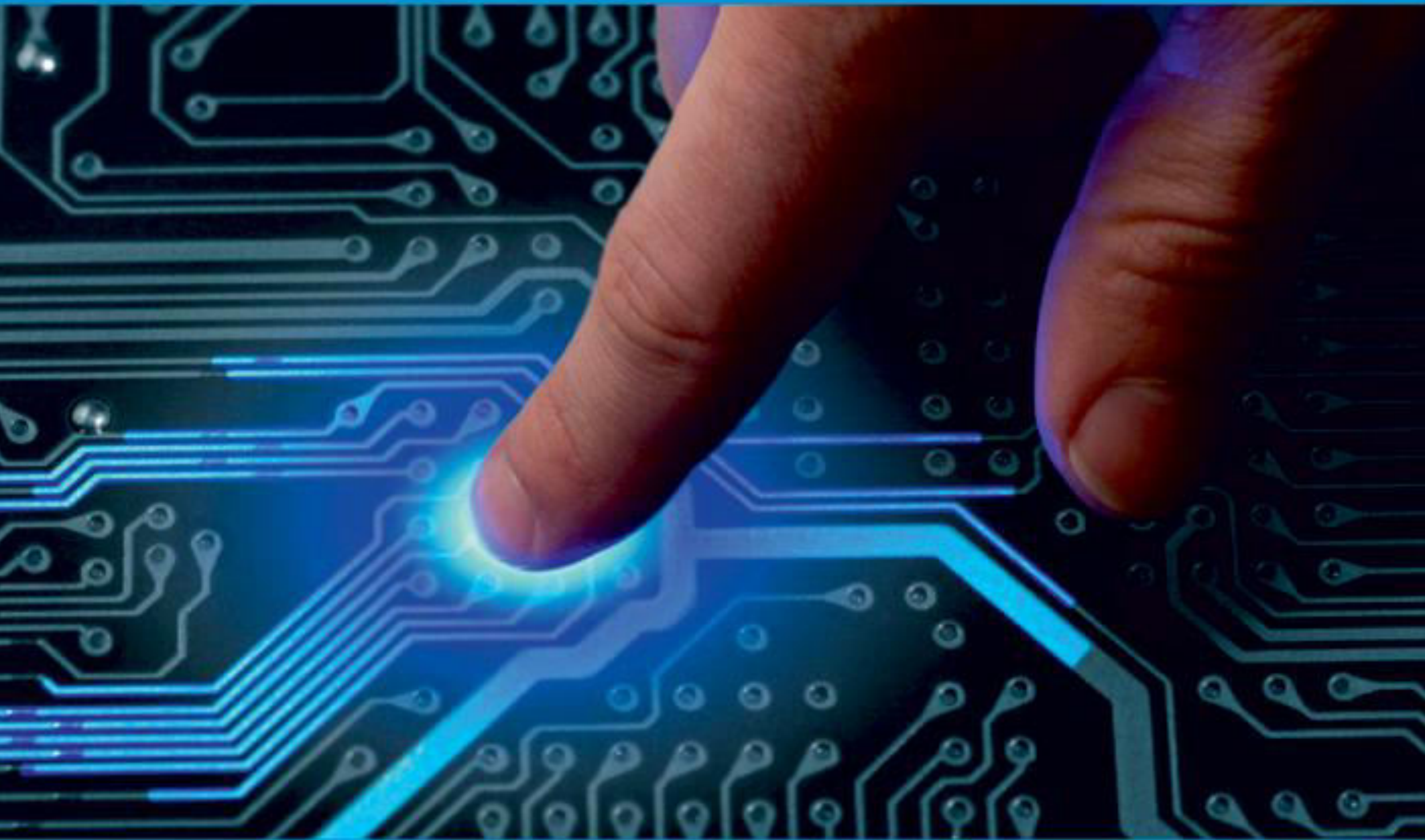
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

Machine Learning Approach for Prediction of Big Mart Sales

Afshan Rather[*1], Kulbir Kaur[2]

PG Scholar, Department of Computer Science and Engineering, Swami Vivekananda Institute of Engineering & Technology, Banur, India[1]

Assistant Professor, Department of Computer Science and Engineering, Swami Vivekananda Institute of Engineering & Technology, Banur, India[2]

[*]Corresponding Author

ABSTRACT*:* The data accessible is expanding step by step and such a gigantic measure of natural data is should have been investigated unequivocally, as it can give exceptionally useful and finely unadulterated inclination results according to current standard necessities. It is right to say likewise with the advancement of Artificial Intelligence (AI) in the course of recent many years, Machine Learning (ML) is additionally on a high speed for its development. Machine Learning is a class of calculations that permits programming applications to turn out to be more precise in foreseeing results without being expressly modified. The fundamental reason of machine learning is to fabricate models and utilize calculations that can get input data and utilize factual investigation to foresee a result while refreshing results as new data opens up. These models can be applied in various regions and prepared to match the assumptions for the board with the goal that precise advances can be taken to accomplish the association's objective. In this work, the case of Big Mart, a one-stop-shopping center, has been discussed to predict the sales of different types of items and for understanding the effects of different factors on the items' sales. Taking various aspects of a dataset collected for Big Mart, and the methodology followed for building a predictive model, results with high levels of accuracy are generated, and these observations can be employed to take decisions to improve sales.

KEYWORDS: ML, Big Mart Sales, AI, Predictive model.,

I. INTRODUCTION

Earlier companies used to produce goods without considering the number of sales and demand. For any manufacturer to determine whether to increase or decrease the production of several units, data regarding the demand for products on the market is required. Companies can face losses if they fail to consider these values while competing on the market. Diffferent companies choose specifific criteria to determine their demand and sales [1].

In today's highly competitive environment and ever-changing consumer landscape, accurate and timely forecasting of future revenue, also known as revenue forecasting, or sales forecasting, can offffer valuable insight to companies engaged in the manufacture, distribution or retail of goods[2]. Short-term forecasts primarily help with production planning and stock management, while long-term forecasts can deal with business growth and decision-making[1].

Sales forecasting is particularly important in the industries because of the limited shelf-life of many of the goods, which leads to a loss of income in both shortage and surplus situations. Too many orders lead to a shortage of products and still too few orders lead to a lack of opportunity. Therefore, competition in the food market is continuously flfluctuating due to factors such as pricing, advertisement, increasing demand from the customers[3].

Managers usually make sales predictions randomly. Professional managers, however, become hard to fifind and not always available (e.g., they can get sick or leave). Sales predictions can be assisted by computer systems that can play the qualifified managers' role when they are not available or allow them to make the right decision by providing potential sales predictions. One way of implementing such a method is to try and model the professional managers' skills inside a computer program[4].

Alternatively, the abundance of sales data and related information can be used through Machine Learning techniques to automatically develop accurate sales predictive models. This approach is much simpler. It is not prejudiced by a single sales manager's particularities and is flflexible, which means it can adapt to data changes.

It has, however, the potential to overestimate the accuracy of the prediction of a human expert, which is normally incomplete. For example, once companies used to produce the products without taking into consideration the number of sales and demand as they faced several problems. Since they don't know how much to sell, for any manufacturer to decide whether to increase or decrease the number of units, data regarding the consumer demand for products is essential. If companies do not consider these principles when competing in the market, they will face losses. Difffferent companies choose difffferent parameters to determine their market and sales.

II. RELATED WORK

Grigorios Tsoumakas had used Machine Learning techniques to perform a survey on the forecasting of food sales. They had addressed data analyst design decisions such as temporal granularity, output variable, and input variables in this survey[4]. In this paper the authors experimented by taking the point of sale (POS) as internal data and even external data by considering difffferent environments to enhance the efficiency of demand forecasting. They considered difffferent Machine Learning algorithms such as Boosted Decision Tree Regression, Bayesian Linear Regression, and Decision Forest Regression for evaluation[24]. The paper's authors had researched interestingly about customers coming to the restaurants using Random Forests, k-nearest neighbor, and XGBoost. They chose two real-world data sets from difffferent booking sites and also made difffferent input variables from restaurant features. The results have shown that XGBoost is the most appropriate model for the dataset[25]. Holmberg and Halldén had observed that regular restaurant sales to be influenced by the weather. They considered two Machine Learning algorithms as XGBoost and neural network, and the results showed that the XGBoost algorithm is more accurate than the other algorithm, and they also found that they had improved their model performance by 2-4 percentage points by taking weather factors into consideration. To improve accuracy, they had considered numerous variables such as date characteristics, sales history, and weather factors[26].

Most of the recent studies focused on sales modeling without considering the relationship between the training and testing data, they used training data directly. This causes many errors which lead to a reduction in accuracy. Recent studies have suggested clustering techniques to separate the entire forecasting data into several clusters of predictable data before designing predictable models to minimize computational time and achieve effffective evaluating performance[27]. In particular, Support Vector Machine(SVM) had been applied to demand forecasting. Garcia et al. (2012), in their study, proposed an intelligent model that relies on supporting vector machines to deal with issues relating to the allocation and revelation of new models. Kandananond (2012) showed that SVM surpassed Artificial Neural Networks in estimating demand for consumer goods[28]. Previously, most of the studies focused on considering the metrics as mean absolute error, mean squared error, median absolute error, and k-fold cross validation is used for training and testing data. Metrics like max error, accuracy, and mean absolute error are considered in this research. In this study stratified K-fold cross validation technique is used for training and testing to increase the effifffifficiency of the results. In this study a suitable algorithm is chosen for sales forecasting.

Ranjitha.P et al Currently, supermarket run-centres, Big Marts keep track of each individual item's sales data in order to anticipate potential consumer demand and update inventory management. Anomalies and general trends are often discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to forecast future sales volume using various machine learning techniques like big mart. A predictive model was developed using Xgboost, Linear regression, Polynomial regression, and Ridge regression techniques for forecasting the sales of a business such as Big-Mart, and it was discovered that the model outperforms existing models.

Big Mart is a large retail business with stores around the world. Big Mart's trends are critical because data scientists analyse them by product and area to find future locations. Data scientists can explore different patterns by shop and product to determine the most effective solutions by using a computer to forecast Big Mart sales. Many businesses rely largely on their data and demand market forecasts. Forecasting requires examining data from a variety of sources, including consumer trends, purchase patterns, and other variables. This study could also aid businesses in better

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

managing their budgets. Bhavana et al Big Marts track everything's sales data to forecast prospective customer demand and update stock management at the moment, shop run-focuses. Inconsistencies and general patterns are routinely mined from the information stockroom's information storage. The following data can be utilised to anticipate future sales volume for retailers like Big Mart utilising AI approaches like giant shop. A predictive model was constructed using Xgboost, Linear relapse, Polynomial relapse, and Ridge relapse techniques to anticipate the deals of a firm, such as Big - Mart, and it was discovered that the model beats existing models. Catchy regressions include Linear Regression, Polynomial Regression, Ridge Regression, and Xgboost Regression.

Machine learning techniques are attracting the interest of numerous stakeholders, including private-sector entities seeking the means to intelligently exploit their data to aid decision-making and enhance their competitive advantage in the market (Dod & Sharma, 2010; Krishna et al., 2017; Tsoumakas, 2019). Kolkman and Van Witteloostuijn (2019) explain that machine learning enables businesses to perform advanced predictive modelling to an extent not possible with traditional statistical techniques (Leo et al., 2019; Van Liebergen, 2017). Machine learning has been widely embraced for a variety of purposes, including financial modelling, health and safety analysis, medical diagnosis, and fraud detection (Crane-Droesch, 2017; Enkono & Suresh, 2020; Gholizadeh et al., 2018; Mohammed et al., 2016). Machine learning techniques have also been embraced for predicting market demand and consumer behaviour (Bajari et al., 2015; Sekban, 2019; Tsoumakas, 2019; Venishetty, 2019). The power of machine learning has attracted significant interest from numerous players, including business owners, data scientists, and econometricians (Bajari et al., 2015; Sekban, 2019; Venishetty, 2019).

III. PROBLEM IDENTIFICATION

In today's modern world, huge shopping centers such as big malls and marts are recording data related to sales of items or products with their various dependent or independent factors as an important step to be helpful in prediction of future demands and inventory management. The dataset built with various dependent and independent variables is a composite form of item attributes, data gathered by means of customer, and also data related to inventory management in a data warehouse. The data is thereafter refined in order to get accurate predictions and gather new as well as interesting results that shed a new light on our knowledge with respect to the task's data. This can then further be used for forecasting future sales by means of employing machine learning algorithms such as the random forests and simple or multiple linear regression model.

To find out what role certain properties of an item play and how they affect their sales by understanding Big Mart sales." In order to help BigMart achieve this goal, a predictive model can be built to find out for every store, the key factors that can increase their sales and what changes could be made to the product or store's characteristics

IV. METHDOLOGY

The steps followed in this work, right from the dataset preparation to obtaining results are represented in Fig.1

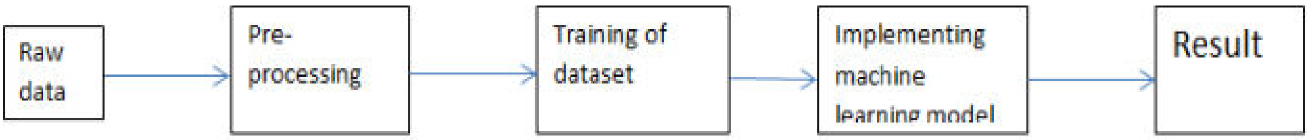


Fig. 1. Steps followed for obtaining results

DATASET AND ITS PREPROCESSING

BigMart's data scientists collected sales data of their 10 stores situated at different locations with each store having 1559 different products as per 2018 data collection.Using all the observations it is inferred what role certain properties of an item play and how they affect their sales.The dataset looks like shown in Fig.4.10on using head() function on the dataset variable.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

```
In [7]: df.head()
        #understanding rows and column
Out[7]:
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Locatic
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	

```
In [7]: df.head()
        #understanding rows and column
Out[7]:
```

at_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.1380
Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.2700
Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	Tier 3	Grocery Store	732.3800
Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052

Fig. 2. Screenshot of Dataset

```
In [9]: df.dtypes
        #tells datatype of column convert data type
Out[9]: Item_Identifier                 object
        Item_Weight                    float64
        Item_Fat_Content                object
        Item_Visibility                float64
        Item_Type                       object
        Item_MRP                       float64
        Outlet_Identifier               object
        Outlet_Establishment_Year        int64
        Outlet_Size                     object
        Outlet_Location_Type            object
        Outlet_Type                     object
        Item_Outlet_Sales              float64
        dtype: object
```

Fig. 3 Various datatypes used in the Dataset

The data set consists of various data types from integer to float to object as shown in Fig.3. In the raw data, there can be various types of underlying patterns which also gives an in-depth knowledge about subject of interest and provides insights about the problem. But caution should be observed with respect to data as itmay contain null values, or redundant values, or various types of ambiguity, which also demands for pre-processing of data. Dataset should therefore be explored as much as possible. Various factors important by statistical means like mean, standard deviation, median, count of values and maximum value etc. are shown in Fig.4 for numerical variables of our dataset.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

```
In [10]: df.describe()
Out[10]:
```

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales
count	7060.000000	8523.000000	8523.000000	8523.000000	8523.000000
mean	12.857645	0.066132	140.992782	1997.831867	2181.288914
std	4.643456	0.051598	62.275067	8.371760	1706.499616
min	4.555000	0.000000	31.290000	1985.000000	33.290000
25%	8.773750	0.026989	93.826500	1987.000000	834.247400
50%	12.600000	0.053931	143.012800	1999.000000	1794.331000
75%	16.850000	0.094585	185.643700	2004.000000	3101.296400
max	21.350000	0.328391	266.888400	2009.000000	13086.964800

Fig. 4Numerical variables of the Dataset

Preprocessing of this dataset includes doing analysis on the independent variables like checking for null values in each column and then replacing or filling them with supported appropriate data types, so that analysis and model fitting is not hindered from its way to accuracy. Shown above are some of the representations obtained by using Pandas tools which tells about variable count for numerical columns and modal values for categorical columns. Maximum and minimum values in numerical columns, along with their percentile values for median, plays an important factor in deciding which value to be chosen at priority for further exploration tasks and analysis. Data types of different columns are used further in label processing and one-hot encoding scheme during model building

ALGORITHMS

Algorithms employed Scikit-Learn can be used to track machine-learning system on wholesome basis[12].Algorithms employed for predicting sales for this dataset are discussed as follows: • Random Forest Algorithm Random forest algorithm is a very accurate algorithm to be used for predicting sales.It is easy to use and understandfor the purpose of predicting results of machine learning tasks.In sales prediction, random forest classifier is used because it has decision tree like hyperparameters.The tree model is same as decision tool. Fig.5 shows the relation between decision trees and random forest.To solve regression tasks of prediction by virtue of random forest, the sklearn.ensemble library's random forest regressor class is used. The key role is played by the parameter termed as n_estimators which also comes under random forest regressor.Random forest can be referred to as a meta-estimator used to fit upon numerous decision trees (based on classification) by taking the dataset's different subsamples.min_samples_split is taken as the minimum number when splitting an internal node if integer number of minimum samples are considered.A split's quality is measured using mse (mean squared error), which can also be termed as feature selection criterion.This also means reduction in variancemae (mean absolute error), which is another criterion for feature selection.Maximum tree depth, measured in integer terms, if equals one, then all leaves are pure or pruning for better model fitting is done forall leaves less than min_samples_splitsamples.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

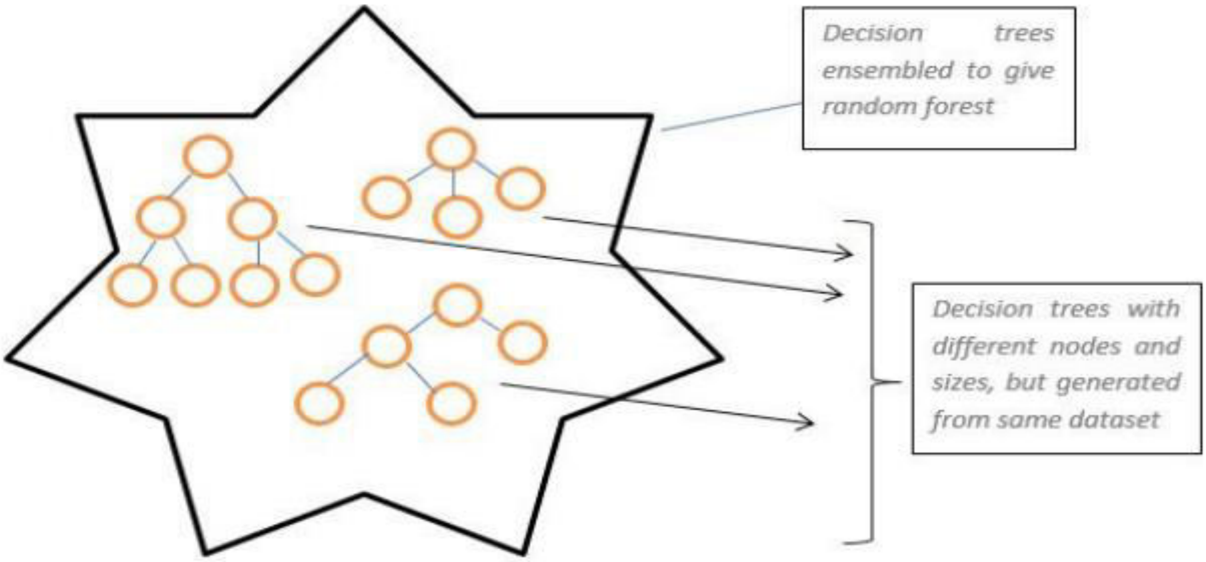


Fig. 5 Relation between Decision Trees and Random Forest

Linear Regression Algorithm Regression can be termed as a parametric technique which is used to predict a continuous or dependent variable on basis of a provided set of independent variables. This technique is said to be parametric as different assumptions are made on basis of data set.

$Y = \beta o + \beta 1X + \in$ (1) Equation shown in eq.1 is used for simple linear regression.

These parameters can be said as: Y -Variable to be predicted X -Variable(s) used for making a prediction βo -When X=0, it is termed as prediction value or can be referred to as intercept term

$\beta 1$ -when there is a change in X by 1 unit it denotes change in Y.It can also be said as slope term $\in$ -The difference between the predicted and actual values is represented by this parameter and also represents the residual value.Howeverefficiently the model is trained, tested and validated, there is always a difference between actual and predicted values which is irreducible error thus we cannot rely completely on the predicted results by the learning algorithm. Alternative methods given by Dietetic can be used for comparing learning algorithms [10].

V. RESULTS AND DISCUSSIONS

In this section, the programming language, libraries, implementation platform along with the data modeling and the observations and results obtained from it are discussed

Data Modeling and Observations Correlation is used to understand the relation between a target variable and predictors. In this work, Item-Sales is the target variable and its correlation with other variables is observed. Considering the case of Item-Weight, the feature item weight is shown to have a low correlation with the target variable Item-Outlet-Sales in Fig.6.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

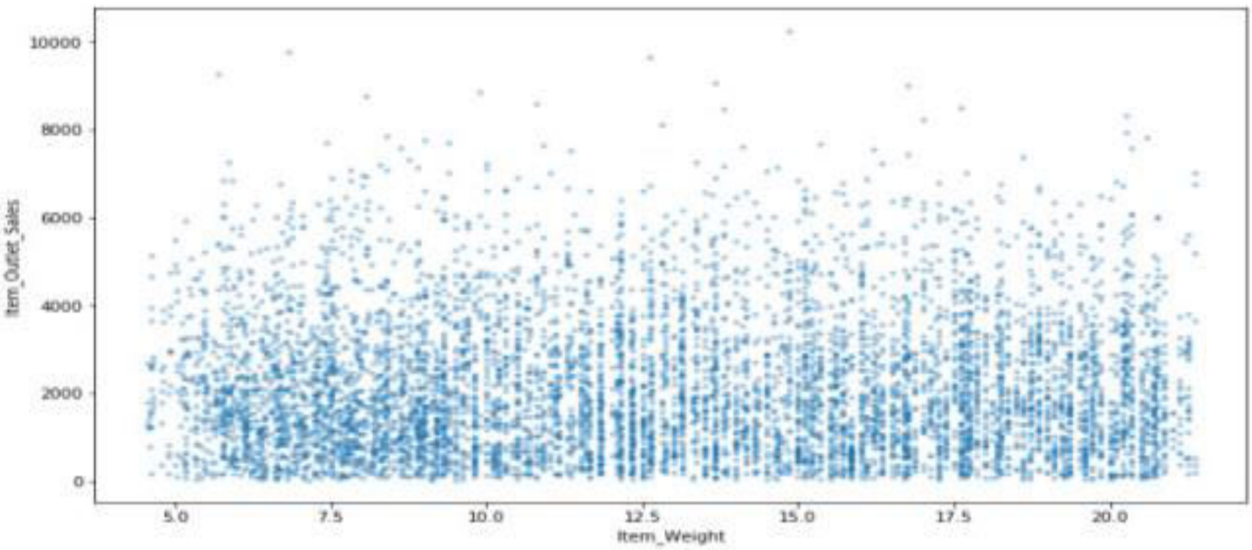


Fig. 6 Correlation between target variable and Item-weight variable

As can be seen from Fig.7 there is no significant relation found between the year of store establishment and the sales for the items. Values can also be combined into variables that classify them into periods and give meaningful results.

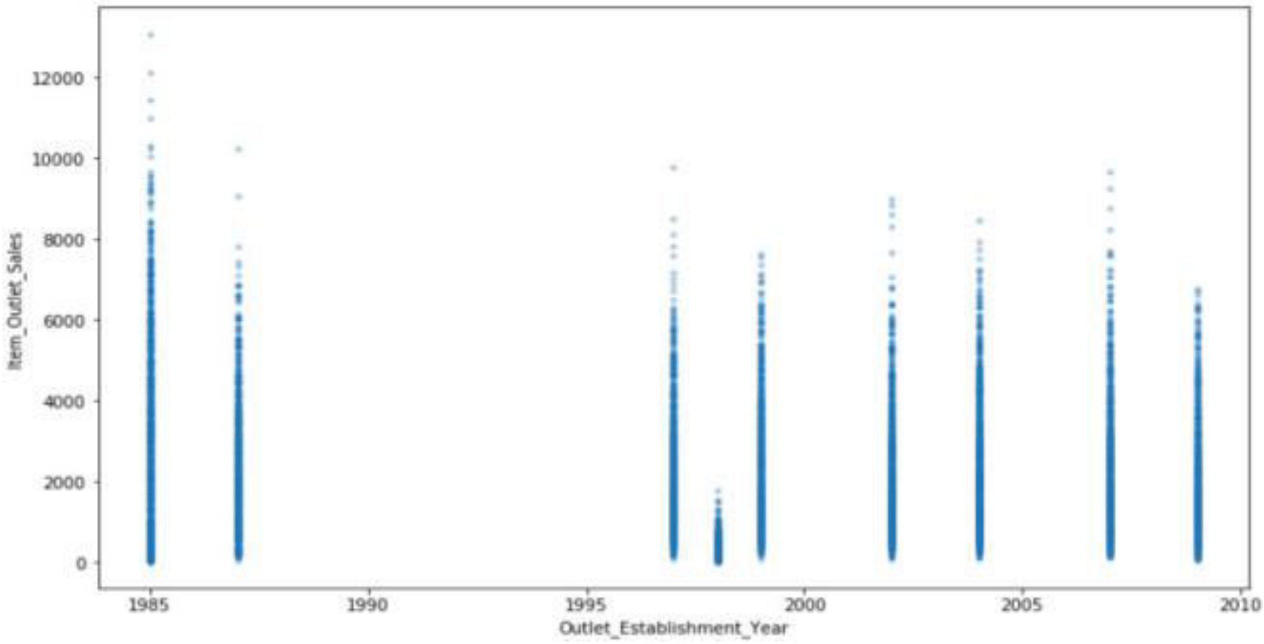


Fig. 7 Correlation between target variable and Outlet-establishment-year variable

The place where an item is placed in a store, referred to as Item_visibility, definitely affectsthe sales.However, the plot chart and correlation table generated previouslyshow that the flow is in opposite side.One of the reasons might be that daily used products don't need high visibility. However, there is an issue that some products have zero visibility, which is quite impossible. Fig.8 shows the correlation between item visibility variable and the target variable.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

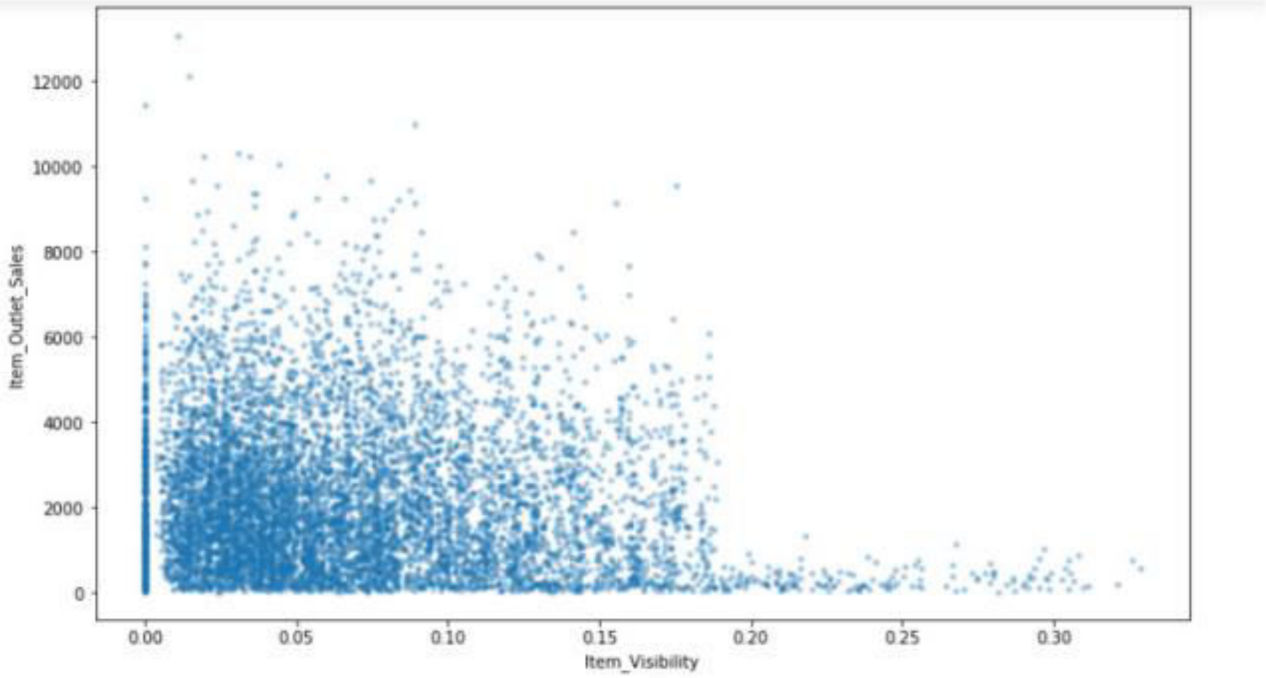


Fig. 8. Correlation between target variable and Item-visibility variable

Frequency for each categorical or nominal variable plays a significant role in further analysis of the dataset, thus supporting and collaborating in data exploration to be performed. As shown in Fig9, various variables in our dataset, with their data type and categories are shown. Here, the ID column and the source column, denoting from where the test or train sample data belongs to, are excluded and not used.

```
In [16]: ct=[x for x in data.dtypes.index if data.dtypes[x]=='object']

In [17]: ct=[x for x in ct if x not in['Item_Identifier','Outlet_Identifier','source']]

In [18]: for col in ct:
             print(data[col].value_counts())

Low Fat     8485
Regular     4824
LF           522
reg          195
low fat      178
Name: Item_Fat_Content, dtype: int64
Fruits and Vegetables    2013
Snack Foods              1989
Household                1548
Frozen Foods             1426
Dairy                    1136
Baking Goods             1086
Canned                   1084
Health and Hygiene        858
Meat                      736
Soft Drinks               726
Breads                    416
Hard Drinks               362
Others                    280
Starchy Foods             269
Breakfast                 186
Seafood                    89
Name: Item_Type, dtype: int64
Tier 3    5583
Tier 2    4641
Tier 1    3980
Name: Outlet_Location_Type, dtype: int64
Medium    4655
Small     3980
High      1553
Name: Outlet_Size, dtype: int64
Supermarket Type1    9294
Grocery Store        1805
Supermarket Type3    1559
Supermarket Type2    1546
Name: Outlet_Type, dtype: int64
```

Fig. 9. Different item categories in the dataset

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

When a predictive model generated from any supervised learning regression method is applied to the dataset, the process is said to be data scoring. The above model score clearly infers about Data Scoring.The probability of a product's sales to rise and sink can be discussed and understood on the basis of certain parameters. The vulnerabilities associated with a product or item and further its sales are also necessary and play a very important role in our problem-solving task. Further, a user authentication mechanism should be employed to avoid access from any unauthorized users and thus ensuring all results are protected and secured.

It is observed that the R-squared value is 0.563 for our dependent variable for 8523 number of observations taken under consideration. This signifies how accurately the built regression model fits.

The largest location did not produce the highest sales. The location that produced the highest sales was the OUT027 location, which was in turn a Supermarket Type3, having its size recorded as medium in our dataset. It can be said that this outlet's performance was much better than any other outlet location with any size provided in the considered dataset. The median of the target variable Item_Outlet_Sales was calculated to be 3364.95 for OUT027 location. The location with second highest median score (OUT035) hada median value of 2109.25. Adjusted R-squared and R-squared values are higher for Linear regression model than average. Therefore, the used model fits better and exhibits accuracy. Also, model accuracy and score of regression model canreach nearly 61% if built with more hypothesis consideration and analysis, as shown by code snippet in Fig.10.

```
from sklearn.ensemble import RandomForestRegressor
X_train = sd.drop(['Item_Outlet_Sales','Item_Identifier','Outlet_Identifier'],axis=1)
Y_train = sd['Item_Outlet_Sales']
X_test = ds.drop(['Item_Identifier','Outlet_Identifier'],axis=1).copy()
rf = RandomForestRegressor(n_estimators=400,max_depth=6, min_samples_leaf=100,n_jobs=4)
rf.fit(X_train,Y_train)
rf_pred = rf.predict(X_test)
rf_accuracy = round(rf.score(X_train,Y_train)*100,2)
print('accuracy of random forest is : %.4g' %rf_accuracy)

accuracy of random forest is : 60.8
```

It can be concluded that more locations should be switched or shifted to Supermarket Type3 to increase the sales of products at Big Mart. Any one-stop-shopping-center like Big Mart can benefit from this model by being able to predict its items' future sales at different locations.

VI. CONCLUSION

In this research work, basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of sales prediction in Big Mart shopping centers at different locations. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other shopping locations should follow similar patterns for improved sales. Multiple instances parameters and various factors can be used to make this sales prediction more innovative and successful. Accuracy, which plays a key role in prediction-based systems, can be significantly increased as the number of parameters used are increased. Also, a look into how the sub-models work can lead to increase in productivity of system. The project can be further collaborated in a web-based application or in any device supported with an in-built intelligence by virtue of Internet of Things (IoT), to be more feasible for use. Various stakeholders concerned with sales information can also provide more inputs to help in hypothesis generation and more instances can be taken into consideration such that more precise results that are closer to real world situations are generated. When combined with effective data mining methods and properties, the traditional means could be seen to make a higher and positive effect on the overall development of corporation's tasks on the whole. One of the main

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

highlights is more expressive regression outputs, which are more understandable bounded with some of accuracy.Moreover, the flexibility of the proposed approach can be increased with variants at a very appropriate stage of regression modelbuilding. There is a further need of experiments for proper measurements of both accuracy and resource efficiency to assess and optimize correctly.

REEFRENCES

1. Patrick Bajari, Denis Nekipelov, Stephen P Ryan, and Miaoyu Yang. Machine learning methods for demand estimation. *American Economic Review*, 105(5):481–85, 2015.
2. Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2016.
3. Ankur Jain, Manghat Nitish Menon, and Saurabh Chandra. Sales forecasting for retail chains, 2015.
4. Grigorios Tsoumakas. A survey of machine learning techniques for food sales prediction. *Artifificial Intelligence Review*, 52(1):441–447, 2019.
5. Xiaogang Su, Xin Yan, and Chih-Ling Tsai. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3):275–294, 2012.
6. Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032, 1988.
7. Zheng Li, Xianfeng Ma, and Hongliang Xin. Feature engineering of machine learning chemisorption models for catalyst design. *Catalysis today*, 280:232–238, 2017.
8. Xinchuan Zeng and Tony R Martinez. Distribution-balanced stratified cross validation for accuracy estimation. *Journal of Experimental & Theoretical Artifificial Intelligence*, 12(1):1–12, 2000.
9. Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. On the stratifification of multi-label data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 145–158. Springer, 2011.
10. Chris Rygielski, Jyun-Cheng Wang, and David C Yen. Data mining techniques for customer relationship management. *Technology in society*, 24(4):483–502, 2002.
11. Krzysztof J Cios, Witold Pedrycz, Roman W Swiniarski, and Lukasz Andrzej Kurgan. *Data mining: a knowledge discovery approach*. Springer Science & Business Media, 2007.
12. Maike Krause-Traudes, Simon Scheider, Stefan Rüping, and Harald Meßner. Spatial data mining for retail sales forecasting. In *11th AGILE International Conference on Geographic Information Science*, pages 1–11, 2008.
13. Stephen Marsland. *Machine learning: an algorithmic perspective*. CRC press, 2015.
14. ML documentation. https://www.mathworks.com/discovery/ machine-learning.html). Accessed: 2020-04-22.
15. Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
16. Arvin Wen Tsui, Yu-Hsiang Chuang, and Hao-Hua Chu. Unsupervised learning for solving rss hardware variance problem in wifufi localization. *Mobile Networks and Applications*, 14(5):677–691, 2009.
17. Bohdan M Pavlyshenko. Machine-learning models for sales time series forecasting. *Data*, 4(1):15, 2019.
18. Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, pages 19–48, 2010.
19. Sanford Weisberg. *Applied linear regression*, volume 528. John Wiley & Sons, 2005.
20. Gradient Boosting documentation. https://turi.com/learn/userguide/supervised-learning/boosted_trees_regression.html). Accessed: 2020- 05-19.
21. JN Hu, JJ Hu, HB Lin, XP Li, CL Jiang, XH Qiu, and WS Li. State-of-charge estimation for battery management system using optimized support vector machine for regression. *Journal of Power Sources*, 269:682–693, 2014.
22. Wangchao Lou, Xiaoqing Wang, Fan Chen, Yixiao Chen, Bo Jiang, and Hua Zhang. Sequence based prediction of dna-binding proteins based on hybrid feature selection using random forest and gaussian naive bayes. *PloS one*, 9(1), 2014.
23. İrem İşlek and Şule Gündüz Öğüdücü. A retail demand forecasting model based on data mining techniques. In *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, pages 55–60. IEEE, 2015.
24. Takashi Tanizaki, Tomohiro Hoshino, Takeshi Shimmura, and Takeshi Takenaka. Demand forecasting in restaurants using machine learning and statistical analysis. *Procedia CIRP*, 79:679–683, 2019.
25. Xu Ma, Yanshan Tian, Chu Luo, and Yuehui Zhang. Predicting future visitors of restaurants using big data. In *2018 International Conference on Machine*
26. *Learning and Cybernetics (ICMLC)*, volume 1, pages 269–274. IEEE, 2018.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007110 |

27. Mikael Holmberg and Pontus Halldén. Machine learning for restaurant sales forecast, 2018.
28. I-Fei Chen and Chi-Jie Lu. Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, 28(9):2633–2647, 2017.
29. Malek Sarhani and Abdellatif El Afifia. Intelligent system based support vector regression for supply chain demand forecasting. In *2014 Second World Conference on Complex Systems (WCCS)*, pages 79–83. IEEE, 2014.
30. Jason Brownlee. *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery, 2017.
31. Python history. https://en.wikipedia.org/wiki/Python_(programming_ language). Accessed: 2020-04-29.
32. Guido Van Rossum et al. Python programming language. In *USENIX annual technical conference*, volume 41, page 36, 2007.
33. Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.
34. Wes McKinney. Pandas, python data analysis library. *see http://pandas. pydata. org*, 2015.
35. Niyazi Ari and Makhamadsulton Ustazhanov. Matplotlib in python. In *2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*, pages 1–6. IEEE, 2014.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com