

## Project Design Phase - II

### Solution Architecture

Date	24 October 2023
Team ID	PNT2023TMID591623
Project Name	Project - Walmart Sales Forecasting
Maximum Marks	4 Marks

#### Solution Architecture:

##### The structure of the software is as follows:

- **User Inputs:** This component provides a user interface for the user to input data into the system.
- **Data Preprocessing:** This component cleans and prepares the data for training and evaluation. This may include tasks such as removing outliers, scaling the data, and converting the data to a format that is compatible with the chosen machine learning algorithm.
- **Model:** This component represents the machine learning model. The model is trained on a set of data and then used to make predictions on new data. Evaluation: This component evaluates the performance of the model on a held-out test set. This helps to ensure that the model is able to generalize to new data.
- **Prediction:** This component uses the trained model to make predictions on new data

##### The software behaves as follows:

- The user inputs data into the system through the user interface.
- The data preprocessing component cleans and prepares the data for training and evaluation.
- The machine learning model is trained on the preprocessed data.
- The model is evaluated on a held-out test set.
- The model is used to make predictions on new data.

##### The software characteristics include:

- **Scalability:** The software should be able to handle large amounts of data and users.
- **Robustness:** The software should be able to handle errors in the data and unexpected inputs.

- **Maintainability:** The software should be easy to maintain and update as new data and requirements become available.

**Other aspects of the software include:**

- **Security:** The software should be secure from unauthorized access and attacks.
- **Explainability:** The software should be able to explain how it makes predictions. This is important for trust and transparency.
- **Fairness:** The software should be fair and unbiased in its predictions. Overall, the software architecture shown in the diagram is a well-established and effective approach for building machine learning applications.

**The features of the machine learning software are:**

- **Data preprocessing:** The software should be able to clean and prepare data for training and evaluation. This may include tasks such as removing outliers, scaling the data, and converting the data to a format that is compatible with the chosen machine learning algorithm.
- **Model training:** The software should be able to train a variety of machine learning models, such as classification, regression, and clustering models.
- **Model evaluation:** The software should be able to evaluate the performance of machine learning models on a held-out test set. This helps to ensure that the model is able to generalize to new data.
- **Prediction:** The software should be able to use trained machine learning models to make predictions on new data.
- **User interface:** The software should provide a user interface for the user to input data into the system and view the results of the predictions.

**The development phases for the machine learning software are:**

- **Requirements gathering:** The first step is to gather requirements from the users of the software. This includes understanding the type of data that will be used, the types of predictions that need to be made, and the performance requirements of the software.
- **System design:** Once the requirements have been gathered, the next step is to design the system architecture. This includes choosing the right machine learning algorithms and tools, and designing the user interface.
- **Implementation:** The implementation phase involves writing the code for the software. This includes developing the data preprocessing, model training, evaluation, prediction, and user interface components.
- **Testing:** The software should be thoroughly tested to ensure that it meets the requirements and works as expected. This includes unit testing, integration testing, and system testing.

- **Deployment:** The final step is to deploy the software to production. This involves making the software available to users and monitoring its performance.

**The solution requirements for the machine learning software are:**

- **Scalability:** The software should be able to handle large amounts of data and users.
- **Robustness:** The software should be able to handle errors in the data and unexpected inputs.
- **Maintainability:** The software should be easy to maintain and update as new data and requirements become available.
- **Security:** The software should be secure from unauthorized access and attacks.
- **Explainability:** The software should be able to explain how it makes predictions. This is important for trust and transparency.
- **Fairness:** The software should be fair and unbiased in its predictions.

**Definition:**

- The solution must be able to train and evaluate a variety of machine learning models, such as classification, regression, and clustering models.
- The solution must be able to make predictions on new data using trained machine learning models.
- The solution must be able to handle large amounts of data and users.
- The solution must be robust to errors in the data and unexpected inputs.
- The solution must be maintainable and easy to update as new data and requirements become available.
- The solution must be secure from unauthorized access and attacks.
- The solution must be able to explain how it makes predictions.
- The solution must be fair and unbiased in its predictions.

**Management:**

- The solution must be managed using a well-defined process that includes requirements gathering, system design, implementation, testing, deployment, and maintenance.
- The solution must be monitored to ensure that it is meeting the performance requirements and that it is free of errors.
- The solution must be updated as needed to reflect changes in the data, requirements, or algorithms.

### Delivery:

- The solution must be delivered to users in a way that is easy to use and efficient.
- The solution must be accompanied by documentation that explains how to use the solution and how to interpret the results.
- The solution must be supported by a team of experts who can help users with any problems they may encounter.

### Here are some additional specifications that may be relevant to your specific project:

- **Data format:** The solution should be able to handle data in a variety of formats, such as CSV, JSON, and Parquet.
- **Machine learning framework:** The solution should be implemented using a popular machine learning framework, such as TensorFlow, PyTorch, or scikit-learn.
- **Deployment environment:** The solution should be able to be deployed in a variety of environments, such as on-premises, cloud, or hybrid.
- **Security requirements:** The solution must meet all applicable security requirements, such as HIPAA compliance or PCI DSS compliance.

### Solution Architecture Diagram:

