International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

ITEM RECOMMENDATION SYSTEM USING MACHINE LEARNING

[1] GEETHA M, [2]UNZILA ANJUM KALMANI

[1]*Assistant Professor, Department of Master of Computer Applications, BIET, Davangere*
[2]*Student, Department of MCA, BIET, Davangere*

Abstract: Every business and organisation today is focused on both meeting client expectations and managing inventories as we achieve new heights of development. By comprehending the pattern from previous sales records, the models they apply enable them to forecast future demand. Everyone has recently given up on using the conventional sales forecasting prediction algorithms since they take so long to produce the desired outcomes. As a result, retailers now keep track of their sales history using a data set that includes information on price tag, outlet type, outlet location, item visibility, item outlet sales, etc.

Keywords: Prediction, Regression, Machine Learning, XG Boost, Big Mart, Data Science.

I. INTRODUCTION

For any store to thrive in the market and compete with other retailers, maintaining customer satisfaction and meeting product demand is crucial. And you can only do these two if you have a projection of future need for developing fresh ideas for a successful firm. A growing population also means that there are more shops and shopping centres, which puts more businesses in rivalry with one another for customers and attention. Businesses, including grocery stores and retail outlets, require analysis to look at trends and forecast future sales. Many companies and organisations track the statistical information about their goods in order to forecast consumer demand. The statistics that are saved include the Any store that wants to remain in business and compete with other retailers must prioritise satisfying its customers and meeting product demand. And you can only do these two if you have a projection of future need for developing fresh ideas for a successful firm. A growing population also means that there are more shops and shopping centres, which puts more businesses in rivalry with one another for customers and attention. Businesses, including grocery stores and retail outlets,

require analysis to look at trends and forecast future sales. Many companies and organisations track the statistical information about their goods in order to forecast consumer demand. The statistics that are saved include which is an important part of a big mart's management.

II.LITERATURE SURVEY

Predictive models have been used by Rohit Sav, Pratiksha Shinde, and Saurabh Gaikwad to measure major mart sales. They used the XG Booster technique after cleaning the collected data. When compared to other algorithms, it was found that XGBost Regressor had the highest accuracy rate. They came to the conclusion that XG boost may be used to anticipate big-mart sales as a result of this.[1]

Dr. Venkata Reddy Medikonda, Inedi Theresa, and K.V. Narasimha Reddyin examines sales forecasting using exploratory machine learning techniques. They completed the entire process by determining the correct phases, which included data gathering, thesis generation to effectively comprehend bugs, and additional data cleaning and processing. To forecast the outcome of the sales, methods like Linear Regression, Decision Tree Regression, Ridge Regression, and Random Forest model were utilised. In comparison to the single model prediction technique, they came to the conclusion that using multiple modelling implementation enabled them to make superior predictions[2].

Rajguru, Ketkar, Ketkar, Shevade, and Kadam in proposed a model that utilises a random forest technique and multiple

linear regression successfully. A specific data collection, including Item_Identifier, Item_Weight, Item_Fat Content, Item_Visibility, Item_Type, Outlet_Identifier, etc., was employed with this model to anticipate huge mart sales predictions [3].

Both Neeta Nain and Gopal Beherain apply the idea behind using GSO to optimise parameters and forecast sales. Their primary focus is on retailers. They have also used XGBoost algorithms to estimate sales and execute Hyperparameter optimization[4].

Rajendra Pamula, Praphula Kumar Jain, and Kumari Punamin A two-level strategy has been developed to predict sales of goods that promise to be more efficient: A Two-Level Statistical Model for Big Mart Sales Prediction. It involves stacking algorithms so that the bottom layer has one or more algorithms and the top layer only has one learning algorithm[5].

Better sales projections are obtained usIng this two-level modelling methodology than using a single model predictive technique. Spandana M. and Ranjitha P. in Xgboost, Linear regression, Polynomial regression, and Ridge regression approaches have been used in Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms to forecast big mart sales[6].

Mr. Bohdan Pavlyshenkoin present the viewpoint of machine generalisation. When there are fewer data points in the system, maybe as a result of the introduction of new goods or outlets, this mostly comes into play. The unique stacking technique was used to build the regression[7].

Better performance and efficiency in sales prediction follow from this. Karan Singh and Nikita Malik To solve the issue statement, which aims to forecast big-mart sales, use the idea of machine learning algorithms. It shows relationships between various features and outlet sizes, which illustrate varying sales rates. They deduce that a specific size with a comparable design will have a comparable rate of sales success[8].

Both Neeta Nain and Gopal Behera have mentioned the xgboost, decision tree, and linear regression techniques. Among these models, XG boost exhibits the highest accuracy, making it a model that is strongly advised. In order to outperform other models, MAE and RMSE were kept within low bounds[9].

Madhuri Ghuge, Saurab Dhawale, Akanksha Dubey, and Archisha Chandel in describes a five-step strategy for predicting big-mart sales. To achieve successful results, data is divided into testing and training groups. Bivariate and univariate analyses of the data are performed. For improved outcomes, data is afterwards pre-processed, changed, and then converted utilising various algorithms[10].

III. EXISTING SYSTEM

The fact that this discussion is still going on demonstrates the growing interest in the significance of completely varied diversity in conceptual frameworks. The development of algorithmic methods for improving overall diversity of vision, which may be intuitively gauged by the number of particular items predicted across all clients, has been the focus of recent study. Accuracy is sacrificed for higher reasonable variety (both individually and collectively). Everyone wants to raise precision when it comes to prophetic frameworks, which is why the Netflix Prize was given out in July for a calculation that improved the administration's expectation calculation's accuracy by 10%. However, creative assorted diversity is a new statistic that computer scientists are learning to use to measure advancement. Once upon a time, accuracy was considered to be the most crucial aspect of future-oriented knowledge, including.

IV.PROPOSED WORKS

The creation of a machine learning software model to analyse customer transactional datasets of online product sales in order to assess and forecast item sold transaction along with customer reviews to determine which business and which item are best to buy.

We described our communitarian, divisive strategy to handling optimal recommendation. Proposal framework aids in the sale of a larger variety of goods when compared to products that are typically given without any kind of

suggestion. Using recommender systems, customers can choose elusive items that may be difficult to find without receiving explicit instructions.

V. DATA SET AND TOOLS USED

Tools used

Anaconda Navigator

Boa constrictor Navigator is a work area graphical UI (GUI) remembered for Anaconda® dispersion that permits you to dispatch applications and effectively oversee conda bundles, conditions, and channels without utilizing order line orders.

Tkinter

Tkinter is a Python authoritative to the Tk GUI toolbox. It is the standard Python interface to the Tk GUI toolkit and is Python's true standard GUI.

Data set

The data set is created by own,This data set cantains item category,item name,item id,item rating.

	A	B	C	D	E	F	G	H	I
1	slno	ica	id	Item	iw	ip	ico	ir	
2	1	Electronics	AD12	Mobile	1	20000	Vivo Y56	1	
3	2	Kitchen	AD13	Pressure Cooker	3	1000	Prestige	1	
4	3	Beauty	AD14	Lipstick	1	500	MyGlamm	1	
5	4	Furniture	AD15	Dining Set	3	25000	RoyalOak	1	
6	5	Fashion	AD16	Shoes	5	3000	Nikee	1	
7	6	Pet Supplies	AD17	Cat Food	2	500	Pedigree	1	
8	7	Fruits	AD18	Apple	8	100	Kashmir	1	
9	8	Daily Essentials	AD19	Brush	1	60	Toilet Brush	1	
10	9	Education	AD20	Books	1	400	Classmate	1	
11	10	Automotive	AD21	Helmates	1	500	Vega	1	
12	11	Office and Professional	AD22	Computers	2	65000	hp	1	
13	12	Sports	AD23	Badminton	1	700	KRB	1	
14	13	Electronics	CD78	Mouse	1	700	hp	1	
15	14	Kitchen	BD77	Plates	3	250	Life	1	
16	15	Beauty	BD56	Lipstick	1	500	Sugar	1	
17	16	Furniture	KR32	Dressing Table	15	24000	RoyalOak	1	
18	17	Fashion	BD55	Sandles	1	800	Flask	1	
19	18	Pet Supplies	ME36	Cat Food	2	500	Pedigree	1	
20	19	Fruits	DF88	Orange	8	150	ABC	1	
21	20	Daily Essentials	AK90	ToothPaste	1	40	Colgate	1	
22	21	Education	KK29	Files	1	30	def	1	
23	22	Automotive	PA45	Gloves	1	300	Vega	1	
24	23	Office and Professional	AR88	Laptops	2	55000	Dell	1	
25	24	Sports	RD33	CricketBat	2	250	KRB	1	
26	25	Fashion	AD15	Shoes	1	3000	Nikee	1	
27	26	Automotive	AD21	Helmates	1	500	Vega	1	

Fig 1 : Data set for item recommendation system

VI.METHODOLOGY

In this work the model is proposed the Xgboost technique. For the proposed model to be built, each stage is essential. Employed an ensemble classifier using Decision trees, Linear regression, Ridge regression, Random forest, and Xgboost after preprocessing and filling in missing values. For estimating the sales in Big Mart, the accuracy metrics MAE and RSME are both used. According to the accuracy measures, the model will forecast most accurately when the MAE and RSME are kept to a minimum.

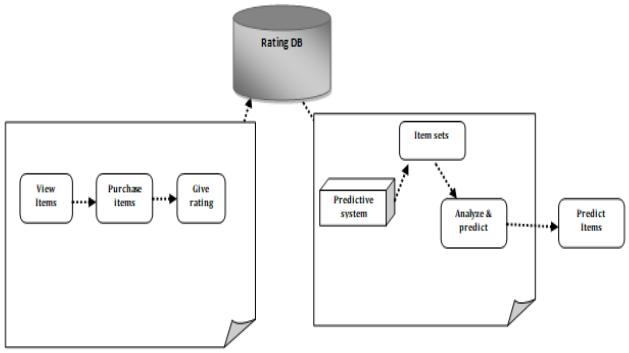

Fig 2 : System Architecture

VII. IMPLEMENTATION

KNN(K-NearestNeighbor):

Algorithm that can be used for classification regression tasks. It is a lazy learning algorithm, which means it does not h` ..
ve a specific training phase,but rather memorizes the training dataset.

The K-NN working can be explained on the basis of the below algorithm:

Step-1: Select the number K of the neighbors.

Step-2: Calculate the Euclidean distance of K number of neighbors

Step-3: Take the K nearest neighbors as per the

calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.

Step-6: Our model is ready.

DecisionTree:

A Decision Tree is a supervised learning algorithm used for classification and regression tasks. It creates a tree-like model of decisions and their possible consequences. The algorithm divides the data set into smaller subsets based on the attributes, which results in a tree-like structure.

The complete process can be better understood using the below algorithm:

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).
Step-3: Divide the S into subsets that contains possible values for the best attributes.
Step-4: Generate the decision tree node, which contains the best attribute.
Step-5: Recursively make new decision trees using the subsets of the dataset created in step-3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

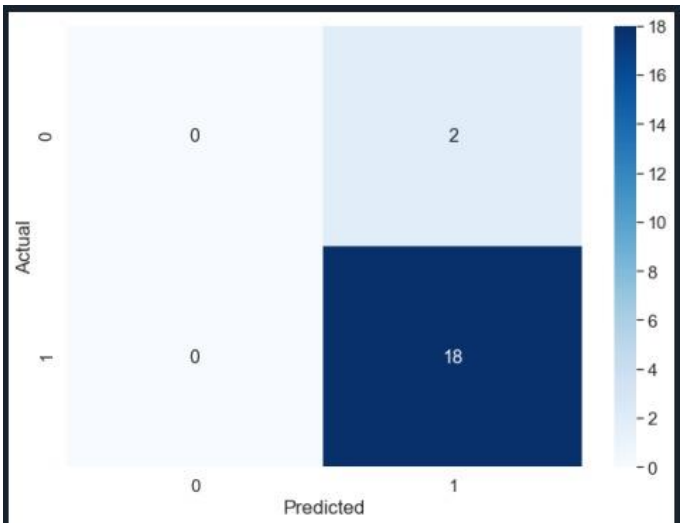

Fig 3 : Confusion Matrix of Drug Usage

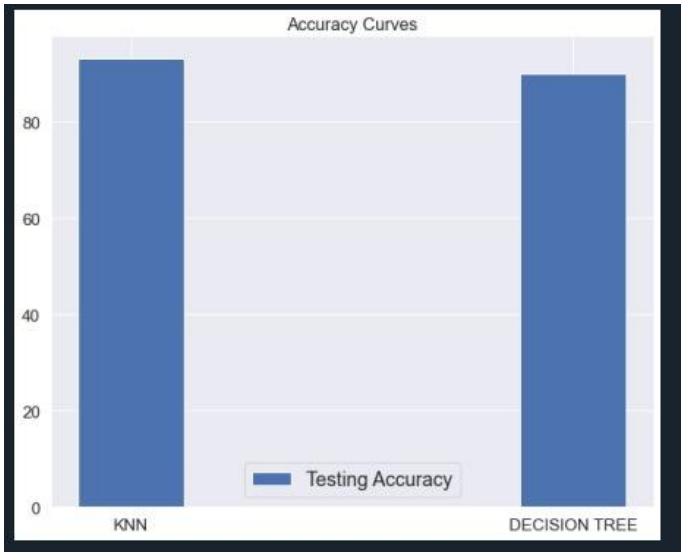

Fig 4 :Testing Accuracy

```
 DECISION TREE ALGORITHM
[[ 0  2]
 [ 0 18]]
accuracy= 0.9

 Accuracy Of KNN
93.10344827586206

 Accuracy Of Decision Tree
accuracy= 90.0
```

Fig 5: Accuracy of KNN and Decision Tree

VIII. DATA PROCESSING

A. Data Gathering

The data was securely gathered by us using a predetermined approach. The process for the data collection varies from client to client and depends on the kind, amount, and availability of the data as well as the client's needs.

B. Data Preprocessing and Cleaning

In order to ensure that the data is properly segmented, detected gaps in the data are filled with the relevant information, making the data compatible, and addressing faults in storage systems that may result in data redundancy, the acquired data is put through a "cleaning" process.

C. Database Modelling

This procedure essentially consists of analysing the given dataset and the objects inside it to gain a clear understanding of the needs that could support our business model. Models are then developed for the project's defined flow based on the examination of data patterns. This flow provides improved support for utilising the project's features in the semi-formal model that was previously agreed upon. Additionally, it offers instructions for understanding the connections between the data objects and other things.

D. Prediction of Data

This technique involves training machine learning prediction models, which are later tested against the data. The preprocessed dataset will then get this application. Linear Regression, Random Forest, Decision Tree, and XG Boost Regressor are a few of the models to be utilised for the prediction. Data visualisation, or E In order for consumers and the admin to draw conclusions and make wise judgements about the data objects and other objects, the data analysis is then further visualised.

IX. RESULTS

Fig 6: Data Classification With different Criteria

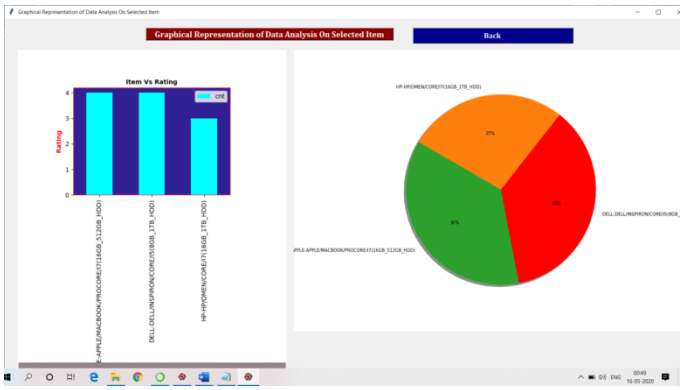

Fig7: Graphical Representation of Analyzed Dataset

CONCLUSION

We used the XGBoost, Random Forest, Linear Regression, and Decision Tree methods. As a consequence of the findings, we can say that when the four algorithms are compared as a whole, XGBoost has the highest accuracy (61.14%). Therefore, XGBoost is the superior algorithm for effective sales analysis. The main users of this methodology are shopping malls, supermarkets, brand stores, etc. In addition to offering a very efficient method of managing sales, data analysis applied to predictive machine learning models generously supports improved decisions and strategy planning based on future demands.In the modern world, this strategy is highly encouraged because it helps many businesses, organisations, researchers, and brands achieve results that lead to management of their earnings, sales, inventory management, data research, and customer demands.

REFERENCES

1. Rohit Sav, Pratiksha Shinde, Saurabh Gaikwad (2021, June). Big Mart Sales Prediction using Machine Learning.2021 International Journal of Research Thoughts (IJCRT).

2. Inedi. Theresa, Dr. Venkata Reddy Medikonda,K.V.Narasimha Reddy. (2020, March). Prediction of Big Mart Sales using Exploratory Machine Learning Techniques 020 International Journal of Advanced Science and Technology (IJAST).

3. Heramb Kadam, Rahul Shevade, Prof. Deven Ketkar , Mr. Sufiyan Rajguru (2018). A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression. (IJEDR).

4. Gopal Behere, Neeta Nain (2019). Grid Search Optimization (GSO) Based Future Sales Prediction for Big Mart. 2019 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS).

5. Kumari Punam , Rajendra Pamula , Praphula Kumar Jain (2018, September 28-29). A Two-Level Statistical Model for Big Mart Sales Prediction. 2018 International conference on on Computing, Power and Communication Technologies

6. Ranjitha P, Spandana M. (2021). Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms.Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021).

7. Bohdan M. Pavlyshenko (2018, August 25). Rainfall Predictive Approach for La Trinidad, Benguet using Machine Learning Classification. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP).

8. Nikita Malik, Karan Singh. (2020, June). SalesPrediction Model for Big Mart

9. Gopal Behere, Neeta Nain. (2019, September). A Comparative Study of Big Mart Sales Prediction.

10. Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge (2019, April). Sales Prediction System using Machine Learning. International Journal of Scientific Research and Engineering Development.

11.https://www.javatpoint.com/machine-learning-random-forestalgorithm

12.https://www.javatpoint.com/machine-learning-decision-treeclassification-algorithm

13. A. Chandel, A. Dubey, S. Dhawale and M. Ghuge,;Sales Prediction System using Machine Learning; International Journal of Scientific Research and Engineering Development, vol. 2, no. 2, pp. 1-4, 2019. [Accessed 27 January 2020].