

Product Sales Forecasting and Prediction Using Machine Learning Algorithm

P. Guru¹, J. Sathyapriya^{2*}, K. V. R. Rajandran³, J. Bhuvaneswari⁴, C. Parimala⁵

Submitted: 08/09/2023

Revised: 22/10/2023

Accepted: 07/11/2023

Abstract: Sales forecasting plays a crucial role for companies involved in various industries such as retailing, logistics, manufacturing, marketing, and wholesaling. The utilization of this approach enables companies to effectively allocate resources, accurately forecast sales revenue, and develop strategic plans that promote long-term organizational success. The conventional employment of statistical techniques for predicting sales at supermarkets has left numerous challenges unattended, leading to the development of models for prediction that exhibit suboptimal performance. The contemporary age of voluminous data, in conjunction with the availability of extensive computational resources, has rendered machine learning a preferred approach for predicting sales. This study demonstrates improved performance in predicting product sales from a specific store compared to existing machine learning algorithms. A K-Nearest Neighbor (K-NN) predictive model was constructed to anticipate sales for a business, specifically Big-Mart. The model's efficacy was then contrasted with that of Linear, Polynomial, and Ridge regression methodologies.

Keywords: Machine Learning, forecasting, predictive model, K-NN, Regression model.

1. Introduction

Sales forecasting is a crucial area of focus that has long been recognized for its importance. The efficacy of marketing organizations relies on the adoption of an effective and optimal forecasting method by vendors. The manual implementation of this task may result in significant errors that could negatively impact the reputation of the organization. In addition, it would take a lot of time, which is not practical in the modern era. The business sectors play a significant role in the global economy by producing sufficient quantities of goods to meet collective demands [1].

The state-of-the-art machine learning approach available today provides tools to foresee or forecast revenues for any business, which is immensely advantageous in mitigating the costs associated with prediction. Accurate predictions are advantageous for the development and enhancement of marketing strategies in the marketplace, rendering them notably beneficial.

The key goals of this machine learning research are to predict future product sales and monitor sales at a single business [2]. This machine learning model is used for supermarket sales forecasting in an effort to identify which shop and product features contribute most to increased sales. Big Mart uses Sales Prediction methods to predict sales of various products at its many stores in different locations across the country.

This research plans to use a machine learning algorithm to look at the features of products and stores that make a big difference in sales growth. The language used for coding in this context is Python, while the tools employed are Jupyter Notebook [3]. The development of this application involves the utilization of machine learning techniques, specifically the tasks of Supervised Learning and Regression. The primary objective of this practice is to forecast the future sales of a company's stores.

2. Literature Survey

Gopagani et al. [4] employ a method to forecast customer purchasing behavior based on the provided elements. This model is developed using a machine learning method known as linear regression. Customers with higher beta values are shown to have higher average order values, as determined by the supervised machine learning algorithms and linear regression model. Customers residing at a considerable distance from the store, those with a higher income, and those who receive catalogs via mail are likely to exhibit a greater propensity to spend. Customers with a single child tend to spend more than those with multiple children.

^{1,2*}Associate Professor, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India

gurup@pmu.edu & sathyapriya@pmu.edu*

³Professor, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India

kvrrajan@pmu.edu

⁴Assistant Professor, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India

bhuvaneswari@pmu.edu

Assistant Professor, Selvamm Arts and Science College (Autonomous), Namakkal, India

eparimala81@gmail.com

In another paper, Odegua [5] conducted an analysis on sales forecasting for a retail chain known as "Chukwudi Supermarkets" utilizing three distinct ML algorithms, namely K-NN, boosting, and random forest. In this study, the author has reported that the random forest learning approach attains a lower Mean Error value compared to the other two methods.

In their publication, Grigorios Tsoumakas et al. (2015) present a synopsis of various machine learning algorithms utilized for the purpose of predicting food sales. The present study examines crucial design choices that a data analyst must make when performing food sales prediction. These include things like picking good input variables for sales forecasting and modeling the sales output variable at the right time level. Furthermore, this study examines ML algorithms that have been utilized for the purpose of predicting food sales and the suitable metrics for assessing their precision. The article ultimately addresses the primary obstacles and prospects pertaining to the practical implementation of machine learning within the context of forecasting food sales.

Kaneko and Yada [7] presented a deep learning methodology for forecasting retail store sales. The purpose of this research is to apply deep learning techniques to the problem of retail sales forecasting, which has recently attracted a lot of attention in the rapidly developing field of ML. By employing a model of this nature for analysis, it is possible to develop a strategy for managing stores. Consequently, a deep learning algorithm incorporating L1 regularization yielded an 86% accuracy rate in predicting sales. In contrast, a reduction of approximately 13% in accuracy was observed upon implementation of the logistic regression model. The findings suggest that Deep Learning (DL) is a highly appropriate approach for developing models that incorporate multiple attributes. The results of the present research show that DL may be used successfully to examine retail POS data.

In their publication, Bajaj et al. (2018) conducted a thorough investigation into the prediction of sales through the implementation of various ML models, including various regressions such as Linear, K-Neighbors, XGBoost, and Random Forest. The present study emphasizes the significance and benefits of utilizing machine learning for predictive purposes. The discussion encompasses a range of machine learning algorithms and their respective performances. Diverse parameters employed for the evaluation of quality were also deliberated.

Elcio Antonio Tarallo [9] conducts an exploratory investigation on the application of machine learning to

forecast demand for consumer products that move quickly, have a short shelf life, and are highly perishable. The findings suggest that machine learning outperforms traditional statistical techniques in terms of accuracy, resulting in improved inventory management across the supply chain. This, in turn, improves the availability of goods for customers, decreases stockout rates at retail locations, and increases revenue.

The primary advantage that was noted pertained to an elevated degree of accuracy in the prediction of demand. As a result, manufacturers can improve inventory balance throughout the supply chain, adjust production volumes and locations, and create more efficient sales and operations plans (S&OP). Retailers have the potential to enhance their operational efficiency by optimizing their in-store management, restocking procedures, and inventory control to minimize shortages or surpluses. When things are readily available, it can increase sales, cut down on waste, and make customers satisfied.

The results of the analysis indicate that the utilization of modern statistical techniques yielded superior sales forecasting outcomes compared to traditional methods. This was evidenced by the attainment of higher levels of accuracy relative to prior models as well as the ability to accommodate an increased number of data parameters and process large data volumes, thus demonstrating greater flexibility. Research has indicated that intricate analyses, such as the evaluation of promotional effects, are most effectively addressed through the utilization of Machine Learning methodologies.

Ranjitha and Spandana (2010) utilize predictive analysis and machine learning algorithms to analyze sales data at Big Mart. This study examines the efficacy of different algorithms in analyzing revenue and review data. We find the best technique for using regression to forecast sales based on past sales data, and we suggest a software solution for doing so. Polynomial regression, Ridge regression, and Xgboost regression can all be used to boost the precision of linear regression forecasts. In conclusion, the author found that the ridge and Xgboost regression models outperformed the linear and polynomial regression models in terms of accuracy, mean absolute error (MAE), and root mean squared error (RMSE).

The article authored by Melvin Tom et al. [11] discusses the utilization of regression analysis for the purpose of predicting supermarket sales. The process of forecasting sales of a product from a specific store is executed in a manner that yields superior efficacy in contrast with other machine learning methodologies. The authors of this study have presented a novel model utilizing the Xgboost

method for the purpose of forecasting sales at companies such as Big Mart. Their findings indicate that this model exhibits superior performance in comparison to other pre-existing models. This paper conducts a comparative analysis of the performance metrics of this model and other models. The present study demonstrates the correlation between various attributes under consideration and elucidates how a specific medium-sized location achieved the highest sales.

3. Proposed Work

3.1 Data Set Preparation

The present study provides a comprehensive description of the Bigmart Sale dataset. The dataset is titled "BigMart". The dataset comprises 12 attributes. Out of these 12 features, Item Outlet Sales serve as the response variable, while the other characteristics are mostly employed as predictors. The dataset comprises 8523 products distributed among various urban centers. Careful deliberation leads to the compilation of data, which is then split 80:20 between a training and a testing dataset [12].

Both the Python programming language and the Jupyter Notebook environment are used. The construction of this model involves the utilization of ML features, in particular the Regression function and the supervised learning function. The primary objective of this exercise is to forecast the forthcoming sales of merchandise sold by the company's retail outlets. The process flow of the suggested model is shown in Figure 3.1.

3.2 Preprocessing of Data

Data preparation is an essential part of any data analysis process, which entails converting unprocessed data into a structure that is appropriate for subsequent analysis and modeling. The objective is to enhance the caliber and applicability of the data by resolving a range of concerns, including absent values, anomalies, incongruous formatting, and additional factors. The following are several prevalent methods utilized in the preprocessing of data:

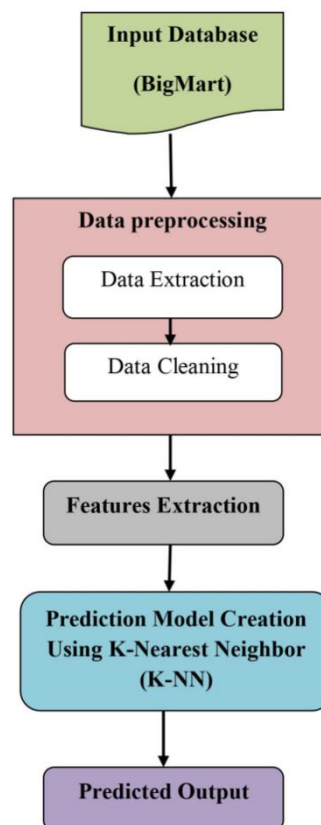


Fig 3.1: Workflow of proposed algorithm

3.2.1 Data Extraction

At this stage, we extract the relevant data we require from the dataset. The goal is to separate the data currently available from the information produced through

hypotheses. The dataset contains 1559 distinct products and 10 exclusive outlets. The Item type comprises a set of 16 distinct and exclusive values. Some things have been seen with the misspelled versions "regular" and "LF" for

"Regular" and "Low Fat" and "High Fat," respectively, despite the fact that there are two different types of item fat content.

3.2.2 Data Cleaning

- **Handling missing data:** One may opt to eliminate observations with absent values, impute absent values with suitable estimations such as mean, median, or mode, or employ sophisticated methodologies such as computation or machine learning algorithms.
- **Removing duplicates:** To mitigate biases and redundancies, it is recommended to identify and remove duplicate data from the dataset.
- **Fixing inconsistent data:** It is important to ensure consistency in formatting, capitalization, and spelling to maintain uniformity.

As mentioned in the previous part, Outlet_Size and Item_Weight have missing data. In our study, if a number for Outlet Size is missing, we fill it in with the function of the attribute that goes with it. Also, when it comes to Item Weight, we fill in the missing numbers with the average of their attribute values. The missing numerical values can be addressed through imputation techniques such as mean and mode replacement. This method has been shown to lower the association between the imputations. Our proposed framework is based on the idea that the decided attribute and the imputed attribute don't depend on each other.

3.3 Feature Extraction

- **Creating new features:** The process of extracting novel features from pre-existing ones is a crucial step in enhancing the efficacy of machine learning algorithms by capturing pertinent information.
- **Dimensionality reduction:** One can employ methods such as principal component analysis (PCA) or methods of feature selection to preserve crucial information while decreasing the number of features.

At the phase of Data exploration, specific characteristics have been noticed in the data set contained within our input database. This stage involves addressing all the intricacies present in the dataset and preparing it for constructing an appropriate model. Throughout the study, it was seen that the "Item visibility" attribute had a value of 0. In such cases, the mean of the item visibility for that particular product was utilized as a substitute for the zero-

value attribute. The resolution of discrepancies in categorical attributes is achieved by transforming all such attributes into a suitable form. In order to mitigate this issue, a third classification called "Item fat content" is established. An item identifier is an attribute that is defined by a unique identity code that starts with either DR, FD, or NC. The "Item Type New" property is set up with values like "Foods," "Drinks," and "Non-consumables". Ultimately, to ascertain the age of an outlet, a supplementary variable, namely year, is appended to the dataset.

3.4 Predicted Model creation

The BigMart dataset is subjected to the k-Nearest Neighbors (k-NN) algorithm for analysis. This section includes a discussion of computations in mathematics and visualization models.

3.4.1 K- Nearest Neighbor algorithm

The proposed K-Nearest Neighbor (K-NN) method is an ML algorithm that is usually regarded as straightforward to execute. The task of forecasting sales can be transformed into a classification problem based on similarity. The sales data with historical significance and the data for testing purposes are transformed into a collection of vectors. Each vector denotes i dimension for each feature. Subsequently, a similarity metric, such as the Euclidean distance (ED), is calculated to make a determination. The description of the K-NN algorithm is given in this section. The K-NN algorithm is classified as a lazy learning technique, as it does not construct a model or function beforehand. Instead, it identifies the k records from the training dataset that exhibit the closest resemblance to the test data or query record. Subsequently, an overwhelming vote is executed among the chosen k records to ascertain the class label, which is subsequently allocated to the query record.

Working of K-NN

Assuming that there are two separate groups, A and B, and that there is a new data point x_1 , it is necessary to figure out which group this data point belongs to. In order to address this particular issue, it is necessary to employ a K-NN algorithm. Using the K-NN method makes it easy to find an array or class of data in a given dataset. Please consider the diagram presented below:

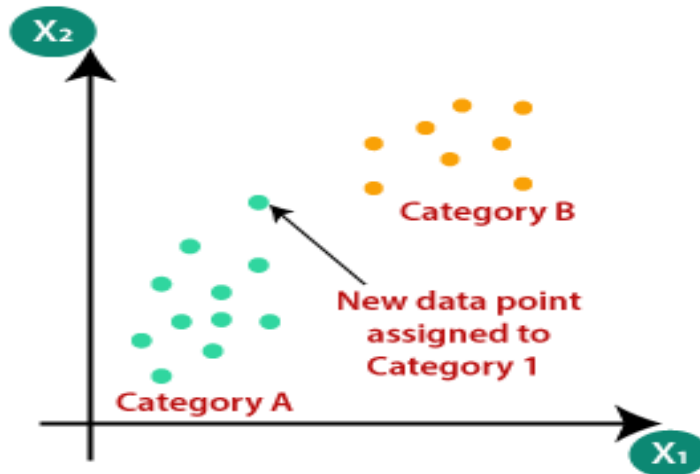


Fig 1: Working of K-NN

Assuming the presence of a novel data point, it becomes necessary to allocate it to the appropriate category. As illustrated in Figure 2.

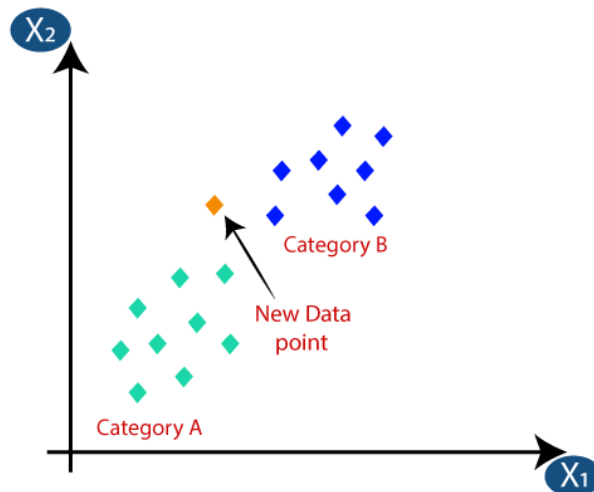


Fig 2: K-NN prediction

- Initially, the selection of the total number of neighbors will be made, with the present illustration opting for $k = 5$. Subsequently, the Euclidean distance will be computed among the given data points. The Euclidean distance, a concept previously explored in geometry, refers to the measure of the distance that exists between two given points. Euclidean geometry is a mathematical system that is based on the work of the ancient Greek mathematician Euclid. A mathematical method is used to figure out how far it is between points A and B:

$$\text{Euclidean Distnce (ED)} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \dots \dots \dots (1)$$

We used Euclidean distance to find out who our closest neighbors were. If three of the nearest neighbors are in

category A and two are in category B, We can see that this new data point must be in category A because the three closest data points are also in category A.

K-Nearest Neighbor Algorithm for sales prediction

- Step 1:** Chose the number of K neighbors
- Step 2:** Use equation 1 to find out the Euclidean distance between K neighbors
- Step 3:** count the number of data points that are shared by these adjacent points
- Step 4:** Set the new data points in the category with the highest number of their neighbors, then assign those categories to each other.
- Step 5:** Predict the sales depend on the maximum number of data points

3.4.2 Pseudo code for K-NN

Algorithm: K-Nearest Neighbor

Input: Sales report in $m \times m$ matrix format and $DM[1\dots m, 1\dots m]$ distance matrix

Output: prediction of sales report of supermarket

```

for n=1 to m do
visited [n]← false
initialize the list with k predicted value
Visited [k]← true
Current value==new k value
for i= to m do
find the lowest element in current row and unmarked
Column in i containing the element
current value←i
visited[i]← true
add i to the end of list

```

```

add k to the end of list
return list

```

The working of proposed algorithm with DataMart data set is given in figure 3.

4. Result Analysis

4.1 Database

The present study aims to forecast customer behavior using the BigMarts Sales dataset gathered in 2013. The input dataset is divided into 2 subsets: the training, which comprises 8523 records and 12 attributes, and the test set, which includes 5681 records and 11 attributes. The training set contains both independent and dependent variables, as listed below.

Table 1: In the input dataset, both the Attribute and the Variable

Item_Identifier	Product ID
Item_Weight	Weight of Product
Item_Fat_Content	Fat content of Product- Low/Regular
Item_Visibility	Parameter to know the visibility/reach of product
Item_Type	Category of Product
Item_MRP	Maximum Retail Price of the Product
Outlet_Identifier	Store ID
Outlet_Establishment_Year	The Year in which store is established
Outlet_Size	Areawise distribution of Stores- Low/Medium/High
Outlet_Location_Type	Type of city in which outlet is located
Outlet_Type	Type of outlet - Grocery store or supermarket
Item_Outlet_Sales	Sale price of product - The dependant variable to be predicted

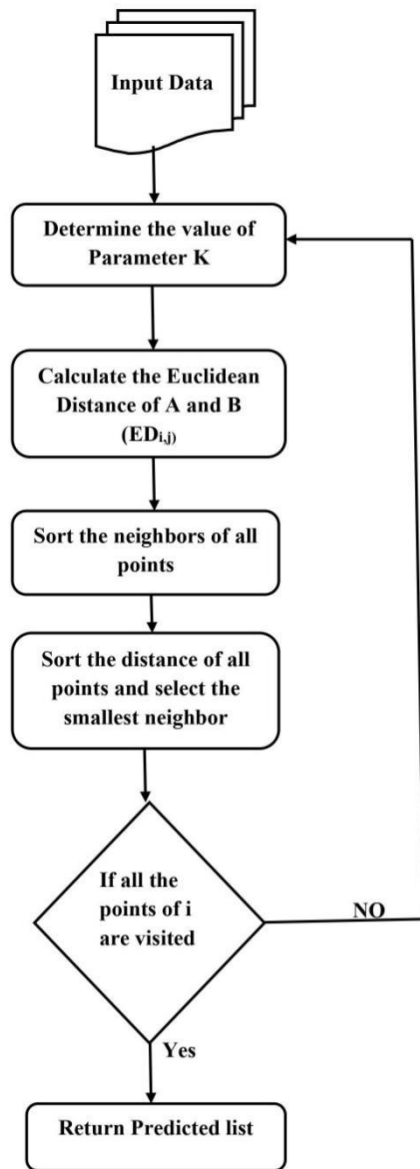


Fig 3: flow chart for proposed work

A. Quality Parameter Analysis

Item Fat Content

The code for predicting sales based on fat content is presented below. Additionally, Figure 4 illustrates that consumers purchased a greater quantity of low-fat products.

```
plt.figure(figsize=(8,5))
```

```

sns.countplot(y="Item_Fat_Content",data=train,palette='Set1',
order=train.groupby("Item_Fat_Content")["Item_Outlet_Sales"].count().sort_values().index)

plt.tight_layout()

plt.show()

```

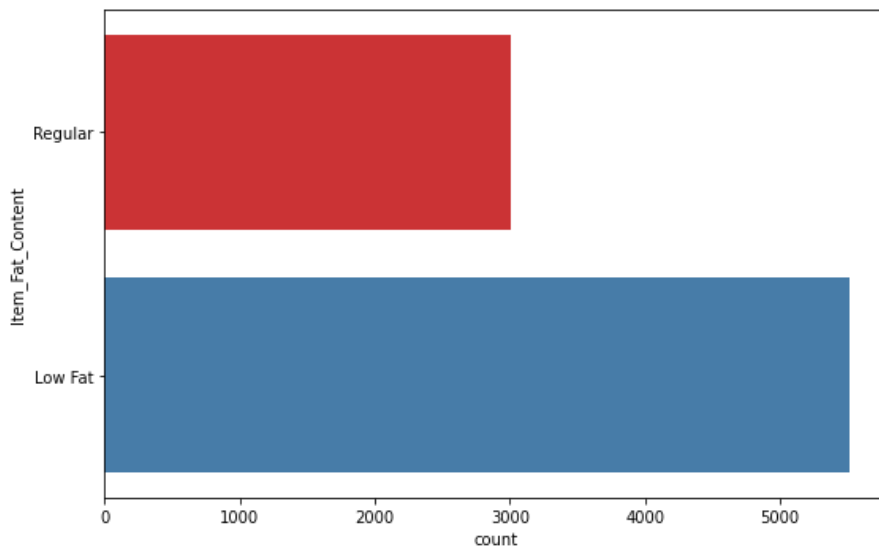


Fig 4: Item Fat content prediction

Item Type Prediction

The aforementioned code is utilized to forecast sales on the input dataset, and Figure 5 provides a lucid representation of the sales prediction across all items. It can be inferred from the data that individuals have a higher propensity to purchase fruits and vegetables in comparison to other commodities. Additionally, snack food is purchased in nearly equal amounts.

```
plt.figure(figsize=(8,5))
sns.countplot(y="Item_Type",data=train,palette='twilight',
order=train.groupby("Item_Type")["Item_Outlet_Sales"].count().sort_values().index)
plt.tight_layout()
plt.show()
```

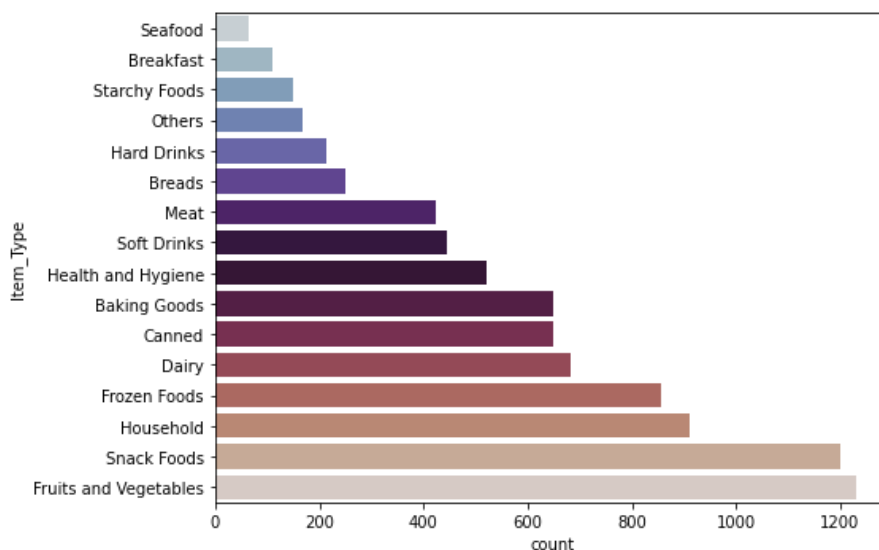


Fig 5: Sale Prediction of over all products

Item size

Figure 6 provides insight into the correlation between the size of an item and its sales performance. It can be inferred that individuals made purchases from small and medium-sized outlets. The higher count of small outlets may be attributed to the substitution of null values.

```
plt.figure(figsize=(8,5))
```

```
sns.countplot(y="Outlet_Size",data=train,palette='Set2',
order=train.groupby("Outlet_Size")["Item_Outlet_Sales"].count().sort_values().index)
plt.tight_layout()
plt.show()
```

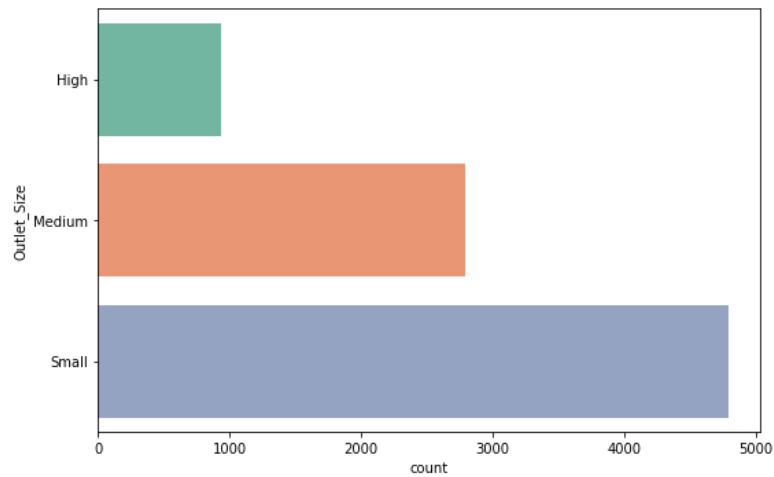



Fig 6: Sale depends on item size

Item MRP to Item Outlet Sales

Figure 7 incorporates the price of the product as a variable in the analysis, and the subsequent prediction is derived from this factor. Products with high prices are marketed at the same level as those with low prices.

```
plt.figure(figsize=(8,6))
sns.scatterplot(x="Item_MRP",y='Item_Outlet_Sales',data=train)
plt.show()
```

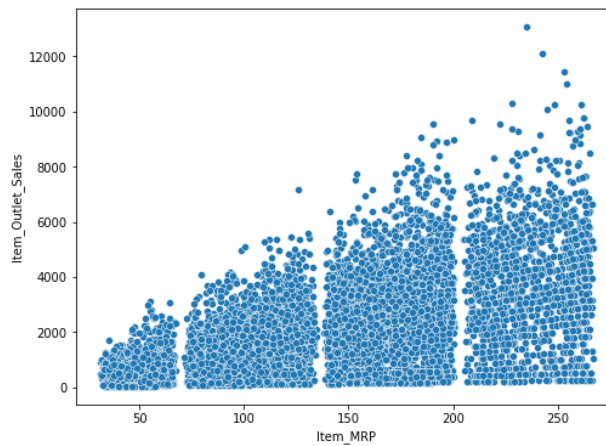


Fig 7: Sale depends on Item MRP

Confusion Matrix

A correlation matrix is a tabular representation that exhibits the correlation coefficients among numerous variables. This methodology is employed to comprehend the interconnections among variables and detect any discernible patterns or interdependencies within the

dataset. The matrix's individual cells each display the correlation between two variables.

The correlation coefficient can range from -1 to 1, and it acts as a statistical measure of the strength as well as the direction of the link that exists between two variables.

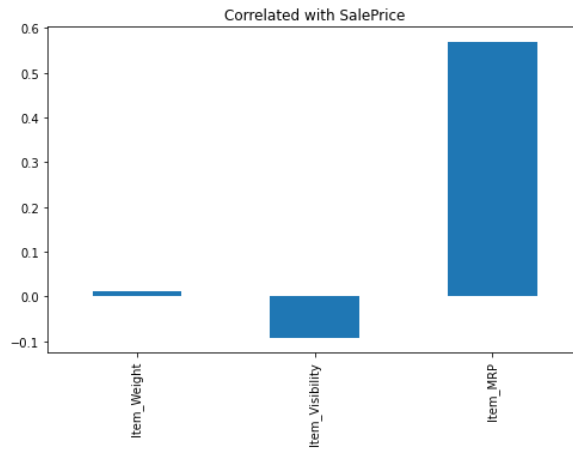


Fig 8: Correlation Matrix of K-NN on sale prediction

Figure 8 depicts the projection of the sales strategy utilizing the K-NN model, with a focus on the correlation among three key factors: item weight, visibility, and MRP.

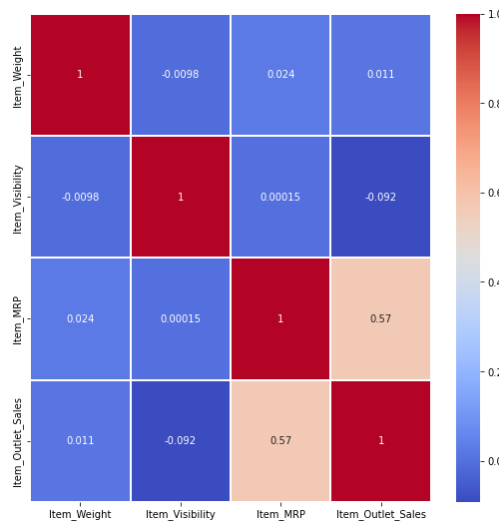


Fig 9: Confusion Matrix

It is common practice to utilize a tabular representation known as the confusion matrix in order to evaluate the accuracy of a classification model. This matrix is also known as the error matrix. The process of comparing predicted values to actual values enables the visualization of a model's performance.

Figure 9 displays the confusion matrix of the proposed model. The present study demonstrates the efficacy of the proposed methodology, yielding a noteworthy accuracy rate of 84.7%. This outcome is comparatively superior to

the results obtained by other existing approaches, as presented in Table 2. The analysis reveals that the target variable, namely Item_Outlet_Sales, exhibits a lower degree of dependence on Item_Visibility while displaying a higher degree of dependence on Item_MRP. It can be observed that there exists an inverse relationship between the Maximum Retail Price (MRP) of a product and its corresponding sales at the outlet, wherein an increase in the former leads to a decrease in the latter.

Table 2: Comparison of Performance of existing methods

Methodology	Accuracy %
Linear Regression Algorithm	56.5
Ridge Regression Algorithm	49
Lasso Regression Algorithm	54
Decision Tree Algorithm	65
Random Forest Algorithm	63
K-Nearest Neighbor algorithm	84.7

5. Conclusion

This project gives an introduction to the core ideas of ML, including the data processing and modeling methods that are pertinent to the field. The focus is on the practical application of these techniques in predicting sales for various Big Mart retail locations. The analysis demonstrates the interrelationship between various attributes and highlights that a medium-sized location achieved the highest sales. This finding implies that other purchasing locations may benefit from adopting similar strategies to enhance their sales performance. The utilization of multiple parameters and diverse factors can be employed to enhance the innovation and efficacy of sales prediction. The role of accuracy is crucial in prediction systems, and it can be substantially enhanced by increasing the parameters employed. Additionally, understanding the functioning of the sub-models can potentially enhance the efficiency of the system.

Given that profits are contingent upon accurate sales predictions, large retail establishments strive to achieve precision in their forecasting efforts to mitigate potential losses. In the current investigation, a K-Nearest Neighbor model is being developed, and it will be applied to the Big Mart 2013 dataset in order to provide projections regarding the sales of a particular outlet's goods. Empirical evidence suggests that our methodology yields superior predictive accuracy in comparison to alternative techniques such as decision trees, ridge regression, and others.

In the future, it is possible that work will be conducted using Deep Learning and Transfer Learning algorithms.

References

- [1] Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." *Journal of Soft Computing Paradigm (JSCP)* 1, no. 01 (2019): 56.- 2. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of D [3] Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 02 (2020): 101- 110.
- [2] Crime Prediction using K Nearest Neighbour Algorithm by Akash Kumar, Aniket Verma,Gandhali Shinde,YashSukhdeve,Nidhi Lal published by 2020 International Conference on Emerging Trends in Information Technology and Engineering.
- [3] House Price Prediction Using Regression Techniques: A Comparative Study by CH.Raga Madhuri, Anuradha G, M.VaniPujitha published by IEEE 6th International Conference on smart structures and systems 2019.
- [4] Deepa Rani Gopagoni, P V Lakshmi and Ankur Chaudhary, "Evaluating Machine Learning Algorithms For Marketing Data Analysis - Predicting Grocery Store Sales" <https://www.researchgate.net/publication/344508907-2019>.
- [5] Rising Odegua, "Applied Machine Learning for Supermarket Sales Prediction", [https://www.researchgate.net/publication/338681895, 2020](https://www.researchgate.net/publication/338681895-2020).
- [6] Grigorios Tsoumakas, "A survey of machine learning techniques for food sales prediction", *Artif Intell Rev*, <https://doi.org/10.1007/s10462-018-9637-z-Springer-2018>.
- [7] Yuta Kaneko and Katsutoshi Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales", *IEEE 16th International Conference on Data Mining Workshops-2016*.
- [8] Purvika Bajaj, Renesa Ray2, Shivani Shedge, Shravani Vidhate, Prof. Dr. Nikhilkumar Shardoor, "sales prediction using machine learning algorithms", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 07 Issue: 06 | June 2020.
- [9] Elcio Tarallo, Getúlio K. Akabane, Camilo I. Shimabukuro, Jose Mello, Douglas Amancio, "Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory

Research”,
<https://www.researchgate.net/publication/338172563>-2019.

- [10] Ranjitha and Spandana, “Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms”, Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021), IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2.
- [11] Melvin Tom, Nayana Raju, Asha Issac,, Jeswin James, Rani Saritha R, “Supermarket Sales Prediction Using Regression”, International Journal of Advanced Trends in Computer Science and Engineering, Volume 10, No.2, March - April 2021.
- [12] Jie Yang and Sakgasit Ramingwong, “Analysis of Sales Influencing Factors and Prediction of Sales in Supermarket based on Machine Learning Technique”, Data Science and Engineering (DSE) Record, Volume 3, issue 1.
- [13] Rajesh, P. ., & Kavitha, R. . (2023). An Imperceptible Method to Monitor Human Activity by Using Sensor Data with CNN and Bi-directional LSTM. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 96–105. <https://doi.org/10.17762/ijritcc.v11i2s.6033>
- [14] Kartika S. (2016). Analysis of “SystemC” design flow for FPGA implementation. International Journal of New Practices in Management and Engineering, 5(01), 01 - 07. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/41>
- [15] Pandey, J.K., Veeraiah, V., Talukdar, S.B., Talukdar, V., Rathod, V.M., Dhablya, D. Smart city approaches using machine learning and the IoT (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 345-362.