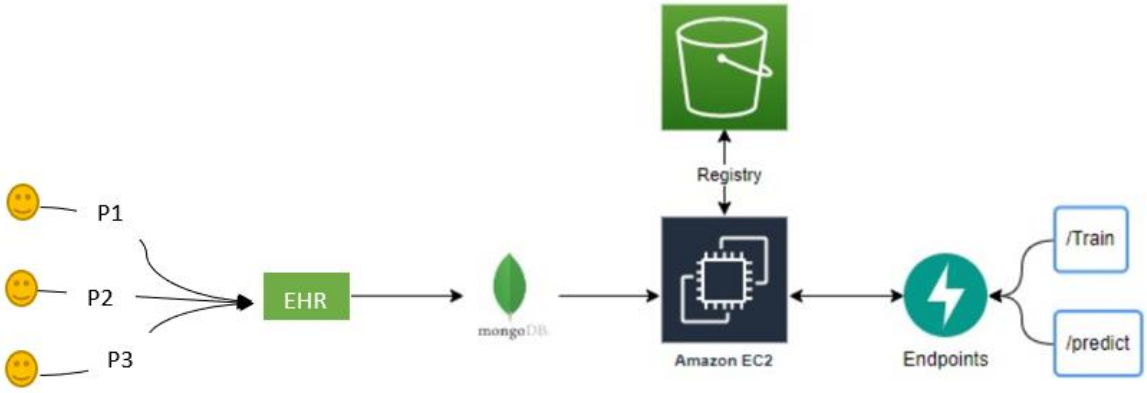
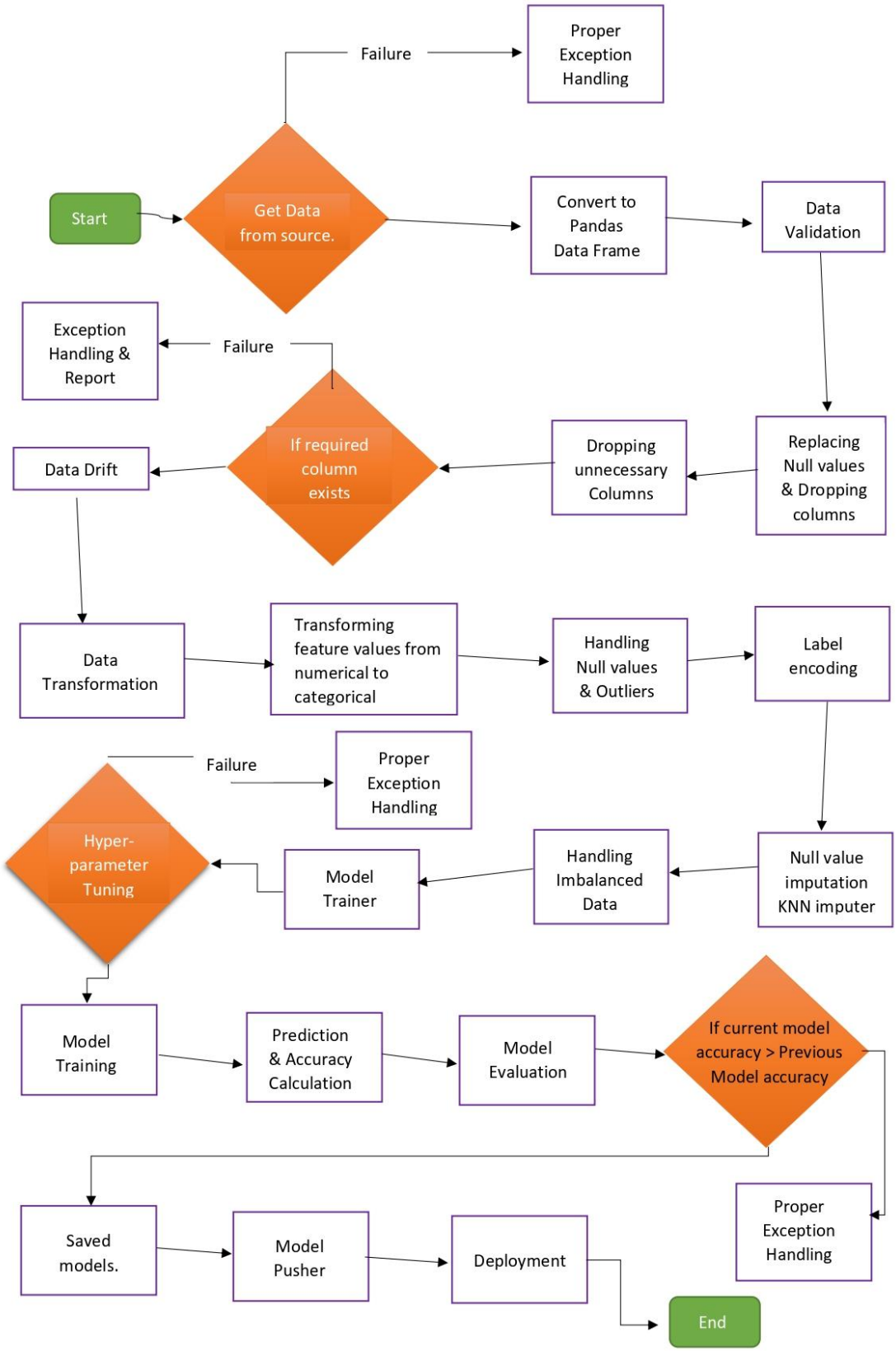
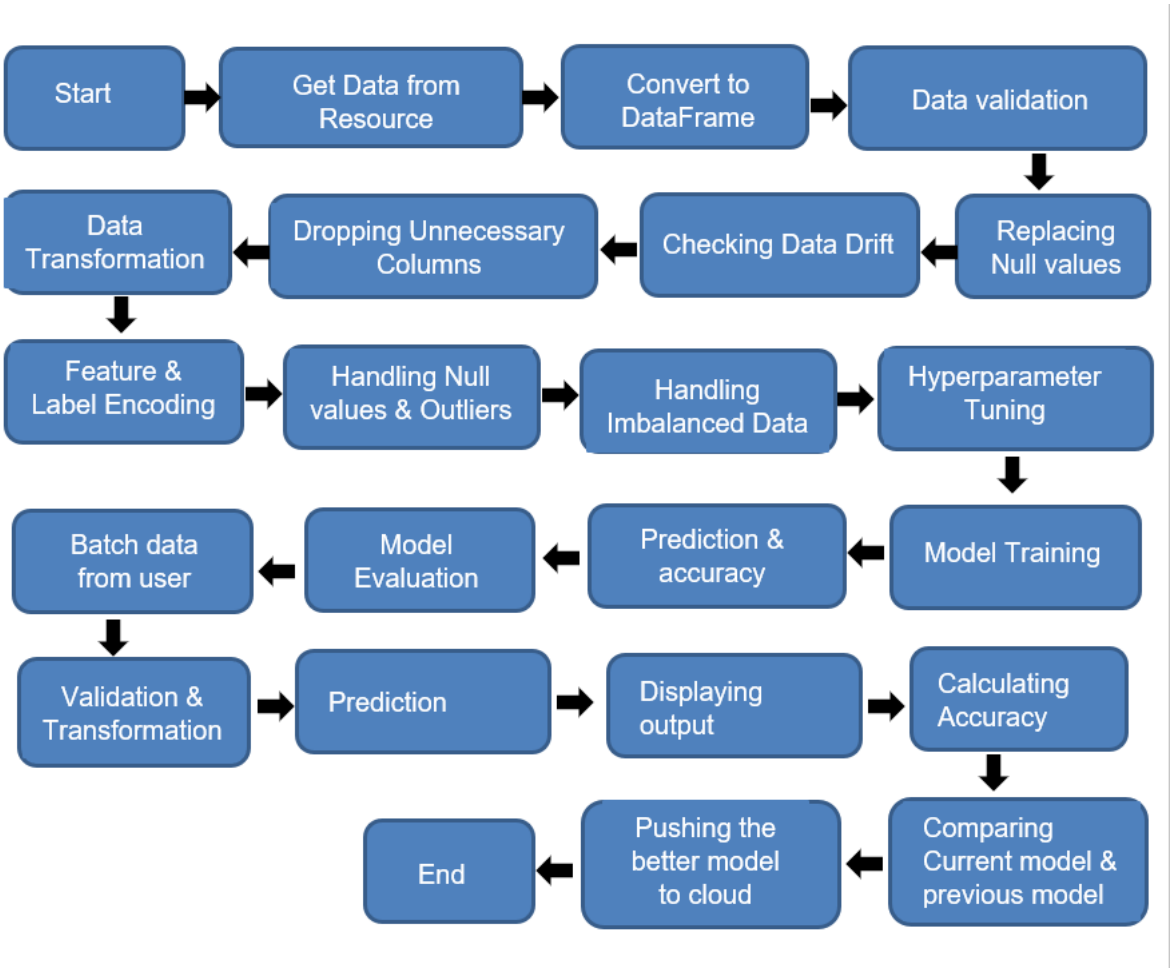
1. Data Collections

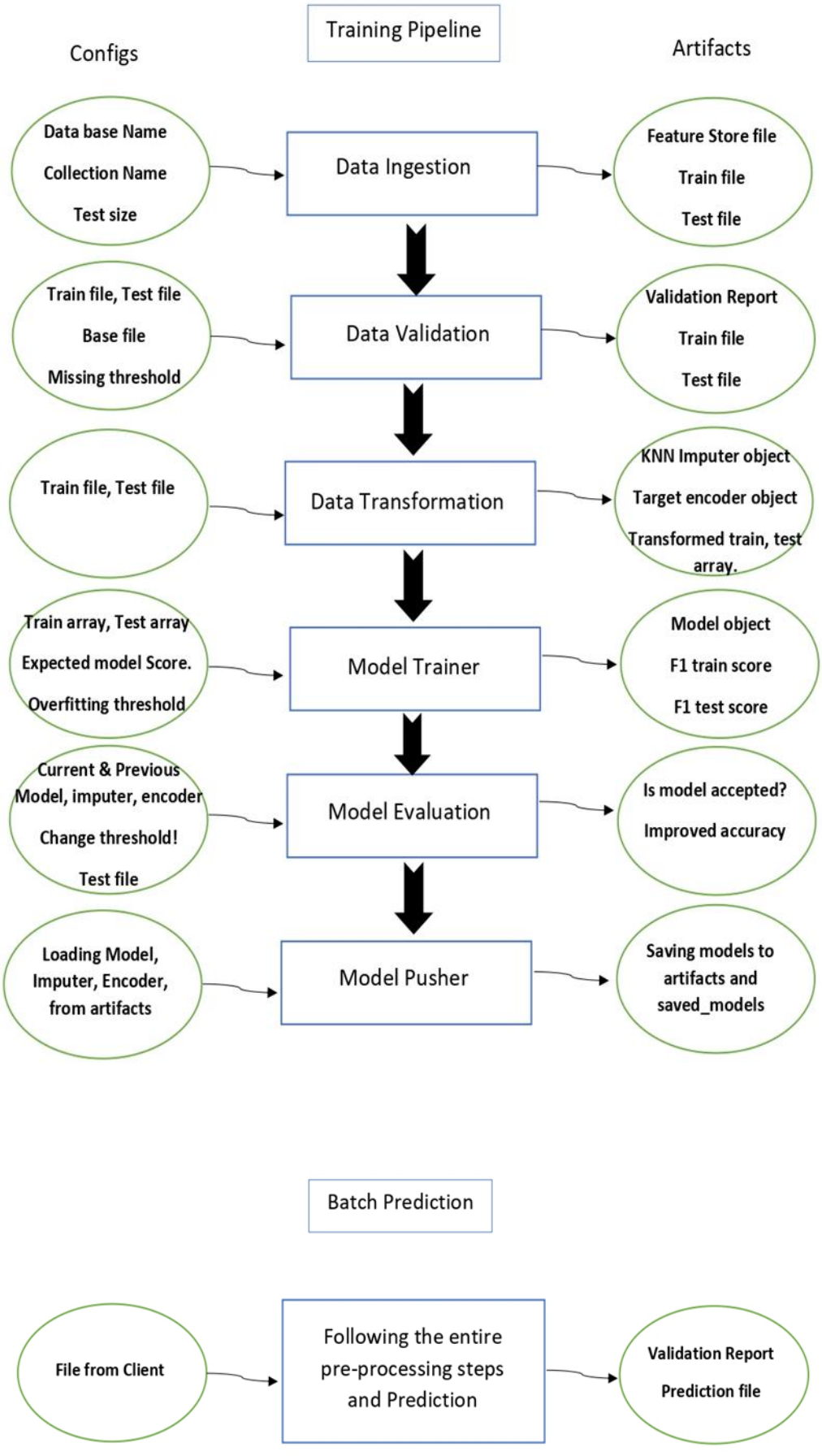


2.Project Architecture

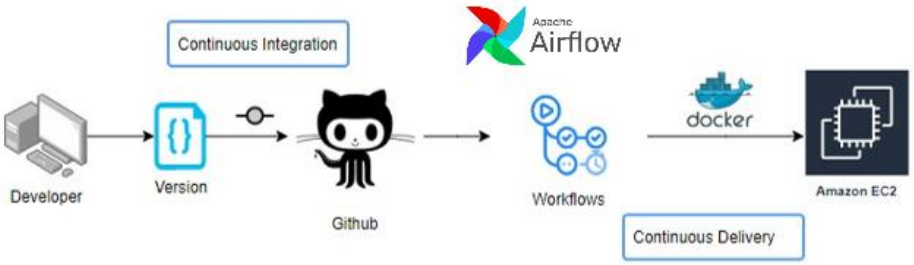


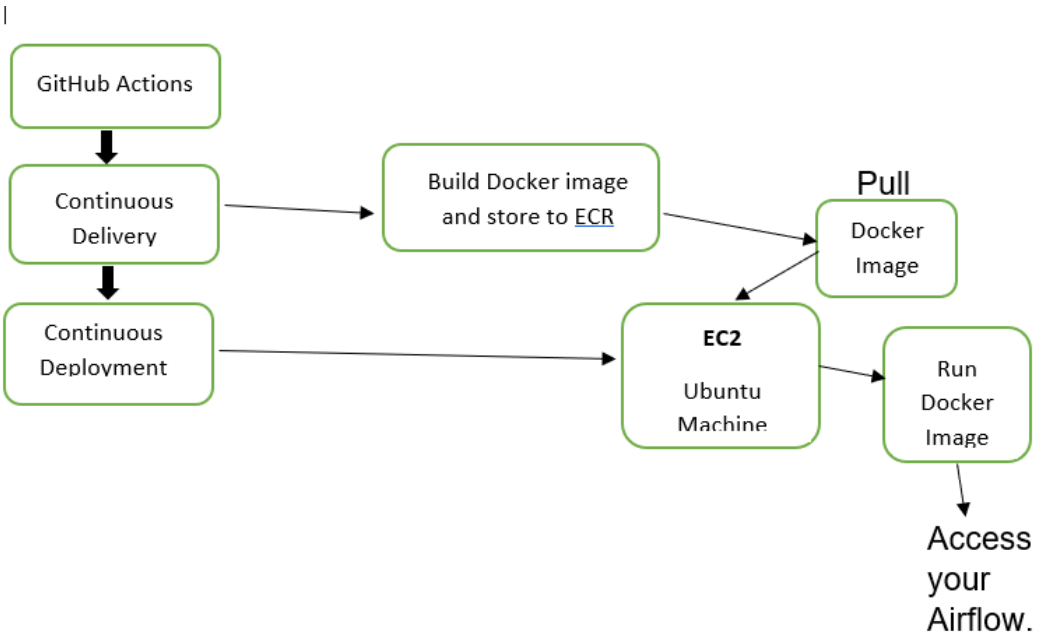
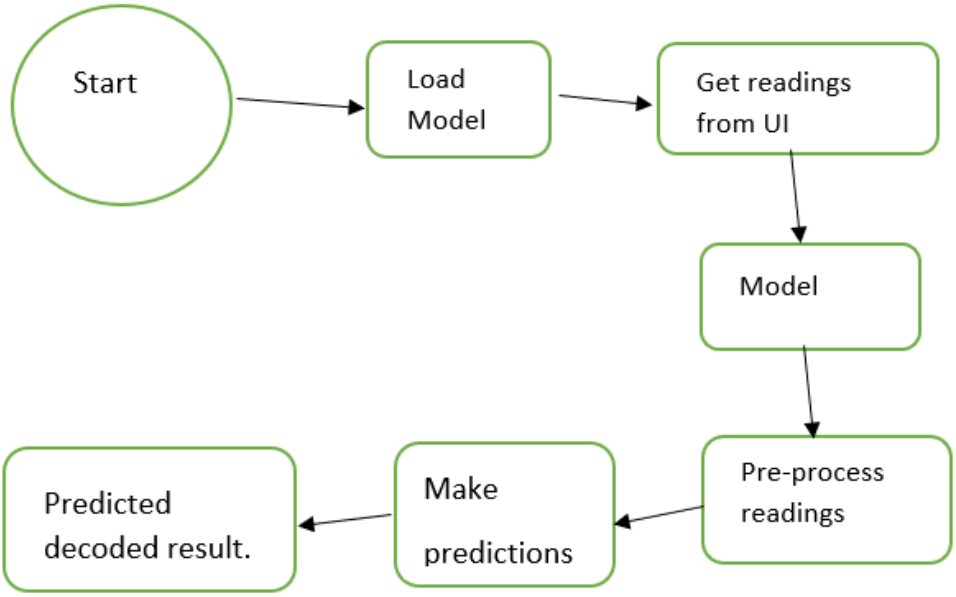

Start
Get Data from Resource
Convert to DataFrame
Data validation
Replacing Null values
Checking Data Drift
Dropping Unnecessary Columns
Data Transformation
Feature & Label Encoding
Handling Null values & Outliers
Handling Imbalanced Data
Hyperparameter Tuning
Model Training
Prediction & accuracy
Model Evaluation
Batch data from user
Validation & Transformation
Prediction
Displaying output
Calculating Accuracy
Comparing Current model & previous model
Pushing the better model to cloud
End

3. Pipeline


Training Pipeline
Configs
Artifacts
Data base Name
Collection Name
Test size
Data Ingestion
Feature Store file
Train file
Test file
Train file, Test file
Base file
Missing threshold
Data Validation
Validation Report
Train file
Test file
Train file, Test file
Data Transformation
KNN Imputer object
Target encoder object
Transformed train, test array.
Train array, Test array
Expected model Score.
Overfitting threshold
Model Trainer
Model object
F1 train score
F1 test score
Current & Previous Model, imputer, encoder
Change threshold!
Test file
Model Evaluation
Is model accepted?
Improved accuracy
Loading Model, Imputer, Encoder, from artifacts
Model Pusher
Saving models to artifacts and saved_models
Batch Prediction
File from Client
Following the entire pre-processing steps and Prediction
Validation Report
Prediction file

4. Deployment Architecture





5.Real Time Prediction