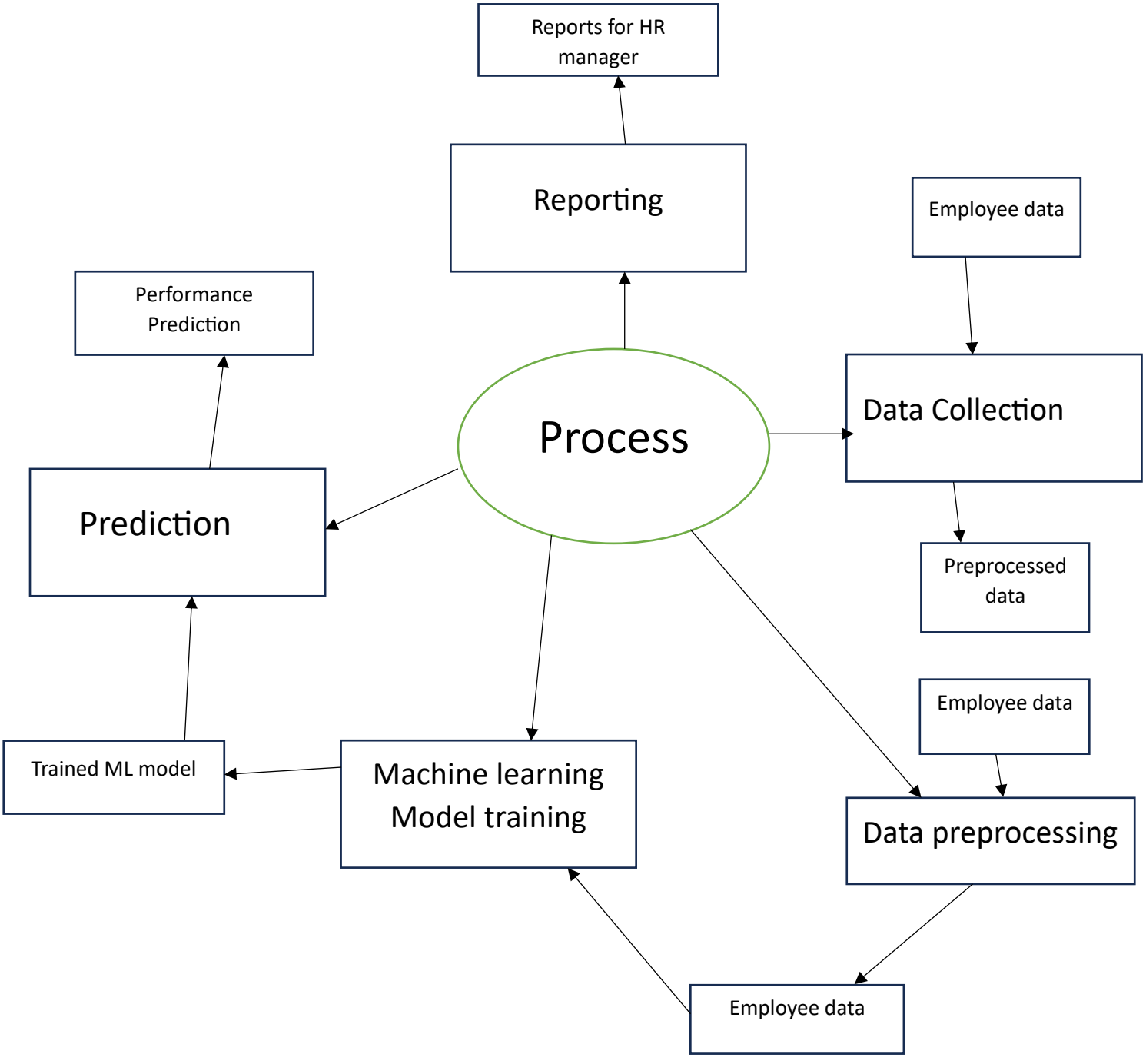
**Project Design Phase-II**
**Data Flow Diagram & User Stories**

| | |
|---|---|
| Date | 03 October 2022 |
| Team ID | Team - 593214 |
| Project Name | Machine Learning approach for Employee Performance Prediction |
| Maximum Marks | 4 Marks |

**Data Flow Diagrams:**

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Stories

Use the below template to list all the user stories for the product.

| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criteria | Priority | Release |
|---|---|---|---|---|---|---|
| HR Manager | Employee Performance Prediction using ML | USN-1 | As an HR manager, I want to upload employee performance data easily, so I can use the machine learning model for predictions. | • There should be an option to upload a CSV file containing employee performance data.<br>• The system should validate the format of the uploaded data.<br>• The uploaded data should be stored securely for processing. | High | 1.0 |
| | | USN-2 | As an HR manager, I want to specify the attributes and metrics to be used for performance prediction, so I can customize the model according to our organization's needs | • There should be a user-friendly interface for selecting the attributes and metrics. | Medium | 1.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | • The selected attributes should be used in the machine learning model for prediction. | | |
| Data Analyst | | USN-3 | As a data analyst, I want to preprocess the employee performance data, including handling missing values and outliers, to ensure data quality and accuracy | • Data preprocessing should include handling missing values, outliers, and data normalization.<br>• The preprocessed data should be stored in a designated location. | High | 1.0 |
| Data Scientist | | USN-4 | As a data scientist, I want to train the machine learning model using the preprocessed data, so I can develop a prediction model. | • The system should have tools and libraries for training machine learning models.<br>• The model should be | High | 1.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | trained using the preprocessed data. | | |
| | | USN-5 | As a data scientist, I want to evaluate the model's performance using various metrics like accuracy, precision, recall, and F1 score, to ensure its reliability. | • The system should provide options to evaluate the model's performance using standard metrics.<br>• Evaluation results should be displayed for analysis. | Medium | 1.0 |
| System Administrator | | USN-6 | As a system administrator, I want to ensure data privacy and security by implementing proper access controls to protect employee data. | • Access to employee data should be controlled based on user roles.<br>• Data should be encrypted and stored securely. | High | 1.0 |
| | | USN -7 | As a system administrator, I want to schedule regular model retraining to keep predictions up to date. | • The system should allow scheduling retraining at | Medium | 1.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | specified intervals.<br>• Retraining should use the most recent data. | | |