Project Report Format

INTRODUCTION

1.1 Project Overview:

The "Walmart Sales Analysis for Retail Industry with Machine Learning" project represents a strategic initiative aimed at unraveling the intricate dynamics of sales performance within Walmart's network of 45 stores. Focused on key holiday periods such as Super Bowl, Labor Day, Thanksgiving, and Christmas, the project employs advanced machine learning algorithms to conduct in-depth sales data analysis. The goal is to discern patterns and correlations that influence consumer behavior during holidays, providing a foundation for precise sales forecasts.

1.2 Purpose:

The primary purpose of this project is to enhance Walmart's understanding of the impact of holidays on sales and to develop accurate sales forecasts through the application of advanced data-driven techniques. By employing machine learning algorithms such as Random Forest, Decision Tree, XGBoost, and ARIMA, the project aims to go beyond traditional analytics, offering unique insights into holiday-specific sales data. This precision in forecasting is not only a technical advancement but also serves a broader purpose by potentially reducing waste and improving customer satisfaction through optimized pricing and promotions. Additionally, the project aligns with Walmart's business objectives, focusing on improving decision-making and sales optimization to potentially increase revenue for the retail giant.

2.1 Existing Problem:

The retail industry, despite its dynamic nature, often grapples with the challenge of accurately predicting sales performance during critical holiday periods. Walmart, with its expansive network of 45 stores, faces the need for a deeper understanding of the impact of holidays, such as Super Bowl, Labor Day, Thanksgiving, and Christmas, on its sales. The existing problem lies in the limitations of conventional analytics to provide precise insights into the intricate patterns of consumer behavior during these peak seasons. The absence of a sophisticated, data-driven approach hinders the ability to optimize sales strategies effectively, potentially leading to suboptimal decision-making, increased waste, and missed revenue opportunities. The "Walmart Sales Analysis for Retail Industry with Machine Learning" project seeks to address and overcome these challenges by leveraging advanced machine learning algorithms for a more nuanced and accurate analysis of holiday-specific sales data.

2.2 References:

The approach and methodologies employed in the "Walmart Sales Analysis for Retail Industry with Machine Learning" project draw inspiration from established research and literature in the fields of machine learning, data science, and retail analytics. Key references guiding the project include seminal works on machine learning algorithms such as "Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman, and "Forecasting: Principles and Practice" by Rob J Hyndman and George Athanasopoulos, which

provide a solid foundation in the application of advanced techniques for predictive modeling and time series analysis.

Additionally, industry-specific research on retail analytics and consumer behavior during holidays, sourced from reputable journals and publications, informs the project's methodology. Notable references include articles from the Journal of Retailing and the International Journal of Research in Marketing, among others, contributing to the project's understanding of the broader context and challenges within the retail sector.

These references serve as a guidepost for the project team, ensuring that the methodologies applied are grounded in established principles and best practices, while also pushing the boundaries of innovation in the pursuit of a more accurate and insightful analysis of Walmart's sales data.

2.3 Problem Statement Definition:

The retail industry, characterized by its inherent volatility, presents a challenge for accurate sales predictions, especially during crucial holiday periods. The existing problem revolves around the limitations of traditional analytics in comprehensively understanding the intricate impact of holidays, such as Super Bowl, Labor Day, Thanksgiving, and Christmas, on the sales performance of Walmart's 45 stores. Conventional methods often fall short in uncovering the nuanced patterns and influences on consumer behavior during these peak seasons, leading to suboptimal sales forecasts. This gap in precision results in challenges for effective sales optimization, potentially contributing to increased waste and missed revenue opportunities. The "Walmart Sales Analysis for Retail Industry with Machine Learning" project aims to address this problem by employing advanced machine learning algorithms to conduct a thorough analysis of holiday-specific sales data, ultimately providing precise and actionable insights for enhanced decision-making and strategic sales optimization.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

https://drive.google.com/drive/folders/17-gAmCon7dVwN2vAmqCA__kpLFJuJmw8?usp=sharing

3.2 Ideation & Brainstorming

https://drive.google.com/drive/folders/17-gAmCon7dVwN2vAmqCA__kpLFJuJmw8?usp=sharing

4. Requirement Analysis

4.1 Functional Requirements:

1. Data Collection and Integration:

- Collect and integrate historical sales data from Walmart's 45 stores, focusing on holiday periods.
- Integrate external data sources, such as holiday calendars and economic indicators, for a comprehensive analysis.

2. Algorithmic Analysis:

- Implement machine learning algorithms, including Random Forest, Decision Tree, XGBoost, and ARIMA, for accurate sales forecasting.
- Continuously update and refine algorithms based on new data for ongoing optimization.

3. Real-time Dashboard:

- Develop a user-friendly dashboard for decision-makers to access real-time insights during holiday periods.
- Include visualizations and key performance indicators for quick and informed decisionmaking.

4. Customer Segmentation:

- Implement algorithms for customer segmentation based on past holiday behavior.
- Customize promotions and discounts for different customer segments to enhance engagement.

5. Predictive Inventory Management:

- Develop a system for predicting inventory needs during holidays, optimizing stock levels, and minimizing waste.
- Integrate inventory data with sales forecasts for proactive stock management.

6.Collaboration with Suppliers:

- Establish a system for collaborative information sharing with suppliers to enhance the supply chain during peak seasons.
- Implement automated communication channels for seamless coordination.

4.2 Non-Functional Requirements:

1. Scalability:

- The system should be scalable to accommodate the growing volume of data and increasing store locations.

2. Performance:

- Ensure timely processing and analysis of data to provide real-time insights, meeting the demands of decision-makers.

3. Security:

- Implement robust security measures to protect sensitive sales and customer data.
- Ensure compliance with data protection regulations.

4. Usability:

- The user interface should be intuitive, requiring minimal training for decision-makers and store managers.

5. Reliability:
- The system should be highly reliable, with minimal downtime during critical holiday periods.
- Implement automated backup and recovery mechanisms.

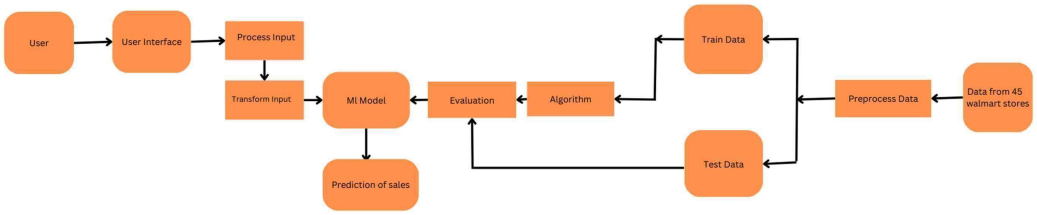
6. Accuracy:
- Algorithms should consistently provide accurate sales forecasts with a low margin of error.
- Regularly validate and refine algorithms to maintain accuracy.

7. Adaptability:
- The system should be adaptable to changes in consumer behavior, economic conditions, and other external factors.
- Implement mechanisms for easy integration with new data sources and algorithms.

These functional and non-functional requirements lay the foundation for developing a robust and effective solution for the Walmart Sales Analysis project, ensuring that it not only meets the current needs but is also adaptable and scalable for future demands in the dynamic retail landscape.

Data Flow Diagrams:

The data flow initiates with data collection from 45 Walmart stores. The data is then preprocessed to clean, transform, and aggregate it. Subsequently, machine learning models (including Random Forest, Decision Tree, XGBoost, and ARIMA) are applied for sales forecasting.Results of the analysis are integrated into a Flask web application hosted on Local server, facilitating user-friendly data visualization and insights. Endusers, encompassing data analysts, business strategists, and store managers, access the application through web browsers, allowing them to make data-driven decisions for improved sales optimization.


User
User Interface
Process Input
Transform Input
ML Model
Prediction of sales
Evaluation
Algorithm
Train Data
Test Data
Preprocess Data
Data from 45 walmart stores

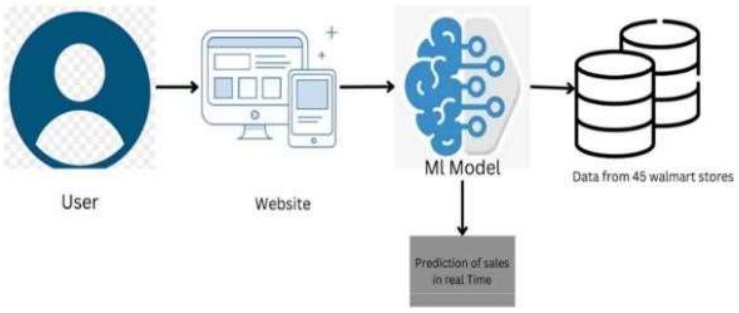
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority
Store Manager	User Interface	USN-1	As a Store Manager, I want user friendly interface	I can navigate easily	High
		USN-2	As a Store Manager, I want expected number of fields to give input	Expected number of fields to give input	High
Regional Manager	Comparative Analysis	USN-3	As a Regional Manager, I view and compare sales across multiple stores	I can view and compare sales across multiple stores through dashboard	High
	Holiday Impact Analysis	USN-4	As a Store manager, I want to access the insights to understand the impact of holidays on sales.	I can access sales analysis and forecasting insights for holidays through dashboard	High
Business Strategist	Trend Analysis	USN-5	As a Strategist, I want to identify patterns and trends in historical sales data for strategic decision making	I can identify trends for decision making	High
Store Manager	Recommendations	USN-6	As a store manager, I want recommendations for improved decision making.	I can access recommendations for sales optimization through dashboard	High
Data Analyst	Prediction	USN-7	As a Data Analyst, I want to forecast future sales trends accurately	Deploy accurate model for forecasting future sales	High
		USN-8	As a Data Analyst, I want clear display of prediction	I can see prediction clearly	High

Solution Architecture:

Data from 45 Walmart stores undergoes rigorous preprocessing which includes Data cleaning, transforming, and aggregating it for analysis. Machine learning models, including Random Forest, Decision Tree, XGBoost, and ARIMA, are applied to the preprocessed data and development of highly accurate sales forecasting models.

To provide accessible insights and recommendations, the results are seamlessly integrated into a user-friendly Flask web application hosted on local system.

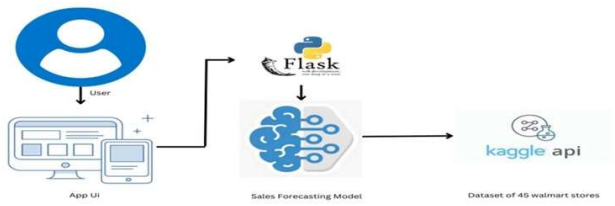
End-users, including data analysts, business strategists, and store managers, access the Flask application via web browsers. This direct interaction allows them to make informed, data driven decisions, enhancing sales optimization and overall decision making processes.



Certainly! Let's break down the information for project planning and scheduling:

6.1 Technical Architecture:

The technical architecture for the "Walmart Sales Analysis for Retail Industry with Machine Learning" project involves the following key components:



User Interface: Collection of input data from like Walmart store Managers,Buisness Strategist through user Interface

Flask: It is python framework which is used for integration of Machine Learning,Deep Learning models into web applications.

Machine Learning Model: Input data collected from the User is passed to this trained model on sales data and output is predicted by this model .

Kaggle Api: It is used for collecting Walmart 45 stores sales data during 2010-2012

6.2 Sprint Planning & Estimation:

Sprint planning involves breaking down the project into smaller, manageable tasks and estimating the effort required for each. For example:

Sprint 1(Start date-25-oct-2023, Estimated End Date-29-oct-2023,Number of days:5days):

User Stories:

USN-1: As a Store Manager, I want user friendly interface
USN-2: As a Store Manager, I want expected number of fields to give input

Sprint 2(Start date-30-oct-2023, Estimated End Date-8-nov-2023,Number of days:8days):

User Stories:

USN-3: As a Regional Manager, I view and compare sales across multiple stores

USN-4: As a Store manager, I want to access the insights to understand the impact of holidays on sales.

USN-5: As a Strategist, I want to identify patterns and trends in historical sales data for strategic decision making

USN-6: As a store manager, I want recommendations for improved decision making.

Sprint 3(Start date-8-nov-2023, Estimated End Date-13-nov-2023,Number of days:6days):

User Stories:

USN-7: As a Data Analyst, I want to forecast future sales trends accurately

USN-8: As a Data Analyst, I want clear display of prediction

6.3 Sprint Delivery Schedule:

Sprint 1(Release Date: 29-oct-2023)

Sprint 2 (Release Date: 7-oct-2023)

Sprint3(Release Date: 13-nov-2023)

7. CODING & SOLUTIONING

7.1 Feature 1: User Interface For Entering the User Input

Code:

```
<form action="/pred" method="POST">
            <label for="store">Store:</label>
            <input type="text" id="store" name="store" required>

            <label for="size">Size:</label>
```

```
            <input type="text" id="size" name="size" required>

            <label for="department">Department:</label>
            <input type="text" id="department" name="department" required>

            <label for="temperature">Temperature:</label>
            <input type="text" id="temperature" name="temperature" required>

            <label for="date">Date:</label>
            <input type="date" id="date" name="date" required>

            <label for="isHoliday">Is Holiday?</label>
            <select id="isHoliday" name="isHoliday" required>
                <option value="1">Yes</option>
                <option value="0">No</option>
            </select>

            <input type="submit">
        </form>
```

Explanation:

This code provides user interface for Entering the input

7.2 Feature 2: Holiday Impact Analysis

Code:

Code for Integrating Holiday.html to application

after Clicking button

```
def another_page1():
    return render_template('Holiday.html')
```

Code of Holiday.html

```
<html>
    <head>
        <style>
           body{
            background-color: orange;
```

```
            }
            #x{

             width:1510px;

            }
            #y{

             width:1510px;

            }
        </style>
    </head>
    <body>
        <center><h1>Holiday Impact</h1></center>

       <img id="x" src="{{ url_for('static', filename='download(3).png') }}"
alt="5"></img></br>
       <img id="y" src="{{ url_for('static', filename='download(4).png') }}"
alt="5"></img>



    </body>
</html>
```

Explanation:

This code is used for Visualising the Holiday impact on Sales in Application

7.3 Feature 3: Comparitive Analysis Visualisation

Code:

Code for Integrating Comparitive .html to

application after Clicking button

```
@app.route('/another_page2')
def another_page2():
    return render_template('Comparitive.html')
```

Code of Comparitive.html

```
<html>

    <head>
        <style>
           body{
            background-color: orange;
           }
           #x{
            width:1510px;


           }
           #y{
            width:1510px;

           }
        </style>
    </head>
    <body>
        <center><h1>Comparitive Analysis</h1></center>

       <img id="x" src="{{ url_for('static', filename='download(2).png') }}"
alt="5"></img></br>
       <img id="y" src="{{ url_for('static', filename='download(11).png') }}"
alt="5"></img>



    </body>
</html>
```

Explanation:

This code is used for Visualising the Comparitive Analysis of 45 Stores in Application

7.4 Feature 4: Trend Analysis Visualisation

Code:

Code for Integrating Trends .html to application

after Clicking button

```
app.route('/another_page3')
def another_page3():
    return render_template('Trend.html')
```

Code of Trend.html

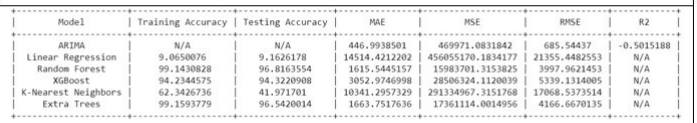
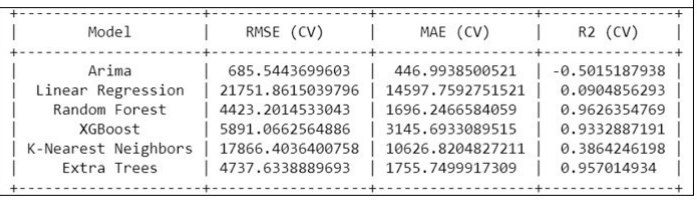
```
<html>
    <head>
        <style>
           body{
            background-color: orange;
           }
           #x{

            width:1510px;

           }
           #y{

            width:1510px;

           }
        </style>
    </head>
    <body>
        <center><h1>Trends Analysis</h1></center>

       <img id="x" src="{{ url_for('static', filename='download(5).png') }}"
alt="5"></img><br>
       <img id="y" src="{{ url_for('static', filename='download(6).png') }}"
alt="5"></img>



    </body>
</html>
```

Explanation:

This code is used for Visualising the Trends Analysis of 45 Stores in Application

8. PERFORMANCE TESTING

8.1 Performace Metrics & Results

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score - **Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report -	
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

Model	Training Accuracy	Testing Accuracy	MAE	MSE	RMSE	R2
ARIMA	N/A	N/A	446.9938501	469971.0831842	685.54437	-0.5015188
Linear Regression	9.0650076	9.1626178	14514.4212202	456055170.1834177	21355.4482553	N/A
Random Forest	99.1430828	96.8163554	1615.5445157	15983701.3153825	3997.9621453	N/A
XGBoost	94.2344575	94.3220908	3052.9746998	28506324.1120039	5339.1314005	N/A
K-Nearest Neighbors	62.3426736	41.971701	10341.2957329	291334967.3151768	17068.5373514	N/A
Extra Trees	99.1593779	96.5420014	1663.7517636	17361114.0014956	4166.6670135	N/A

Model	RMSE (CV)	MAE (CV)	R2 (CV)
Arima	685.5443699603	446.9938500521	-0.5015187938
Linear Regression	21751.8615039796	14597.7592751521	0.0904856293
Random Forest	4423.2014533043	1696.2466584059	0.9626354769
XGBoost	5891.0662564886	3145.6933089515	0.9332887191
K-Nearest Neighbors	17866.4036400758	10626.8204827211	0.3864246198
Extra Trees	4737.6338889693	1755.7499917309	0.957014934

8. ADVANTAGES & DISADVANTAGES

8.1 Advantages:

1. Improved Sales Forecasting:

- The utilization of advanced machine learning algorithms enables more accurate sales forecasting.

- Patterns and correlations identified through data analysis contribute to precise predictions.

2. Enhanced Decision-Making:

- In-depth analysis of sales data provides valuable insights for strategic decision-making in inventory management, marketing, and resource allocation.

3. Seasonal and Holiday Pattern Recognition:

- The project focuses on key holiday periods, allowing for the identification of seasonal trends and their impact on consumer behavior.

4. Utilization of Multiple Models:

- The inclusion of various machine learning models such as Linear Regression, XGBoost, Random Forest, and ARIMA provides flexibility and the ability to choose the most suitable model for different scenarios.

8.2 Disadvantages:

1. Data Dependency:

- The accuracy of the models heavily relies on the quality and completeness of the input data. Inaccuracies or missing information may affect the performance of the models.

2. Model Complexity:

- The use of multiple models may introduce complexity, making it challenging to interpret and implement the most effective solution in a real-world setting.

3. Resource Intensiveness:

- Implementing machine learning models, especially more complex ones, may require significant computational resources and time.

9. CONCLUSION

In conclusion, the "Walmart Sales Analysis for Retail Industry with Machine Learning" project represents a comprehensive approach to understanding and predicting sales dynamics. By leveraging advanced machine learning techniques, the project aims to enhance sales forecasting accuracy, improve decision-making, and uncover valuable insights into consumer behavior during holidays.

The inclusion of various visualizations and exploration of different models provides a holistic view of the data, allowing for a more informed analysis. The project's success lies in its ability to harness the power of machine learning to unravel complex sales patterns and provide actionable insights for Walmart's retail strategy.

10. FUTURE SCOPE

The project lays the groundwork for further advancements and improvements. Future developments could include:

1. Dynamic Model Tuning:

- Implementing mechanisms for dynamic model tuning to adapt to changing market conditions and consumer behavior.

2. Integration with External Data:

- Incorporating external data sources, such as economic indicators or competitor information, to enhance the models' predictive capabilities.

3. Real-time Sales Forecasting:

- Moving towards real-time sales forecasting to enable more responsive decision-making.

4. Advanced Feature Engineering:

- Exploring additional features and advanced feature engineering techniques to capture more nuanced patterns in the data.

11. APPENDIX

11.1 Source Code
https://drive.google.com/file/d/1PqnUOWd_nYRKa_SlP_m1LQPvnwYMehbz/view?usp=sharing

13.2 Dependencies

Ensure you have the required Python libraries installed. You can install them using the following:

bashCopy code

13.3 Data Sources

The project relies on several datasets, including:

- **train.csv**: Training data containing weekly sales information.
- **features.csv**: Additional features related to the stores.
- **stores.csv** : Store-specific information.

13.4 Additional Resources

For a more detailed understanding and organization of the code, it is recommended to refer to the complete project repository, which may include additional documentation, notebooks, or scripts.

14. REFERENCES

Include any references, datasets, or external sources used during the project.

- Walmart Sales Dataset
- PMDARIMA Documentation
- XGBoost Documentation
- Scikit-Learn Documentation