

High-Level Design (HLD) Document

Project: Adult census income Prediction

Document Version: 1.0

Date: 25-Sept-2023

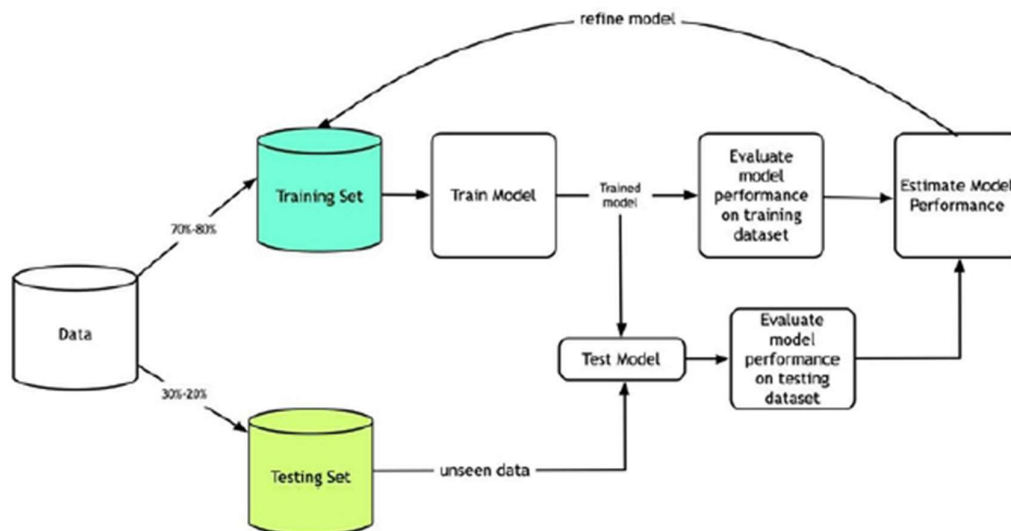
Author: Abhishek Pandit

Table of Contents

1. [Introduction](#)
2. [Design Overview](#)
3. [Architecture](#)
4. [Components](#)
5. [Data Structures](#)
6. [Functionality](#)
7. [Security](#)
8. [Performance](#)
9. [Deployment](#)
10. [Maintenance](#)
11. [Conclusion](#)

1. Introduction

The High-Level Design (HLD) document provides an overview of the architectural design and key components of the Machine Learning Income Prediction App. It outlines the system's structure, data flow, user interface, data structures, functionality, security measures, performance considerations, deployment strategies, and maintenance procedures.

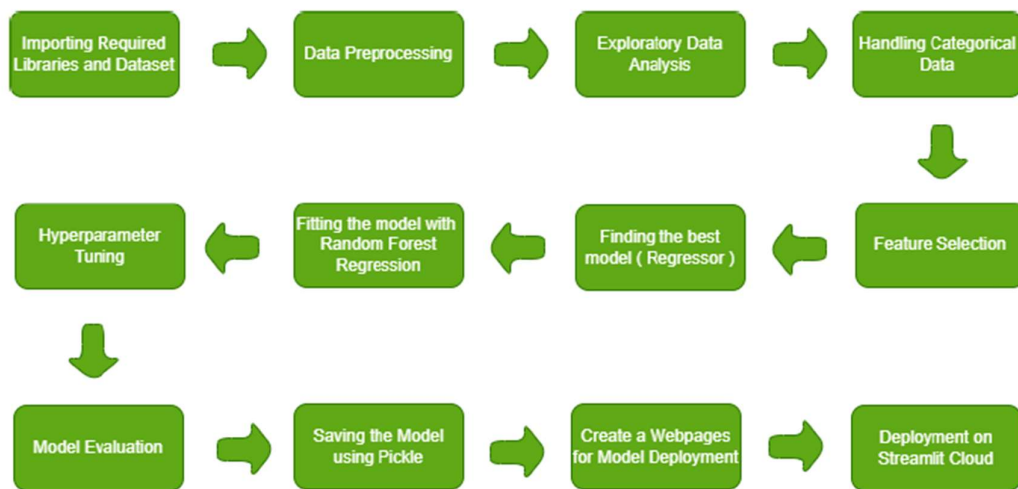


2. Design Overview

Objectives:

- Create a user-friendly income prediction application.
- Integrate machine learning models for accurate predictions.
- Ensure scalability, security, and maintainability of the system.

3. Architecture



Modular Architecture:

The application follows a modular architecture with the following components:

- Data Ingestion
- Data Transformation
- Model Training
- Prediction Pipeline
- User Interface (Streamlit)

Technology Stack:

- Programming Language: Python
- Libraries: pandas, scikit-learn, Streamlit

4. Components

Data Ingestion:

- Responsible for data loading and initial preprocessing.
- Reads data from external sources (e.g., CSV file).
- Ensures data quality and completeness.

Data Transformation:

- Performs feature engineering and data preprocessing.
- Includes pipelines for scaling, encoding, and imputing missing values.
- Prepares data for model training.

Model Training:

- Trains machine learning models using pre-processed data.
- Evaluates model performance and selects the best model.
- Stores the trained model for future predictions.

Prediction Pipeline:

- Handles user input and data preprocessing for predictions.
- Utilizes the trained model to make income predictions.
- Returns prediction results to the user interface.

User Interface:

- Developed using Streamlit for a web-based UI.
- Provides a user-friendly interface for data input and result display.
- Integrates seamlessly with other components.

5. Data Structures

Input Data:

- User input data as a DataFrame.

Preprocessing Objects:

- Preprocessor object for data transformation.

Model Objects:

- Trained machine learning model for income prediction.

6. Functionality

Data Ingestion (data_ingestion.py):

- Loads and preprocesses raw data.
- Ensures data quality and completeness.

Data Transformation (data_transformation.py):

- Performs data transformation and preprocessing.
- Includes pipelines for feature scaling, encoding, and imputation.

Model Training (model_trainer.py):

- Trains machine learning models on pre-processed data.
- Evaluates model performance and selects the best model.

Prediction Pipeline (prediction_pipeline.py):

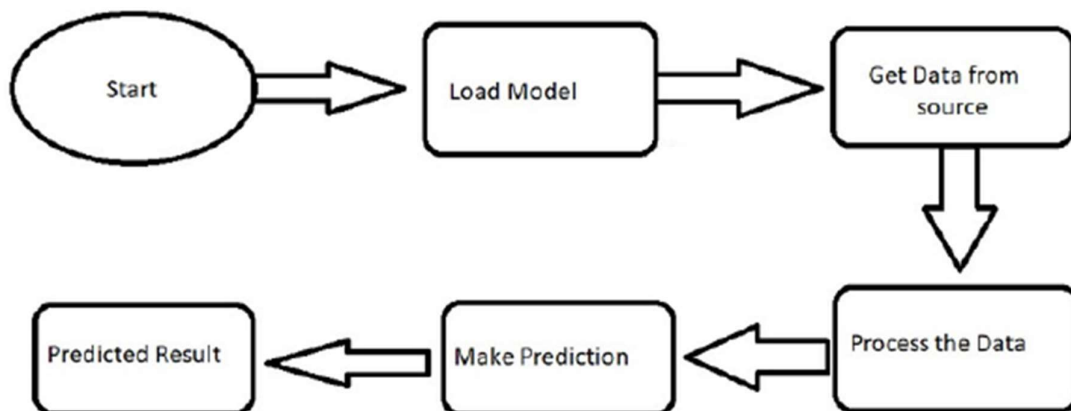
- Handles user input and data preprocessing.
- Makes income predictions using the selected model.

7. Security

- Ensures secure handling of user data.
- Implements access controls and authentication mechanisms.
- Protects the model and preprocessor objects from unauthorized access.

8. Performance

- Optimizes code and algorithms for efficient processing.
- Monitors application performance and responds to bottlenecks.
- Ensures responsive user interactions.

9. Deployment

- Supports deployment on various hosting platforms (e.g., Heroku, AWS, or Docker).
- Utilizes containerization for scalability and easy deployment.
- Implements CI/CD pipelines for automated deployment.

10. Maintenance

- Establishes a maintenance plan for regular updates and improvements.
- Monitors and addresses issues and bugs.
- Keeps dependencies up to date.

11. Conclusion

The High-Level Design (HLD) document provides an overview of the architectural design, components, data flow, user interface, data structures, functionality, security measures, performance considerations, deployment strategies, and maintenance procedures of the Machine Learning Income Prediction App. This document serves as a reference for the development and maintenance of the system.