Walmart's Inventory Management and Demand Forecasting

Prepared by :

Faris Arief Mawardi

Michael Nathaniel

Nadia Nabilla Shafira

Noufal Rifata Reyhan

Introduction

Walmart's Inventory Management and Demand Forecasting

Our project focuses on precise sales forecasting to enhance store efficiency and customer satisfaction. Key correlated factors influencing Walmart retail sales include **Fuel Price**, **Consumer Price Index (CPI)**, **Store Size**, **Store Types**, **Environmental Temperatures (Season)**, and **Price Markdowns.**

Utilizing **Random Forest Regression**, we achieved robust performance with a **0.9774 Data Train Evaluation** and **0.9389 Data Test Evaluation** using the **R2 Score Matrix**, making it the optimal model for sales forecasting in Walmart retail.

Source : https://petapixel.com/assets/uploads/2020/05/walmart.jpg

Background

"Determine what your customers need, and work backwards"
- Jeff Bezos

[1]

[2]

[3]

[4]

Importance of inventory management ➡ impacting the profitability of retail enterprises. [5]

Source :
1. https://thumbs.dreamstime.com/b/supermarket-isometric-concept-footage-people-available-k-uhd-fullhd-hd-d-video-animation-87609130.jpg
2. https://th.bing.com/th/id/OIP.fQgYRueFiBUlCbHxtpU8hQHaEK?w=326&h=183&c=7&r=0&o=5&dpr=1.3&pid=1.7
3. https://th.bing.com/th/id/OIP.Nw69UK2E0SPqk6Jas3leVgHaD4?w=302&h=180&c=7&r=0&o=5&dpr=1.3&pid=1.7
4. https://th.bing.com/th/id/OIP.CxKeD-6Fjl8xWzyDsLW8YQAAAA?w=198&h=172&c=7&r=0&o=5&dpr=1.3&pid=1.7
5. Khurshid Ali & Numaira Showkat (2022). Impact of Inventory Management on Operating Profits: Evidence from India. Journal of Finance and Economics. DOI: 10.12691/jfe-10-2-3.

Problem Statement

Problem Statement :

Enhancing the accuracy in sales forecasting at Walmart stores to increase operational efficiency and customer satisfaction.

Objective :

Built an accurate sales forecasting model to optimize inventory levels, improve store operational efficiency and enhance customer satisfaction.

Project Timeline

HACKTIV8

Discovering Business Objectives

Explore retail business fundamentals to analyze methods for boosting profitability and establishing project objectives and methods.

Phase 2

Dec 2nd of 2023

Data Exploration and Model Development

Analyzing the datasets employed in this project to extract insights and proceed to the development process of the sales forecasting model.

Phase 4

Dec 7th of 2023

Phase 1

Nov 30th - Dec 1st of 2023

Data Extraction

Extracting raw datas for subsequent cleaning to facilitate the project's progression
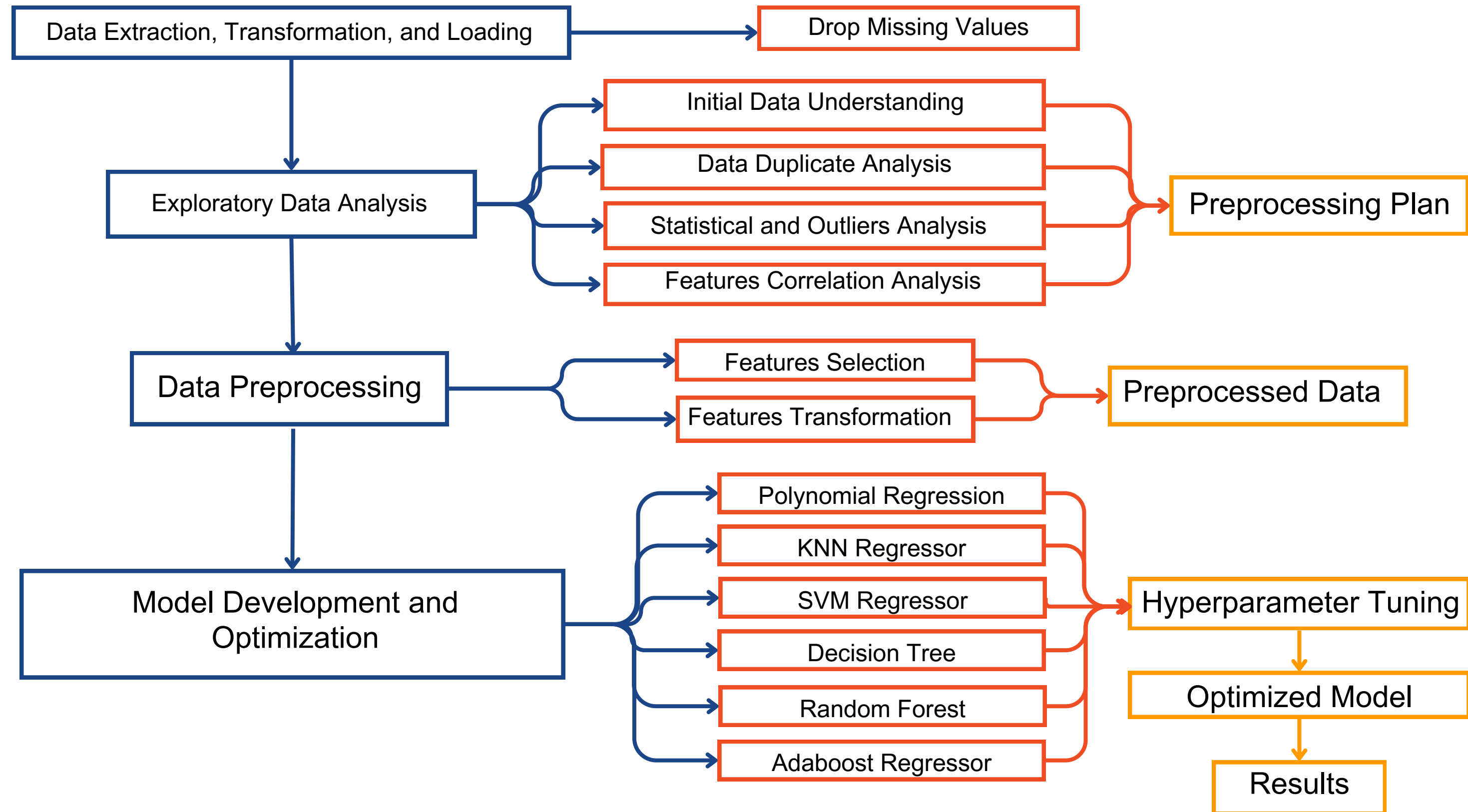
Phase 3

Dec 3rd - Dec 6th of 2023

Interpreting The Results and Generating Reports

Analyzing the findings derived from the project and crafting comprehensive reports.

Methods

The Dataset

Source: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data

WALMART · RECRUITMENT PREDICTION COMPETITION · 10 YEARS AGO

Late Submission •••

Walmart Recruiting - Store Sales Forecasting

Use historical markdown data to predict store sales

Overview   Data   Code   Models   Discussion   Leaderboard   Rules   Team   Submissions

Dataset Description

You are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains a number of departments, and you are tasked with predicting the department-wide sales for each store.

In addition, Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this competition is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data.

stores.csv

This file contains anonymized information about the 45 stores, indicating the type and size of store.

train.csv

This is the historical training data, which covers to 2010-02-05 to 2012-11-01. Within this file you will find the following fields:

Files
5 files

Size
3.22 MB

Type
zip, csv

License
Subject to Competition Rules

sampleSubmission.csv
- 115.064 Data Entries
- 2 Columns

features.csv*
- 8.190 Data Entries
- 12 Columns

stores.csv*
- 45 Data Entries
- 3 Columns

train.csv*
- 421.570 Data Entries
- 5Columns

test.csv
- 115.064 Data Entries
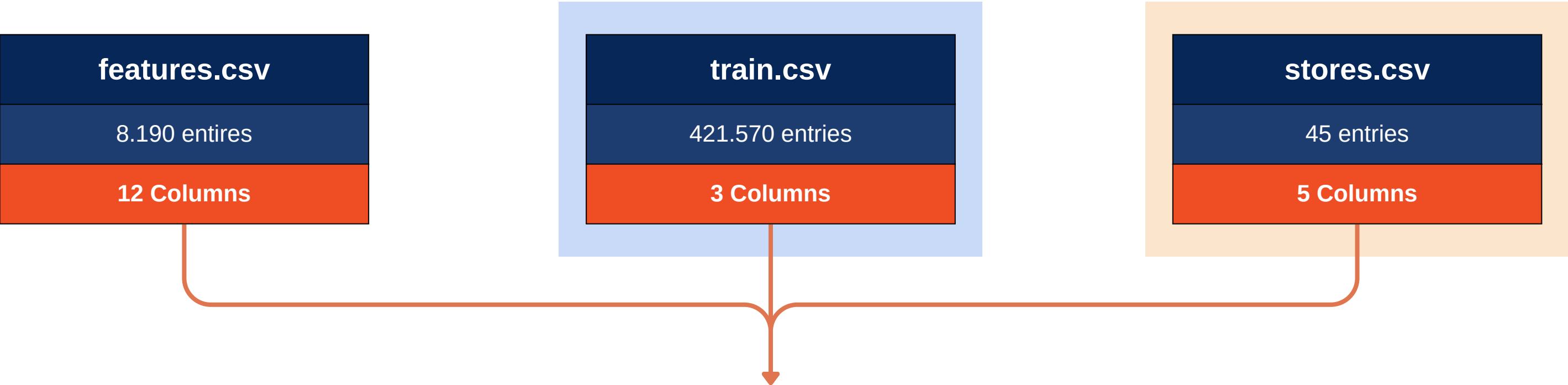- 4Columns

The (.csv*) is the **main CSV file.**

Data Engineering

features.csv

8.190 entires   12 columns

Column Name	Non-Null Count	Data Type
Store	8.190	int64
Date	8.190	object
Temperature	8.190	float64
Fuel_Price	8.190	float64
Markdown1	4.032	float64
Markdown2	2.921	float64
Markdown3	3.613	float64
Markdown4	3.464	float64
Markdown5	4.050	float64
CPI	7.605	float64
Unemployment	7.605	float64
IsHoliday	8.190	bool

train.csv

421.570 entries   5 columns

Column Name	Non-Null Count	Data Type
Store	421.570	int64
Dept	421.570	int64
Date	421.570	object
Weekly_Sales	421.570	float64
IsHoliday	421.570	bool

stores.csv

45 entries   3 columns

Column Name	Non-Null Count	Data Type
Store	45	int64
Type	45	object
Size	45	int64

8

Data Engineering

features.csv
8.190 entires
12 Columns

train.csv
421.570 entries
3 Columns

stores.csv
45 entries
5 Columns

"Merging **all three different dataset into one new dataset**."

Walmart's Inventory Management and Demand Forecasting.csv			
423.325 entires			
16 Columns			
Store	Date	Temperature	Fuel_Price
Markdown1	Markdown2	Markdown3	Markdown4
Markdown5	CPI	Unemployment	IsHoliday
Type	Size	Dept	Weekly_Sales

"The merging process is done with **pd.merge** method."

Data Engineering

Duplicated

"There is no need for duplicated data handling since **no single evidence showing our dataset has duplicated data.**"

"We are using...

data.duplicated().sum()

...method to check if there is any duplicated data in our dataset."

Missing Values

	Missing Values	Missing %
Markdown1	270892	63.991496
Markdown2	310793	73.417115
Markdown3	284667	67.245497
Markdown4	286859	67.763302
Markdown5	270138	63.813382
CPI	585	0.138192
Unemployment	585	0.138192
Dept	1755	0.414575
Weekly_Sales	1755	0.414575

"Removed missing values in certain columns **based on the absence of markdown data before November 2011** and **the absence of CPI, Unemployment, Dept, and Weekly_Sales data after October 2013**."

"We are using...

data.dropna()

...method to drop data containing missing values."

Results

Seasonal Sales Pattern

- **Peak Sales Periods**: Notable sales spikes occur during specific periods like **Year-end** (November–December), **Winter-to-Spring transition** (February–March), and **Summer months** (June–August).

- **Sales Dips**: Declines are observed during the **Spring season** (March–May) and the transition from **Summer to Fall** (August–October).

Total Sales per Month
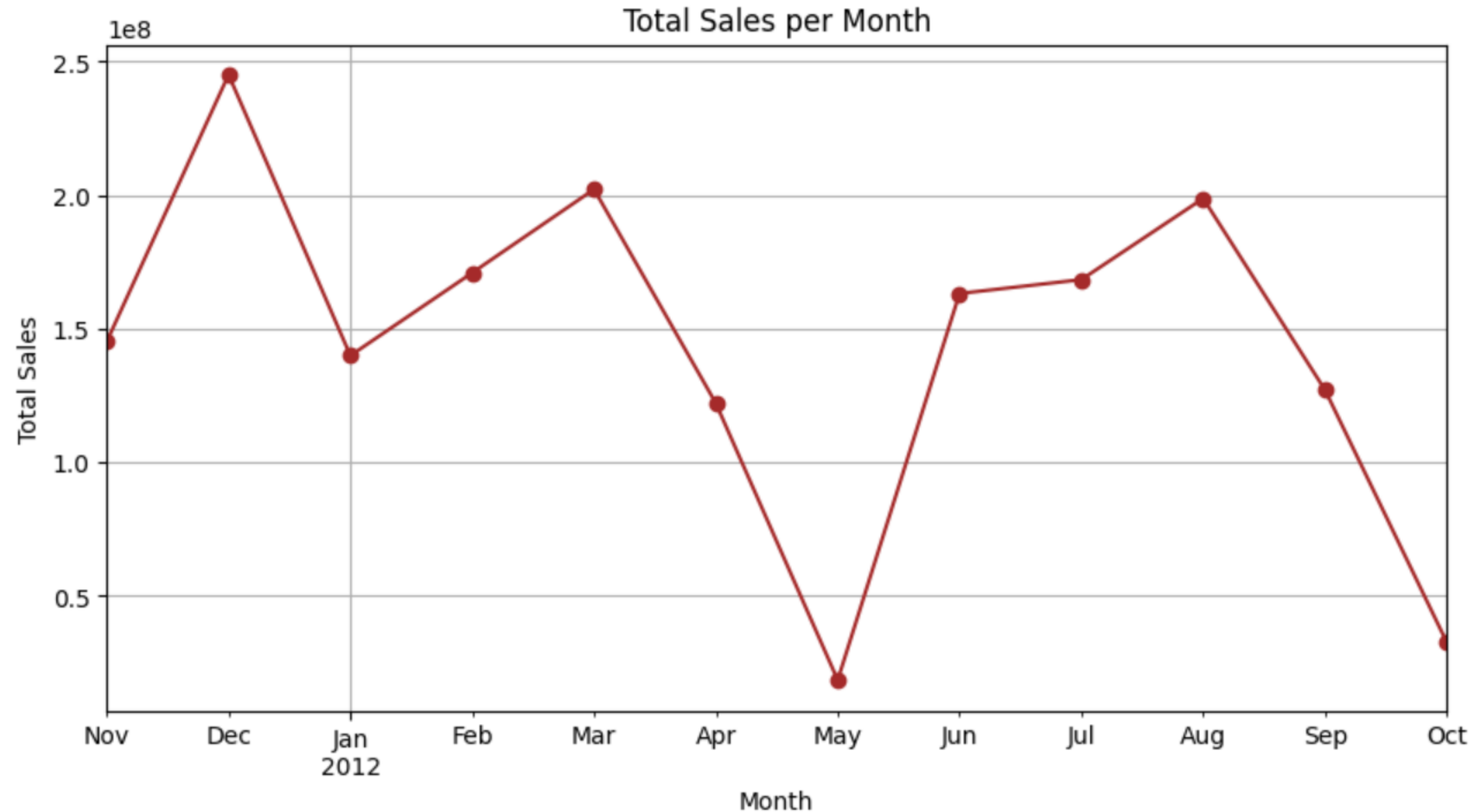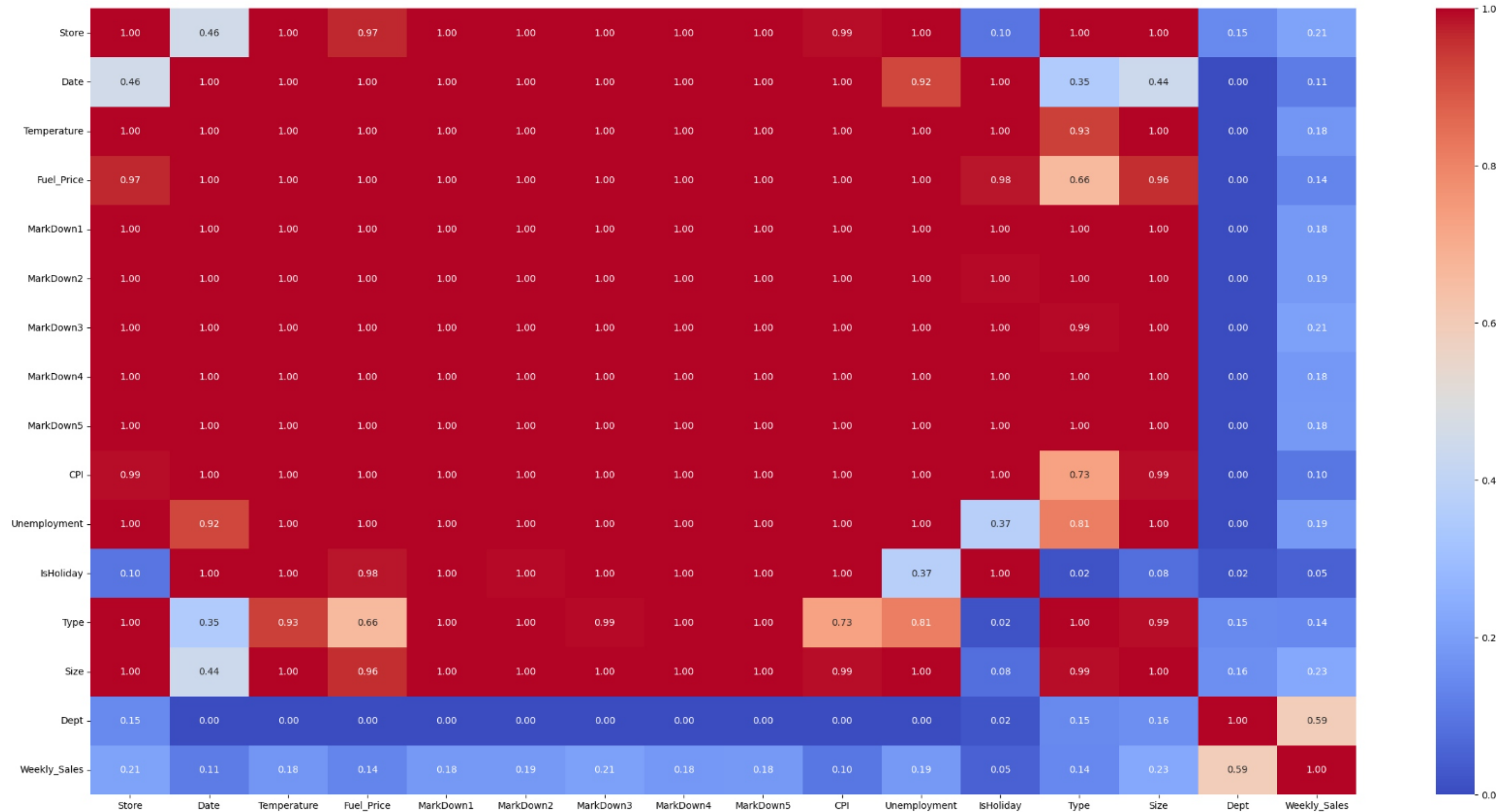
Impactful Features for Sales Prediction

Fig.5. Phik Correlations Analysis of Dataset's Features Related to The Target Variable (Weekly Sales)

Model Performance

Random Forest and **Decision Tree** showcased a significant dominance through Cross Validation utilizing the MAE & R2 Metrics for Model Evaluation.

	Mean of MAE - CV	Mean of R2 - CV
Polynomial Regression	13974.41	20.79%
KNN Regressor	7677.44	65.42%
SVM Regressor	14253.42	-10.66%
Decision Tree Regressor	**3300.41**	**87.93%**
Random Forest Regressor	**2520.77**	**93.75%**
Adaboost Regressor	24853.27	-43.65%

Fine Tuned Models Evaluation

- The **Random Forest Regressor** outperforms the Decision Tree Regressor in terms of both evaluation matrices

- A **minor overfitting** tendency is observed post hyperparameter tuning, but it remains within acceptable limits.

- **Continuous monitoring** and **adjustments** are recommended to mitigate potential escalating concerns.

	Decision Tree	Random Forest
MAE Train	4602.43	**1406.80**
MAE Test	4955.72	**2488.92**
MAE Deviation	**353.29**	1082.12
R-squared Train	89.28%	**97.74%**
R-squared Test	83.30%	**93.89%**
R-squared Deviation	5.97%	**3.85%**

Challenges

Limitations within the dataset columns, particularly concerning markdowns 1-5 and department demonstrate diverse impacts on sales interpretations

1. **Markdowns**
   a. Markdowns 1, 2, 4, and 5 display a trend where lower markdown prices correlate with increased weekly sales.
   b. Markdown 3 shows that higher markdown values result in increased sales.
   c. These disparities suggest varying promotional strategies for different product categories.
2. **Dept (values : 1-98)**
   a. The wide range of department values (1-98) indicates a diverse array of products with distinct sales patterns.
   b. Our analysis highlights Department 72 as the highest sales contributor, implying a potential for high-demand or best-selling products.

A deeper understanding of the specific products within each department and product categories within each markdowns **would provide more nuanced insights into these trends**.

Success Story

Robust Model Evaluation Results

a. Exceptional Training Performances
- The random forest model showcases exceptional proficiency on the training dataset boasting an impressive R-squared score of 97.74%. This remarkable performance signifies the model's exceptional ability to understand and learn from historical data, enabling it to predict inventory outcomes accurately

b. Generalization to New Data
- The success is not limited to the training dataset. Achieving an R-squared score of 93.89% when confronted with unseen data test, the model maintains its commendable performance indicating a robust capacity to generalize well to new data, and ensuring reliable predictions beyond the training dataset.

Conclusions

1. **Impactful Features for Sales Prediction:**
   - 'Dept'   - 'Fuel_Price'
   - 'Size'   - 'MarkDown 1-5'
   - 'Store'   - 'Unemployment'
   - 'Date'   - 'Type'
   - 'Temperature'

2. **Seasonal Sales Patterns:**
   - **Peak Sales Periods:** Notable sales spikes occur during specific periods like Year-end (November-December), Winter-to-Spring transition (February-March), and Summer months (June-August).
   - **Sales Dips:** Declines are observed during the Spring season (March-May) and the transition from Summer to Fall (August-October).

3. **Model Performance Insights:**
   - The Random Forest exhibits an R-squared score of 97.74% on data train evaluation and 93.89% on data test evaluation showcasing its **exceptional ability to predict inventory outcomes accurately**

Business Insight

Unlocking Operational Excellence

The implementation of sophisticated predictive models, particularly the Random Forest algorithm, could possibly revolutionized the inventory management strategy, **leading to significant advancements and impactful insights.**

The Model can also helps on forecasting sales which enables inventory levels optimization, improves operational efficiency, and optimizing profitability of retail enterprises

Thank You

Faris Arief Mawardi
Data Analyst

0812 1932 6001
farismawardi96@gmail.com

Michael Nathaniel
Data Engineer

0812 9491 9185
michael.nathaniel20@gmail.com

Nadia Nabilla Shafira
Data Scientist

0852 2232 3200
nadnabilla@gmail.com

Noufal Rifata Reyhan
Data Scientist

0812 1216 4554
noufalreyhan1@gmail.com

Hacktiv8 Headquarter

Menara BCA Lt. 8-10, Jakarta

+62 21 629 93 64
www.hacktiv8.com