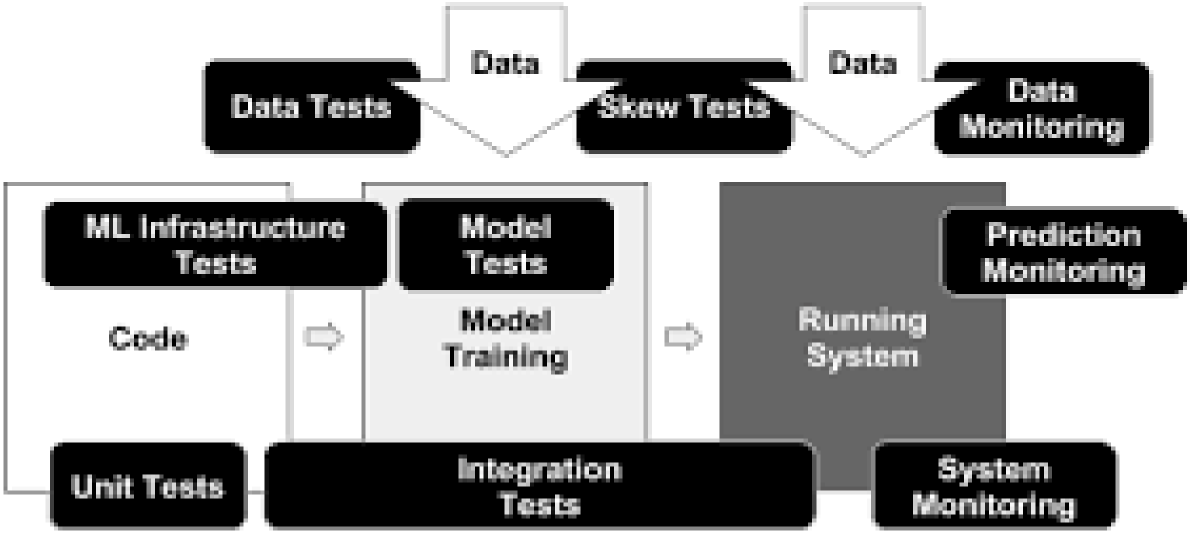
Machine Learning Model Deployment Project Documentation

PHASE 5: DOCUMENTATION



ML-Based System Testing and Monitoring

Project Objective

The objective of this project is to develop and deploy a machine learning model for real-time predictions. The model will be used for a predictive use case, and the project will encompass the following key components:

1.Use Case: Predicting customer churn for a subscription-based service.

2.Dataset: Selection and preparation of a historical customer data set.

3.Model Development: Training and evaluating a machine learning model to predict customer churn.

4.Model Deployment: Deploying the trained model to make real-time predictions.

5.Integration: Integrating the deployed model into the business application to enable seamless utilization for decision-making.

Design Thinking Process:

Understanding the Problem:

The initial phase of the project involved a thorough understanding of the problem statement - predicting customer churn. It was essential to comprehend the context and the factors contributing to churn in the subscription-based service.

Data Collection and Preparation:

The next step was to identify and obtain a relevant dataset that contains historical customer data, including customer attributes and churn labels. Data preprocessing was performed to clean, transform, and engineer features to make it suitable for model training.

Model Selection and Training:

Several machine learning algorithms were evaluated to select the best performing model. The dataset was split into training and testing sets for model training and evaluation. Hyperparameter tuning and cross-validation were performed to optimize the model's performance.

Model Deployment and Integration:

Once the model was trained and evaluated, the focus shifted to deploying the model in a production environment and integrating it into the business application. This phase involved ensuring the model's accessibility and scalability for real-time predictions.

Development Phases:

Predictive Use Case:

The use case involves predicting customer churn to proactively identify and retain at-risk customers in a subscription-based service.

Dataset Selection:

A historical customer dataset was selected, containing customer attributes such as demographics, subscription history, usage patterns, and churn labels. The dataset was preprocessed to handle missing values, encode categorical variables, and create relevant features.

Model Training

The selected model (e.g., Random Forest, Gradient Boosting) was trained using the preprocessed dataset. Hyperparameters were optimized, and the model's performance was evaluated using appropriate metrics like accuracy, precision, recall, and F1-score.

Model Deployment

The trained model was deployed using a containerized service (e.g., Docker) and hosted on a cloud platform (e.g., AWS, Azure, or GCP). This allowed the model to be accessible via an API endpoint.

Integration

The API endpoint was integrated into the business application, ensuring that it could receive input data and return real-time predictions. The integration included error handling, input validation, and logging for monitoring.

Real-Time Predictions

The deployed model can be accessed and utilized for real-time predictions through a RESTful API. The following steps outline how it can be used:

1.Input Data: Users provide customer data in a structured format to the API, including customer attributes.

2.API Request: Send an HTTP POST request to the model's API endpoint with the input data as a JSON payload.

3.Model Inference: The deployed model processes the input data and returns predictions in real-time.

4.Prediction Output: The API responds with the prediction results, indicating whether the customer is likely to churn.

5.Utilization: The business application can utilize these predictions to take appropriate actions, such as offering retention incentives to at-risk customers.

****