# BIGMART SALES PREDICTION

## Akansha Mishra, Asst. Prof. Suchi Sharma, Shiwani Singh, Ananya Srivastava, Devika Tripathi

1,3,4,5 Students, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology

2 Assistant Professor, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology

## ABSTRACT

These data stores basically contain a large number of customer data and individual item attributes in a data warehouse. Further, anomalies and frequent patterns are detected by mining the datastore from the data warehouse. In this paper, we propose a predictive model using the Xg boost technique for predicting the sales of a company like BIGMART and found that the model produces better performance as compared to existing models.

## INTRODUCTION

Every day a feeling of competition among several shopping malls as well as BIGMART is becoming very serious due to the growth of the malls and several websites that are running online. Present machine learning algorithms are very sophisticated and provide a technique to predict or forecast the future demand of sales for an organization which also helps in overcoming the cheap availability of computing and storage systems.

In this paper, we are solving the issues of BIGMART SALES PREDICTION for forecasting customers' future demand in different BIGMART across various locations and products based on the previous records. Several machine learning algorithms like linear regression analysis, random forest, etc are used for the prediction or forecasting of sales volume.

As good sales are the life of every organization another forecasting of sales is playing an important role in any shopping complex. Always a better prediction is helpful, to develop well as to improve the knowledge of the marketplace.

## PROBLEM STATEMENT

As we complete preprocessing and fill missing values, we used ensemble classifier using Decision trees, Linear regression, and Random forest. Both MAE and RSME are used as accuracy metrics for the prediction of sales in BIGMART. From the accuracy metrics, it was found that the model will predict best using minimum MAE and RSME. We have used the 2013 Sales data of BIGMART as the dataset. Where the dataset consists of 12 attributes like Item Fat, Item Type, Item MRP, Outlet Type, Item Visibility, Item Weight, Outlet Identify, Outlet Size, Outlet Establishment Year, Outlet Location Type, Item identity and Item Outlet Sales. The data set consists of 8523 products across different cities and locations. This data-set is also based on hypotheses of store level and product level. Where store level involves attributes.

# PROPOSED SYSTEM

After preprocessing and filling missing values, we used ensemble classifier using Decision trees, Linear regression, Random forest Both MAE and RSME are used as accuracy metrics for predicting the sales in BIGMART. In our work, we have used the 2013 Sales data of BIGMART as the dataset. Where the dataset consists of 12 attributes like Item Fat, Item Type, Item MRP, Outlet Type, Item Visibility, Item Weight, Outlet Identify, Outlet Size, Outlet Establishment Year, Outlet Location Type, Item identity and Item Outlet

```html
<html>
<head>
<body bgcolor=#99ffdd>
<title>itm</title></head>
<form method="POST" action="mth">
{% csrf_token %}

<h1>BigMart Sales Prediction</h1>
<table>
<tr>
<td>ItemMrp:</td>
<td><input type="number" name="itemmrp"> </br></td>
</tr>


<tr>
<td>Outletyears:</td>
<td><input type="number" name="outletyears"> </br></td>
</tr>


<tr>
<td>OUTLET TYPE:(SMT1,GROCERY,SMT2,SMT3)</td>
<td><input type="text" name="listbox1"> </br></td>
</tr>
<tr>
<td>OUTLET SIZE:(High,Medium,Low)</td>
<td><input type="text" name="os"> </br></td>
</tr>


<tr>
<td><input type="submit" value="Submit"></td>
</tr>
</table>
</form>
</body>
</html>
```

Sales. Out of these attributes, the response variable is the Item Outlet Sales attribute and the remaining attributes are used as the predictor variables.

## CODE

```html
<html>
<head>

</head>
<body bgcolor=#99ffdd>

<h1 align=left>Outlet Sales Prediction</h1>


<h2>Outlet Sales is     {{data}}</h2>




</body>
</html>
```

## CONCLUSION

In the present era of a digitally connected world, every shopping mall desires to know the customer demands beforehand to avoid the shortfall of sale items in all seasons. Day to day the companies or the malls are predicting more accurately on target variable item outlet sale Impact of item visibility on target variable item outlet sale compared to more visibility items as outlet contains daily used items which contradict the null hypothesis.

Fig. 8. Impact of item type on target variable item outlet sale.

Fig. 9. Distribution of outlet size. The number of outlet sizes is available in the dataset.

Table 1. Comparison of Cross-Validation Score of different Model Cross-Validation Score (Mean)Cross-Validation

Score(Std)

Linear Regression 1129 43.24

Decision Tree 1091 45.42

Ridge Regression 1097 43.41

Table 2. Comparison of MAE and RMSE of the proposed model with other Model MAE

RMSE

Linear Regression 836.1 1127

Decision Tree 741.6 1058

Ridge Regression 836 1129

xgboost 739.03 1052

A Comparative Study of BIGMART SALES PREDICTION.

# REFERENCE

1- www.w3schools.com
2-  www.tutorialspoint.com
3- en.wikipedia.org
4- www.javatpoint.com
5- MacKay, D. J., & Mac Kay, D. J. (2003). Information theory, inference and learning algorithms.

6- Daumé III, H. (2012). A course in machine learning. Publisher, ciml. info, 5, 69.

7- Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.