 

# Graduation Project Document

Supply chain with tabular data

**Introduced By**

# Abdelrahman Ibrahim Kamal Eldin

# Supervised By

Eng. Ahmed azab



**supply chain**

## 

## Introduction:

## This documentation explores a project that applies **data analytics** and **machine learning techniques** to a **supply chain management** problem. Supply chain management involves overseeing the flow of goods and services, including all processes that transform raw materials into final products. Efficient supply chain management is essential for reducing costs, improving delivery times, and maintaining customer satisfaction.

## Goal: This project aims to leverage data to enhance decision-making processes in supply chain operations by applying machine learning models such as Random Forest Regression and Logistic Regression for Shipment Mode Prediction

## Overview: This documentation presents two machine learning approaches—Logistic Regression and Random Forest Regression—applied to an Orders and Shipments Dataset.

* Logistic Regression is used to predict the Shipment Mode (a categorical classification problem).
* Random Forest Regression is used to predict Profit (a continuous regression problem).

These two models are commonly used in real-world business applications for predicting categorical outcomes (classification) and continuous variables (regression), respectively. This guide provides an end-to-end explanation, from loading the dataset to performing preprocessing, model training, evaluation, and visualizing the results.

### **Logistic Regression:**

* **Task**: Predict the **Shipment Mode** from order and customer details.
* **Type**: Classification (categorical outcome).
* **Algorithm**: Logistic Regression.
* **Evaluation Metrics**: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

### **Random Forest Regression:**

* **Task**: Predict the **Profit** from sales-related features.
* **Type**: Regression (continuous outcome).
* **Algorithm**: Random Forest Regressor.
* **Evaluation Metrics**: Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² score.

A black text on a white background  Description automatically generated

## Step-by-Step Guide ****classification Shipment Mode Prediction And Profit Predict****:

* **Step 1: Importing Libraries:**
* We begin by importing the required libraries for data manipulation, model training, and evaluation.

## Step 2: Loading and Preprocessing the Dataset

* + We clean column names to avoid issues related to spaces
  + **Drop missing values**: Rows where **Profit** is missing are removed.
  + **Label Encoding**: Categorical variables (like product and customer information) are encoded into numerical values using **LabelEncoder**.

## Step 3: Feature Selection and Train-Test Split

## ****classification Shipment Mode:****

* + **X**: Features that might influence shipment mode.
  + **y**: The target variable, **Shipment Mode**, to predict.
  + We split the dataset into training (70%) and testing (30%) sets

1. **Profit Predict:**
   * **X**: Features that might influence **Profit**
   * **y**: The target variable, **Profit**, for prediction
   * The dataset is split into 80% training and 20% testing.

## Step 4: Model Training and Fine-Tuning

* + **Choose the Model:** Train a language model (model name) on the preprocessed data
  + We train the **Logistic Regression model for the shipment mode** on the training data
  + We train the **Random Forest Regressor for profit prediction** with 100 trees (**n\_estimators=100**) on the training set.

## Step 5: Evaluating the Model

## We evaluate the model using **MAE**, **MSE**, and **R² score** to measure how well the model predicts the continuous values

## **Accuracy**, **Precision**, **Recall**, **F1-Score**, and the **Confusion Matrix** help evaluate the model's classification performance

## Step 6: Visualization: Actual vs Predicted Profit

* + A **scatter plot** is used to visualize the actual vs predicted profits. The red diagonal line represents perfect predictions, helping us assess the performance visually.

## A black text on a white background Description automatically generated

## Summary:

Logistic Regression was applied for the Shipment Mode Prediction, a classification problem. After training, the model’s performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix.

Random Forest Regression was used for predicting Profit, a continuous variable. The model's performance was assessed using regression metrics such as MAE, MSE, and R² score.

Both models performed reasonably well on their respective tasks, providing valuable predictions for shipment mode and profit estimation.

**Conclusion:**

This documentation showcases two different machine learning models:

Logistic Regression for classification tasks, helping businesses predict categorical outcomes like shipment modes.

Random Forest Regressor for regression tasks, making it easier to estimate continuous variables like profit.

By following this step-by-step approach, businesses can apply similar models to a wide range of real-world problems, improving decision-making processes through data-driven insights.

**Benefits of These Models:**

**Logistic Regression is:**

Simple and interpretable.

Effective for binary or multiclass classification tasks.

Computationally efficient for large datasets.

**Random Forest Regression is:**

Highly accurate due to the ensemble of multiple trees.

Resistant to overfitting.

Capable of handling non-linear relationships in data.

Both models are valuable tools in any data science toolkit, providing flexibility for tackling different types of predictive tasks.

A black text on a white background  Description automatically generated