A CASE STUDY (IEEE Format)

Software Requirements Specification Document

SALES FORECAST PREDICTION SYSTEM

Submitted by:

(102103373) ANUPRIYA LATHEY

(102103377) NITLEEN KAUR

(102103394) PIA GUPTA

BE Third Year, COE14

Group No: 5

Submitted to:

Ms. ANAMIKA SHARMA

Lab Instructor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

TABLE OF CONTENTS

1. Introduction

1.1 Purpose of this Document

The purpose of this SRS document is to provide a detailed overview of our software product, its parameters and goals. This document describes the project's target audience and its user interface, hardware and software requirements. It defines how our client, team and audience see the product and its functionality.

1.2 Scope of the Development Project

Sales Forecast Prediction using Machine Learning is a pivotal project that addresses the pressing need for businesses to make accurate sales predictions. The essence of this project is to leverage historical sales data, combined with cutting-edge machine learning techniques, to enable companies to forecast future sales trends with precision. In doing so, this project seeks to empower decision-makers with invaluable insights, thereby facilitating more informed and strategic choices in areas such as inventory management, marketing strategies, and financial planning. The project's primary goal is to develop a robust and reliable system that harnesses the power of data to provide forecasts for sales performance across various products, time frames, and geographic locations. By analyzing patterns, trends, and historical data, this system will become a strategic tool for businesses in optimizing their operations, improving customer service, and achieving sustainable   growth

1.3 Definitions, abbreviations and acronyms

<u>Definitions</u>

Table 1 gives explanation of the most commonly used terms in this SRS document.

Table 1: Definitions for most commonly used terms

S.No.	Term	Definition
1	Forecasting	a method of making informed predictions by using historical data as the main input for determining the course of future trends.
2	Hyperparameters	parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.
3	Scikit-learn, TensorFlow, or PyTorch	Python Frameworks
4	Apportioning	To divide and assign according to a plan
5	XG Boost regressor	Extreme Gradient Boosting (XGBoost) is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm.

Abbreviations

Table 2 gives the full form of most commonly used mnemonics in this SRS document.

Table 2: Full form for most commonly used mnemonics

S.No.	Mnemonic	Full Form
1	CSV	Comma Separated Values
2	MAE	Mean Absolute Error
3	RMSE	Root Mean Square Error
4	ML	Machine Learning
5	SRS	Software Requirement Specification
6	UML	Unified Modelling Language
7	GDPR	General Data Protection Regulation
8	DFD	Data Flow Diagram
9	XG Boost	Extreme Gradient Boost

1.4 References

[1] Sales Forecasting guide. **Link**: https://www.anaplan.com/blog/sales-forecasting-guide/

[2] XG Boost regressor. **Link**: https://machinelearningmastery.com/xgboost-for-regression/

[3] Software Engineering , A Practitioner's Approach by Roger S. Pressman

[4] UML Diagrams. **Link**: https://www.lucidchart.com/blog/types-of-UML-diagrams

1.5 Gantt Chart WBS (Work Breakdown Structure):

S.No.	Task Name	Duration	Start	End	August							September																						
					21 22 23 24 25 26 27	28 29 30 31	1 2 3	4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1	Planning & Feasibility Study	1 week	21.08.23	27.08.23	███			
2	Requirement Analysis	4 weeks	28.08.23	24.09.23			████████████████████████

					September		October	
					25 26 27 28 29 30 1	2 3 4 5 6 7 8	9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
3	Design	2 weeks	25.09.23	08.10.23	██████████		
4	Development	5 weeks	09.10.23	14.11.23			████████████████████████████████

					Oct		November		Dec
					30 31	1 2 3 4 5 6 7 8 9 10 11 12 13	14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	1 2 3
4	Development	5 weeks	09.10.23	12.11.23	████████████████		
5	Testing	2 weeks	13.11.23	26.11.23		█████████████	
6	Documentation & Deployment	1 week	27.11.23	03.12.23			███████

1.6 Overview

The remaining sections of this document provide a general description, including characteristics of the users of this project, the product's hardware, and the functional and data requirements of the product. General description of the project is discussed in section 2 of this document. Section 2 gives the functional requirements, data requirements and constraints and assumptions made while designing the multi-utility system. It also gives the user viewpoint of product use. Section 3 gives the specific requirements of the product. Section 3.0 also discusses the external interface requirements and gives detailed description of functional requirements.

2. Overall Description

2.1 Product Perspective

Step 1: ***User Authentication***

The user, whether an administrator, sales manager, or executive, accesses the system and provides their login credentials. The system verifies the user's credentials by checking them against the user database. If the credentials are valid, access is granted; otherwise, access is denied.

Step 2: ***Dashboard Show***

After successful authentication, the user is presented with a dashboard screen.

Step 3: ***Data Import***

Sales managers can choose to import historical sales data for analysis and forecasting. If selected, the system provides a data import feature that allows sales managers to upload historical sales data files (e.g., CSV). Behind the scenes, the system processes the uploaded data, cleans it, and prepares it for analysis.

Step 4: ***Data Analysis and Forecasting***

Users can view sales forecasts and reports on the dashboard. The system uses machine learning algorithms to analyze the imported historical sales data. It then generates sales forecasts based on the analyzed data and user-defined parameters. The results are displayed on the dashboard in the form of charts, graphs, and textual information.

Step 5: ***Security and Audit Logs***

User Action: Security measures are enforced transparently during user interactions with the system. The system enforces security behind the scenes by using authentication and authorization mechanisms.

2.2 Product Functions

The product should be able to perform the following operations:

- **Data Collection**: Gather comprehensive historical sales data, encompassing factors such as product details, sales quantities, customer demographics, and market conditions.

- **Model Selection**: Implement a range of machine learning models including linear regression, decision trees, random forests and neural networks. Conduct comprehensive model selection and comparison exercises, taking into account performance metrics and business objectives.

- **Model Training, Testing and Validation**: Train selected models on the training dataset, ensuring proper validation and hyperparameter tuning to optimize predictive accuracy. Implement cross-validation techniques to assess and improve the robustness of models. Evaluate model performance on the validation and testing datasets, employing relevant metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Confirm the model's ability to generalize to unseen data.

- **Visualization and Reporting**: Create intuitive and interactive dashboards and reports for visualizing historical sales data, comparing actual vs. predicted sales, and identifying emerging trends. Enable users to export reports in multiple formats (e.g., PDF, Excel, graphs).

- **User Management and Documentation**: Implement role-based access control and user management functionalities to control and track user access to the forecasting system. Provide comprehensive documentation and training materials, ensuring that users and administrators can effectively navigate and utilize the forecasting system.

2.3 User Characteristics

These user types are listed below as follows:

1. **Administrators:**
   Data Collection and Management: Administrators are responsible for gathering and managing the data required for training and testing the sales prediction model. They ensure data quality, handle missing values, and preprocess the data for analysis.

   Model Selection and Training: They choose the appropriate machine learning algorithms or techniques for the task, train the model using historical sales data, and fine-tune hyperparameters.

   Monitoring and Maintenance: Administrators continuously monitor the model's performance and retrain it periodically to adapt to changing trends.

2. **Sales Managers:**
   Model Evaluation: Sales managers interact with the system to evaluate the performance of the sales prediction model. They analyze metrics such as accuracy, precision, recall, and F1 score to assess how well the model is predicting sales.

   Visualizations: They use visualizations provided by the system to understand historical sales trends and predictions, helping them make informed decisions.

3. **Sales Representatives:**
Sales Projections: Sales representatives can access the system to view projected sales for specific products or timeframes. This information aids in setting sales targets and prioritizing their efforts.

4. **Executives and Decision Makers:**
Performance Reports: Executives use the system to access comprehensive reports on overall sales performance, broken down by products, regions, or other relevant segments.

Long-Term Planning: The system's predictions can assist in long-term planning, helping executives allocate resources, adjust marketing strategies, and make informed business decisions.

Scenario Analysis: Decision makers can use the system to simulate different   scenarios and assess how changes in various factors (e.g., pricing, marketing spend) might impact future sales.

5. **IT Support:**
Infrastructure and Deployment: IT support ensures that the ML-based sales prediction system is properly deployed and integrated with other systems. They manage server resources, scalability, and security.

Troubleshooting: If users encounter technical issues while interacting with the system, IT support provides assistance in resolving these problems.

6. **Customers (limited interaction):**
Online Shopping Platforms: If the ML system is integrated with online shopping platforms, customers might indirectly interact with it by receiving personalized recommendations based on their browsing and purchasing history.

2.4 General Constraints, Assumptions and Dependencies

CONSTRAINTS:

1. **Data Quality:** The accuracy of sales forecasts is heavily dependent on the quality of historical sales data. Any inconsistencies, inaccuracies, or missing data can impact the effectiveness of the forecasting system.

2. **Computational Resources:** The availability of computational resources, including processing power and memory, may constrain the complexity of machine learning models and the size of datasets that can be used.

3. **Historical Data Availability:** The availability of historical sales data for certain products or regions may be limited, affecting the accuracy of forecasts for those specific areas.

4. **Competitive Intelligence:** Limited access to competitor data and market intelligence can constrain the project's ability to factor in external competitive pressures into forecasts.

5. **Model Complexity:** The project may be constrained by the computational resources required to implement highly complex machine learning models, impacting the choice of algorithms and techniques.

6. **Data Availability for New Products:** Forecasting for new product launches can be challenging due to the lack of historical sales data, potentially limiting the accuracy of forecasts for these items.

7. **Seasonality and External Events:** The presence of seasonality or external events (e.g., holidays, economic fluctuations) can introduce constraints, as these factors may require specialized modelling techniques.

8. **Data Imbalance:** Imbalanced datasets, where certain products or regions have significantly fewer sales data points than others, can constrain the model's ability to make accurate predictions for underrepresented categories.

ASSUMPTIONS:

1. **Historical Data Availability:** It is assumed that a sufficient amount of historical sales data is available for training and testing the machine learning models.

2. **Data Relevance:** The assumption is made that historical sales data is relevant to future sales patterns and can be used as a basis for forecasting.

3. **Stable Business Environment:** The project assumes that the business environment, including market conditions and product offerings, will remain relatively stable during the forecasting period.

4. **User Training:** It is assumed that users will receive the necessary training to effectively use the forecasting system and interpret its results.

5. **Data Privacy Compliance:** The project assumes that data privacy regulations will be adhered to, and any necessary consents or permissions for data usage have been obtained.

6. **Customer Behaviour Consistency:** The project assumes that customer purchasing behaviour remains relatively consistent over time, without significant shifts in preferences or buying habits.

7. **Model Generalization:** It is assumed that the selected machine learning models will generalize well across different products, regions, or market segments, even when training data is limited for certain subsets.

DEPENDENCIES:

1. **Data Sources**: The project is dependent on the availability and reliability of data sources, including historical sales data and any external data used for contextual information.

2**. Machine Learning Libraries:** Dependencies exist on machine learning libraries and frameworks, such as scikit-learn, TensorFlow, or PyTorch, for model development.

3. **Feedback Loop:** The project may depend on establishing a feedback loop for continuous improvement, where insights from forecast results are used to refine the models and strategies.

4. **IT Infrastructure:** Dependencies exist on the IT infrastructure, including server availability, database performance, and network stability, to support the forecasting system.

5. **Software Updates:** The project relies on specific software or libraries, hence dependencies exist on the maintenance and updates of those tools to ensure compatibility and security.

6. **Data Cleaning and Validation Tools:** Dependencies on tools and processes for data cleaning, validation, and anomaly detection to maintain data quality.

7. **Training Data Updates:** Dependencies on processes to periodically update training data to account for changing sales patterns and emerging trends.

2.5 Apportioning of requirement

The Sales Forecasting Predicting System is to be implemented in the following three phases:

i. **Pilot Phase:** Focus on core functionalities, such as data import, model selection, and basic reporting. Ensure stability and performance. Implement basic user roles (e.g., administrators, sales managers) and their associated permissions for data access and configuration.

ii. **Institute wide deployment:** Following the successful completion of the pilot phase, we plan to deploy the same across more granular access controls, such as team-level permissions and executive-level access.

iii. **Expansion of data sources:** In future versions, consider expanding data sources to include external data feeds and third-party integrations based on user requirements.

3. Specific Requirements

3.1 External Interface Requirements

The following list presents the external interface requirements:

- The product requires very limited graphics usage. User inputs through the keypad, such as login credentials, data selection, and configuration settings.

- Displayed information and text-based responses on the screen; no graphics, sound, or animation are required.

- The system's user interface should be tailored to fit the specified screen resolution without graphical elements or animations.

3.2 Detailed Description of Functional Requirements

1. **Data Collection:** Gather comprehensive historical sales data, encompassing factors such as product details, sales quantities, customer demographics, and market conditions.

2. **Data Preprocessing:** Rigorously clean and preprocess the collected data, addressing issues such as missing values, outliers, and data inconsistencies. Implement advanced feature engineering techniques to derive meaningful predictors for model training.

3. **Data Splitting:** Divide the dataset into training, validation, and testing subsets to facilitate rigorous model evaluation and fine-tuning.

4. **Model Selection:** Implement a range of machine learning models, including but not limited to linear regression, decision trees, random forests, XGBoost and neural networks. Conduct comprehensive model selection and comparison exercises, taking into account performance metrics and business objectives.

5. **Model Training:** Train selected models on the training dataset, ensuring proper validation and hyperparameter tuning to optimize predictive accuracy. Implement cross-validation techniques to assess and improve the robustness of models.

6. **Model Testing and Validation:** Evaluate model performance on the validation and testing datasets, employing relevant metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Confirm the model's ability to generalize to unseen data.

7. **Forecast Generation:** Deploy the trained models to generate sales forecasts for future time periods, offering flexibility in customizing forecast horizons and granularity (daily, weekly, monthly).

8. **Visualization and Reporting:** Create intuitive and interactive dashboards and reports for visualizing historical sales data, comparing actual vs. predicted sales, and identifying emerging trends. Enable users to export reports in multiple formats (e.g., PDF, Excel).

9. **Alerting Mechanism:** Implement an automated alerting system to notify stakeholders when sales figures significantly deviate from forecasts, facilitating rapid response and adaptation.

10. **Integration:** Seamlessly integrate the forecasting system with existing Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM) systems to ensure smooth data flow and decision support.

11. **User Management**: Implement role-based access control and user management functionalities to control and track user access to the forecasting system.

12. **Documentation and Training:** Provide comprehensive documentation and training materials, ensuring that users and administrators can effectively navigate and utilize the forecasting system.

3.3 Performance Requirements

1. **System Deployment:** The Sales Forecast Prediction System will be deployed on standard desktop and server hardware. It should be accessible from a web browser, ensuring compatibility with a wide range of devices and platforms.

2. **Concurrency:** The system should support concurrent user access. It must be able to handle multiple users simultaneously, regardless of the number of terminals, ensuring a responsive user experience.

3. **Data Handling:** The system will process both numerical and textual data related to sales and forecasting. It should efficiently handle varying data volumes, scaling to accommodate large datasets when necessary.
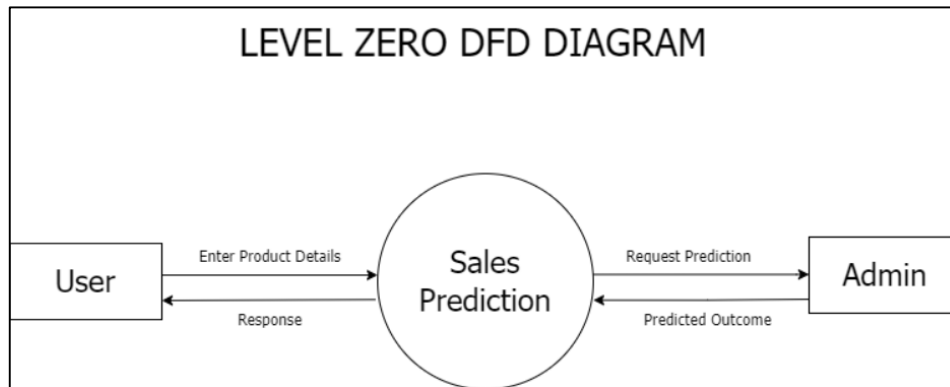
4. **Response Time:** Under normal operating conditions, at least 95% of user requests for sales forecasts should be processed and responded to within 5 seconds. This ensures that users can quickly retrieve critical information to make informed decisions.

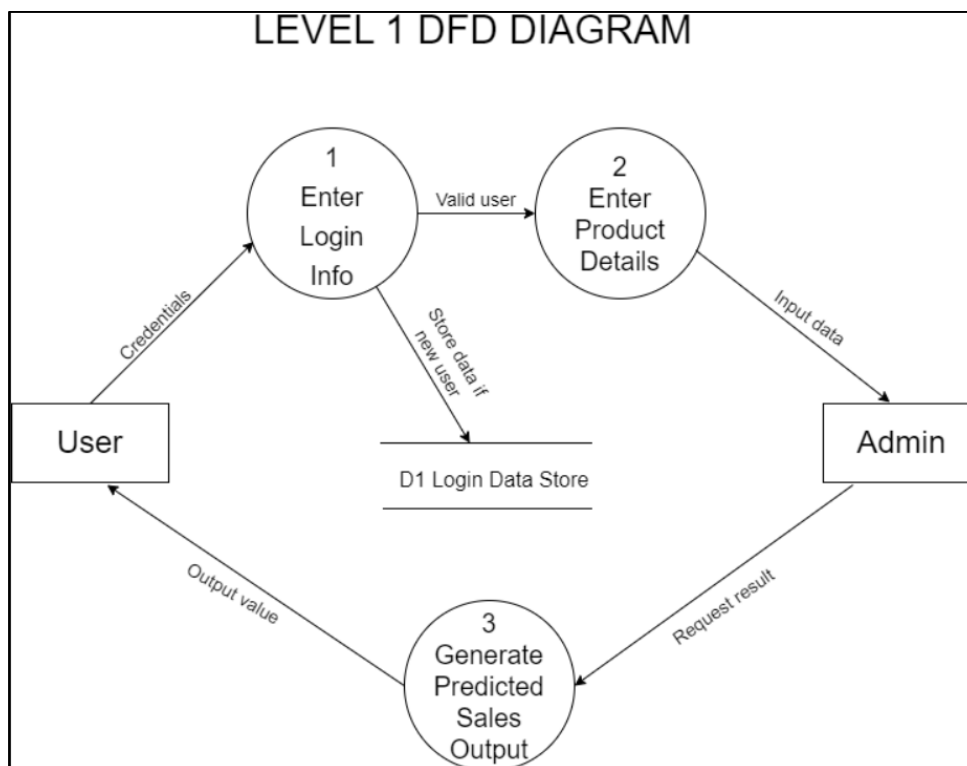## 3.4 Non-Functional Requirements:

1. **Accuracy:** The system must consistently achieve a high level of forecasting accuracy to minimize errors in predictions, underpinning its value to the business.

2. **Scalability:** The system should be designed to scale gracefully, accommodating large datasets and accommodating increasing data volume over time.

3. **Performance:** Forecast generation should be efficient, with minimal delays, to facilitate timely decision-making.

4. **Robustness:** The system should demonstrate resilience to data anomalies, changing market conditions, and unexpected sales patterns.

5. **Security:** Implement robust data security measures and user authentication protocols to safeguard sensitive sales information.

6. **Usability:** The user interface should be highly intuitive and user-friendly, requiring minimal training for users to operate effectively.

7. **Compatibility:** Ensure compatibility with various data formats and seamless integration with common business software and systems.

8. **Maintainability:** The system should be designed for easy maintenance, including regular model retraining and updates to accommodate evolving business needs.

9. **Reliability:** The forecasting system should be highly reliable, with minimal downtime or disruptions to business operations.

10. **Regulatory Compliance:** Adhere to any relevant data protection and privacy regulations (e.g., GDPR) if applicable, ensuring the responsible use of customer and sales data.

11. **Auditability:** Maintain detailed logs and audit trails for user actions and system processes, enabling traceability and accountability.

12. **Cost-Efficiency:** Optimize resource usage to minimize operational costs, making the system sustainable and budget-friendly for the organization.

3.5 Logical Database Requirements

1. Level 0 DFD

LEVEL ZERO DFD DIAGRAM

User — Enter Product Details → Sales Prediction — Request Prediction → Admin
Sales Prediction — Response → User
Admin — Predicted Outcome → Sales Prediction

2. Level 1 DFD

LEVEL 1 DFD DIAGRAM

1 Enter Login Info
2 Enter Product Details
3 Generate Predicted Sales Output
D1 Login Data Store

User — Credentials → Enter Login Info
Enter Login Info — Valid user → Enter Product Details
Enter Login Info — Store data if new user → D1 Login Data Store
Enter Product Details — Input data → Admin
Admin — Request result → Generate Predicted Sales Output
Generate Predicted Sales Output — Output value → User

3. Use Case Diagram

USE CASE DIAGRAM

4. Use Case Template

1. Use Case Title	Enter Product Details
2. Abbreviated Title	Enter Product Details
3. Use Case ID	1
4. Actors	User
5. Description	User must enter product specifications like Item Identifier, Item Weight, Item Fat Content, Item Visibility, Item Type, Item MRP, Outlet Identifier, etc to predict its sales.
5.1. Pre-Conditions	Every product must have predefined attributes like Item Identifier, Item Weight, Item Fat Content, Item Visibility, Item Type, Item MRP, Outlet Identifier, etc to predict its sales.
5.2. Task Sequence	a) Product form will be displayed on the home page. b) User must enter corresponding details.
5.3. Post Conditions	Foremost step in prediction.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

1. Use Case Title	Login
2. Abbreviated Title	Login
3. Use Case ID	2
4. Actors	User
5. Description	User must enter personal information like name, employee id, etc. Enables Admin to find out which user is valid to access the model.
5.1. Pre-Conditions	Set attributes which user must enter.
5.2. Task Sequence	a) User Info form will be displayed on the page. b) User must enter corresponding details.
5.3. Post Conditions	If successful, user is logged in for the next step.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

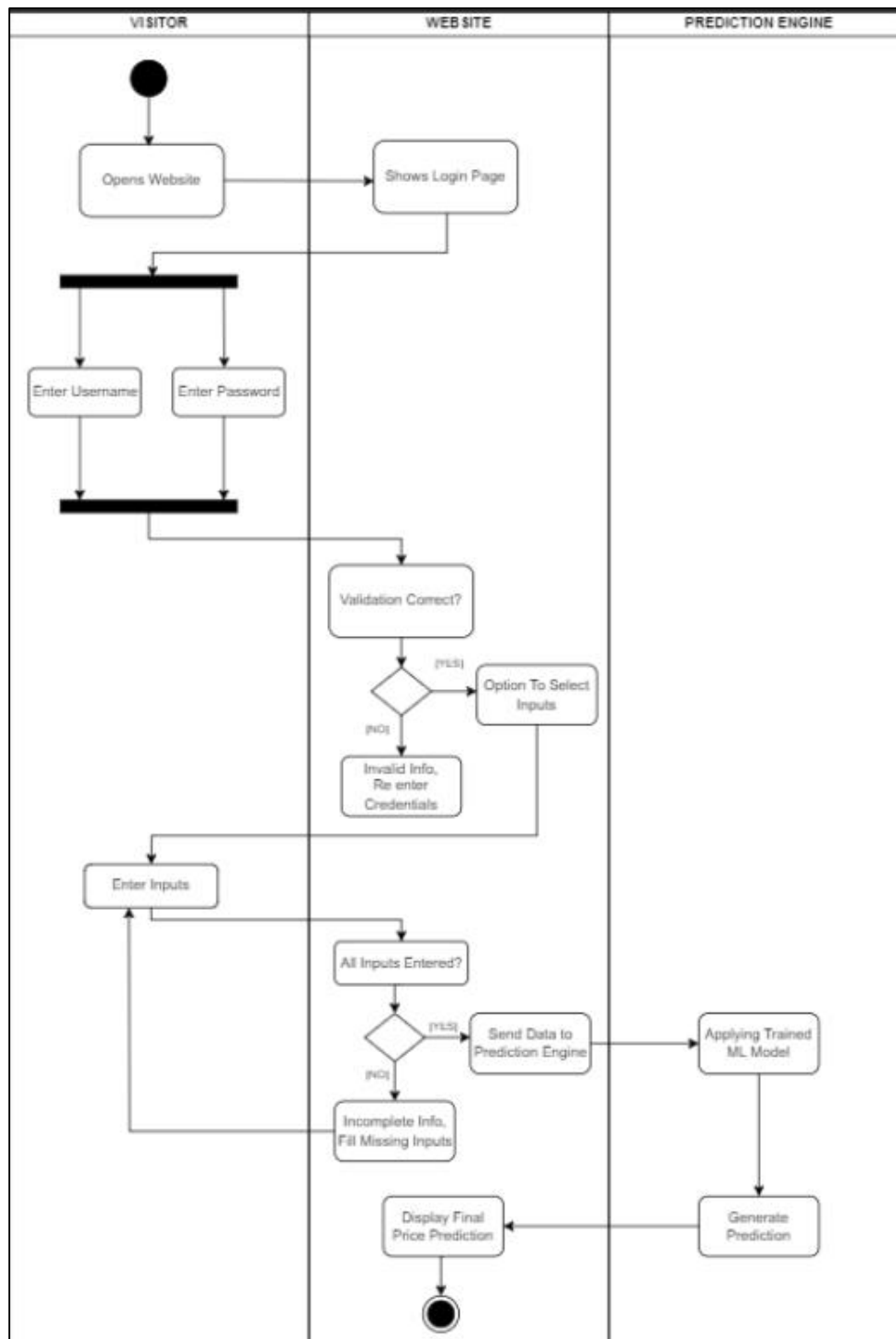|	
1. Use Case Title	Save Login Information
2. Abbreviated Title	Save Login Info
3. Use Case ID	3
4. Actors	User
5. Description	User can choose whether or not to store his login info for future login.
5.1. Pre-Conditions	User must have entered some information prior to this.
5.2. Task Sequence	a)User must select the option whether or not he wants his information saved for future. b)Click login button
5.3. Post Conditions	User information is saved for future.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

1. Use Case Title	Store Login Information
2. Abbreviated Title	Store Login Info
3. Use Case ID	4
4. Actors	Admin
5. Description	Login info entered by user is stored as a record of who logged in.
5.1. Pre-Conditions	User must have entered some data using Login and clicked the Login button.
5.2. Task Sequence	Admin will update the user login history.
5.3. Post Conditions	Data of user who accessed system is stored.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

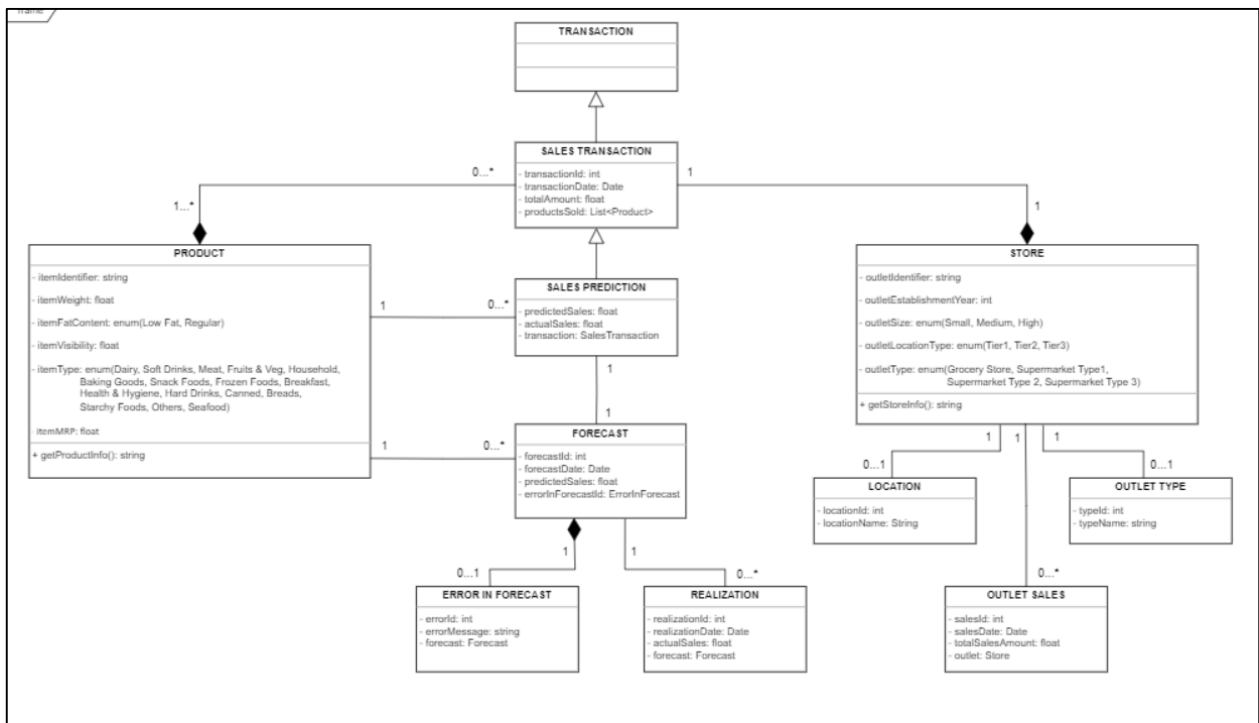| 1. Use Case Title	Check Valid User
2. Abbreviated Title	Check Valid User
3. Use Case ID	5
4. Actors	Admin
5. Description	Admin must check if user is allowed access or not depending upon his employee ID.
5.1. Pre-Conditions	A set sequence of Employee IDs for users who are to be allowed access.
5.2. Task Sequence	Followed by entering of Login details. Admin must display message if user invalid.
5.3. Post Conditions	If valid, user accepted for next step.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

1. Use Case Title	Manage Data
2. Abbreviated Title	Manage Data
3. Use Case ID	6
4. Actors	Admin
5. Description	Admin must manage and store database on which the model is built and if there are any changes to be made on it.
5.1. Pre-Conditions	Pre-defined data set.
5.2. Task Sequence	A dataset must be pre-processed and uploaded.
5.3. Post Conditions	Model is learned to be implemented.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

1. Use Case Title	Predict Output
2. Abbreviated Title	Predict Output
3. Use Case ID	7
4. Actors	Admin
5. Description	Output for which the model is built is calculated.
5.1. Pre-Conditions	All processes must be executed successfully.
5.2. Task Sequence	a) Admin must calculate the result. b) The result must be displayed on the screen, visible to the user.
5.3. Post Conditions	User can apply this model for sales predictions.
6. Modification History	September 20, 2023
7. Author	Nitleen Kaur, Anupriya Lathey, Pia Gupta

5. Activity/ Swimlane Diagram

VISITOR	WEB SITE	PREDICTION ENGINE

Opens Website → Shows Login Page

Enter Username | Enter Password

Validation Correct?

[YES] → Option To Select Inputs

[NO] → Invalid Info, Re enter Credentials

Enter Inputs

All Inputs Entered?

[YES] → Send Data to Prediction Engine → Applying Trained ML Model

[NO] → Incomplete Info, Fill Missing Inputs

Display Final Price Prediction ← Generate Prediction

6. Class Diagram

3.6 Quality Attributes

The Sales Forecast Prediction System is designed to cater to a diverse range of users, including sales managers, marketing professionals, and data analysts. The product is expected to exhibit robust performance across various hardware configurations and operating systems. It must be capable of handling a wide array of input scenarios, including unexpected data inputs or user interactions.

3.7 Other Requirements

None at this time

4. Change History

202309	Version 1.0 – Initial Release

5. Document Approvers

SRS for Sales Forecast Prediction System is approved by:

_____

Name:

Designation:

Date: