental store. The database and the data can be up to date to the brand new imminent versions. Thus, the gadget can be altered in accordance with the future necessities and advancements. System overall performance comparison need to be monitored now not solely to decide whether or not they operate as format however also to decide if they have to have to meet modifications in the facts wanted for the departmental store.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2023 ||

| DOI:10.15680/IJIRSET.2023.1203033 |

## REFERENCES

[1] Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan FS Mohammed, and Mahmood A. Rashid."Walmart's Sales Data Analysis-A Big Data Analytics Perspective." In 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp. 114-119. IEEE, 2017.
[2] Sekban, Judi. "Applying machine learning algorithms in sales prediction." (2019).
[3] Panjwani, Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar, and Rupali Hande. Sales Prediction System Using Machine Learning.No. 3243.EasyChair, 2020.
[4] Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. "Intelligent Sales Prediction Using Machine Learning Techniques."In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), pp. 53-58.IEEE, 2018.
[5] Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." ACM SIGKDD Explorations Newsletter 10, no. 2 (2008): 84-89.
[6] Baba, Norio, and Hidetsugu Suto. "Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system."In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks.IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 6, pp. 565-570. IEEE, 2000.
[7] Ragg, Thomas, Wolfram Menzel, Walter Baum, and Michael Wigbers. "Bayesian learning for sales rate prediction for thousands of retailers." Neurocomputing 43, no. 1-4 (2002): 127-144.
[8] Fawcett, Tom, and Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling."In KDD, pp. 8-13. 1996

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details