

# PROJECT\_REPORT: Black Friday Sale Prediction

## 1. Introduction

The **Black Friday Sale Prediction** project aims to predict customer spending behavior during Black Friday sales events. This is valuable for retailers looking to optimize marketing, inventory management, and sales strategies. The project utilizes machine learning models to forecast how much a customer will spend, based on customer demographics, product categories, and other factors.

## 2. Objective

The primary objectives of this project are:

- To predict the total purchase amount by customers during the Black Friday sale.
- To analyze which customer and product attributes influence purchasing behavior.
- To evaluate the accuracy of different machine learning models using evaluation metrics like MAE, MSE, and  $R^2$  scores.

## 3. Tools and Technologies

The project uses the following tools and libraries:

- **Python:** The primary language used for data analysis and modeling.
  - Pandas:** For data manipulation and preprocessing.
  - NumPy:** For numerical computations.
  - Matplotlib** and **Seaborn:** For data visualization.
  - Scikit-learn:** For implementing machine learning models for that we import following regressions from scikit-learn:

**1.Linear Regression**

**2.Decision Tree Regressor**

**3.Random Forest Regressor**

- **Jupyter Notebook:** For interactive coding and analysis.

## 4. Methodology

1. **Data Collection:** The dataset contains customer demographic details (age, gender, occupation), product information, and purchase amounts.
2. **Data Preprocessing:**

Load the dataset using **Pandas**.

Handle missing values and encode categorical variables using one-hot encoding or label encoding.

Split the data into training and test sets to evaluate the models effectively.

### 3. Exploratory Data Analysis (EDA):

Visualizations like bar charts and histograms are used to explore the relationship between purchase amounts and demographic features like age, gender, and product categories.

Identify patterns, distributions, and potential outliers in the data.

### 4. Model Building:

**Linear Regression:** Assumes a linear relationship between the independent variables and the purchase amount.

**Decision Tree Regressor:** Splits data into branches based on feature values to predict the target variable.

**Random Forest Regressor:** An ensemble model that averages multiple decision trees to improve accuracy and reduce overfitting.

### 5. Model Evaluation: The models are evaluated using the following metrics:

**Mean Absolute Error (MAE):** Measures the average magnitude of errors in predictions without considering their direction.

**Mean Squared Error (MSE):** Measures the average squared difference between predicted and actual values, penalizing larger errors more.

**R<sup>2</sup> Score (Coefficient of Determination):** Indicates how well the independent variables explain the variance in the target variable. A higher R<sup>2</sup> score means a better fit.

## 5. Conclusion

The **Black Friday Sale Prediction** project successfully predicts customer spending using machine learning models such as Linear Regression, Decision Tree Regressor, and Random Forest Regressor. The models are evaluated using MAE, MSE, and R<sup>2</sup> scores to ensure accuracy. Typically, the **Random Forest Regressor** outperforms other models, providing more accurate predictions due to its ensemble technique, which helps in reducing overfitting and improving the generalization of predictions.