

# SALES PREDICTION

M.Usha Sri: 21B01A05A6 : CSE-B

N.Harshita: 21B01A05C1 : CSE-B

N.Santhi Priya: 21B01A05C4 : CSE-B

**SHRI VISHNU ENGINEERING  
COLLEGE FOR WOMEN**

February 16, 2024

# Table Of Content

- ➊ Introduction
- ➋ Problem Statement
- ➌ Dataset Description
- ➍ Model Building
- ➎ Evaluation Metrics
- ➏ Deployment
- ➐ Accuracy Results
- ➑ Result
- ➒ References

# Introduction

- Sales prediction is crucial for businesses as it provides foresight into future revenue, enabling effective resource allocation, inventory management, and strategic planning.
- Our project aims to develop a machine learning model for accurate sales prediction, providing businesses with valuable insights to optimize strategies and enhance overall performance.

# Problem Statement

- Develop a predictive model to forecast item outlet sales, utilizing pertinent data and features to optimize inventory management and strategic decision-making in retail.
- Formulate the problem as a regression task, with the goal of predicting item outlet sales to enable businesses to proactively adapt to market demands and enhance overall sales performance.

# Dataset Description

- The dataset includes item-specific details:
  - **Item Identifier**
  - **Item Weight**
  - **Item Fat Content**
  - **Item Visibility**
  - **Item Type**
  - **Item MRP**
- The dataset also contains outlet information:
  - **Outlet Identifier**
  - **Outlet Establishment Year**
  - **Outlet Size**
  - **Outlet Location Type**
  - **Outlet Type**
- The target variable is **Item Outlet Sales**, representing the actual sales figures of items within the respective outlets.

# Model Building

Three classification models have been implemented in the project:

- Linear Regression
- Random Forest Regressor
- Random Forest Regressor with k-fold cross-validation

The performance of the models is assessed using common classification metrics:

- Root Mean Squared Error
- R-squared score

# Deployment

- Framework : gradio
- Programming Language: Python
- Version Control : Git Hub



# Accuracy Results

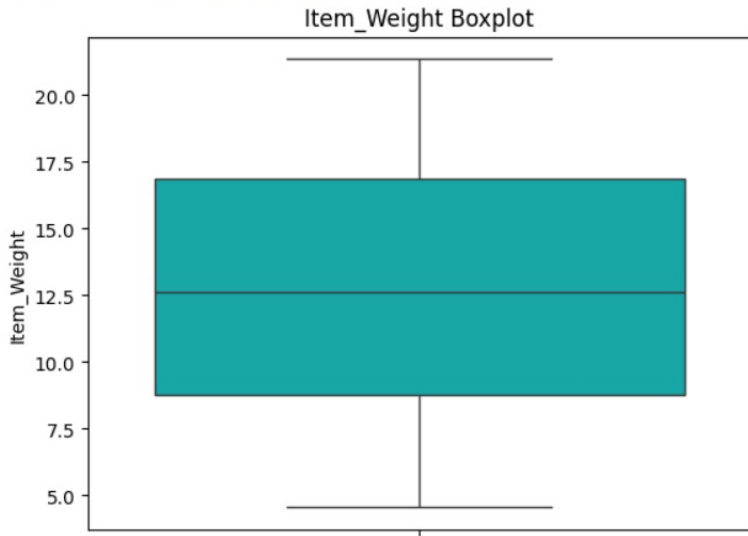
Table 1: Models and  $R^2$  Score

Model	$R^2$ Score
Linear Regression	0.524
Random Forest Regressor	0.564
Random Forest Regressor with k-fold cross-validation	0.689

- Selected Model: Random Forest Regressor with k-fold cross-validation

# Checking the Normal Distribution

```
Text(0.5, 1.0, 'Item_weight Boxplot')
```



# After Optimization Random Forest Regressor Accuracy

```
Optimized Random Forest Regressor with k-fold cross-validation:  
RMSE: 950.4019975224165  
R^2 Score: 0.6897918170898717  
Best parameters: {'max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 10, 'n_estimators': 100}
```

**Figure 1:** Accuracy Image After Optimization Random Forest Regressor.

## Sales Prediction

Item Weight

9

Item Fat Content



Low Fat



Regular

Item Visibility

0.1456

Item Type

Meat

Item MRP

450

Outlet Establishment Year

1900

- Dataset Link:
  - <https://www.kaggle.com/uniabhi/bigmart-sales-data>
- Regression Models for Sales Predictions:
  - Medium Article: <https://medium.com/mlearning-ai/sales-prediction-using-a-linear-regression-model-ffeec8>
  - HubSpot Blog: <https://blog.hubspot.com/sales/regression-analysis-to-forecast-sales>