High Level Design (HLD)

High Level Design (HLD)

[**Adult\_income\_Prediction**](https://github.com/mohiteyashprogrammer/Adult_income_Prediction/tree/main)

Revision Number – 1.0

Last Date of Revision – 26-07-2023

**ARUNKUMAR V**

1

High-Low Level Designing (LLD)(HLD)

2

High Level Design (HLD)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Contents** | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |
| [**Abstract**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.gjdgxs) | | | | | |  | | | | | | | | | | | | | | | | [4](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.gjdgxs) | | | |
|  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| [**INTRODUCTION**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.30j0zll) | | | | | | | | | |  | | | | | | | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.30j0zll) | | | |
|  |  |  |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | [**Why this HLD documentation?**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1fob9te) | | | | | | | | | | | | | | | | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1fob9te) | | |  |
|  |  |  |  | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| [**1 Description**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2et92p0) | | | | | | | |  | | | | | | | | | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2et92p0) | | | |
|  |  | |  |  |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | [**1.1**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.tyjcwt) | |  |  | [**Problem Perspective**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.tyjcwt) | | | | | | | | | | | | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.tyjcwt) | | |  |
|  |  | |  | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
|  | [**1.2 Problem Statement**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3dy6vkm) | | | | | | | | | | | | | | |  | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3dy6vkm) | | |  |
|  |  | | | | | |  | |  | |  |  |  |  |  |  |  | |  |  |  |  |  | |  |
|  | [**1.3 Proposed Solution**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1t3h5sf) | | | | | | | | | | | | | | |  | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1t3h5sf) | | |  |
|  |  | | | | | |  | |  | |  |  |  |  |  |  |  | |  |  |  |  |  | |  |
|  | [**1.4 Solution Improvements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.4d34og8) | | | | | | | | | | | | | | | | | | | | | [5](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.4d34og8) | | |  |
|  |  | | | | | |  | |  | |  |  |  |  |  |  |  | |  |  |  |  |  | |  |
|  | [**1.5 Technical Requirements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2s8eyo1) | | | | | | | | | | | | | | | | | | |  | |  | [6](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2s8eyo1) | |  |
|  | [**1.6 Data Requirements**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.17dp8vu) | | | | | | | | | | | | | | |  | | | | | | [6](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.17dp8vu) | | |  |
|  |  | | | | | |  | |  | |  |  |  |  |  |  |  | | | |  |  |  | |  |
|  | [**1.7 Tools Used**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3rdcrjn) | | | | | | | | | |  | | | | |  | | | | | | [7](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3rdcrjn) | | | |
|  |  | |  |  | | |  | |  | |  |  |  |  |  | |  | | | |  |  |  |  | |
|  | [**1.8**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.26in1rg) | |  | [**Constraints**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.26in1rg) | | | | | | | |  | | | | | | | | | | [7](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.26in1rg) | | | |
|  |  | | | | | |  | |  | | |  |  |  |  | |  | | | |  |  |  | | |
|  | [**1.9 Assumptions**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.lnxbz9) | | | | | | | | | | | | | | | | | | | | | [7](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.lnxbz9) | | | |
|  | |  | | | | |  | |  | | |  |  |  |  | |  | | | |  |  |  | | |
| [**2 Design Flow**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.35nkun2) | | | | | | | | |  | | | | | | | | | | | | | [7](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.35nkun2) | | | |
|  |  | | | | | |  | |  | | |  |  | |  | |  | | | |  |  | |  |  |
|  | [**2.1 Modelling Process**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1ksv4uv) | | | | | | | | | | | | | | | | | | | | | [8](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1ksv4uv) | | |  |
|  |  | | | | | |  | | | | |  |  | |  | |  | | | |  |  |  | |  |
|  | [**2.2 Deployment Process**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.44sinio) | | | | | | | | | | | | | | | | | | | | | [8](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.44sinio) | | |  |
|  | |  | |  | | |  | | | | |  |  | | | |  | | | |  |  |  | |  |
| **2.3** | | | | **Logging** | | | | | | | | | | | | | | | | | | 8 | | |  |
| **2.2** | | | | **Error Handling** | | | | | | | | | | | | | | | | | | 8 | | |  |
| [**3 Performance Evaluation**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.4i7ojhp) | | | | | | | | | | | | | | | | |  | | | | | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.4i7ojhp) | | |  |
|  | | |  |  | | |  | | | | |  |  | | | |  | | | |  |  |  | |  |
|  | [**3.1**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2xcytpi) | |  | [**Reusability**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2xcytpi) | | | | | | | |  | | | | |  | | | | | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2xcytpi) | | | |
|  |  | |  |  | | |  | | | | |  |  | | | |  | | | |  |  |  | |  |
|  | [**3.2**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1ci93xb) | |  | [**Application Compatibility**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1ci93xb) | | | | | | | | | | | | | | | | | | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.1ci93xb) | | |  |
|  | [**3.3**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3whwml4) | |  | [**Resource Utilization**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3whwml4) | | | | | | | | | | | | |  | | | |  |  | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.3whwml4) | |  |
|  |  | | | | | |  | | | | | |  | | | |  | | | | |  |  | |  |
|  | [**3.2 Deployment**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2bn6wsx) | | | | | | | | | | | | | | | |  | | | | | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.2bn6wsx) | | | |
|  |  | | | | | |  | | | | | |  | | | | | | | | |  |  | | |
| [**Conclusion**](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.qsh70q) | | | | | | | | | | | | | | | | | | | | | | [9](https://docs.google.com/document/d/1ZLlZEVdMuxYl3o0PziQjZcUbNmAcz-PR/edit#heading=h.qsh70q) | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

3

High Level Design (HLD)

**Abstract**

Predicting adult income is a complex task that involves analyzing various factors related to an individual's demographics, education, occupation, and other relevant variables. While I can provide you with some general information, please note that predicting income accurately for specific individuals can be challenging due to the wide range of factors involved and the inherent uncertainty of future outcomes.

4

High Level Design (HLD)

**1 Introduction**

**1.1 Why this High-Level Design Document?**

The main purpose of this HLD documentation is to feature the required details of the project and supply the outline of the machine learning model and also the written code. This additionally provides the careful description on however the complete project has been designed end-to-end.

**1.2 Description**

**Problem Perspective**

Predicting adult income is a complex task that involves analyzing various factors related to an individual's demographics, education, occupation, and other relevant variables. While I can provide you with some general information, please note that predicting income accurately for specific individuals can be challenging due to the wide range of factors involved and the inherent uncertainty of future outcomes.

**1.3 Problem Statement**

The objective of this project is to develop a predictive model that can accurately estimate the income level of adults based on various demographic, educational, and occupational features. The model should analyze a given individual's characteristics and provide a prediction of whether their income exceeds a certain threshold, such as $50,000 per year.

The problem involves leveraging historical data on adult income along with associated attributes to train the predictive model. The trained model will then be used to make income predictions for new individuals based on their input features.

The goal is to create an accurate and robust income prediction model that can assist in making informed decisions, such as resource allocation, policy planning, and financial assessments. The model should take into account factors such as education, occupation, work experience, demographics, and any other relevant attributes that may contribute to an individual's income level.

The success of the project will be evaluated based on the model's predictive performance, measured by appropriate evaluation metrics such as accuracy, precision, recall, or area under the receiver operating characteristic curve (AUC-ROC). The model should aim to achieve high accuracy and minimize false predictions.

The developed model should be applicable to diverse populations and exhibit fairness, ensuring that predictions are not biased or discriminatory based on factors such as gender, race, or ethnicity. Ethical considerations should be taken into account throughout the development process to ensure responsible and unbiased predictions.

Overall, the project aims to provide a reliable and interpretable income prediction model that can assist in understanding income disparities, identifying potential opportunities for improvement, and informing decision-making processes related to income distribution and economic policies.

**1.4. Project Solution**

Project requires the desired input of user from the created interface and method all the provided information to satisfy the wants of the machine learning model and at last show the expected output.

5

High Level Design (HLD)

**1.6 Technical needs**

There are not any hardware needs needed for victimization this application, the user should have AN interactive device that has access to the web and should have the fundamental understanding of providing the input. And for the backend half the server should run all the package that's needed for the process and provided information to show the results.

**1.7 Information needs**

The info demand is totally supported the matter statement. and also, the information set is accessible on the Kaggle within the type of standout sheet(.xlsx), because the main theme of the project is to induce the expertise of real time issues, we have a tendency to once more mercantilism {the information into the prophetess data base and commerce it into csv format.

**1.8Tools Used**

* Python 3.8 is employed because the programming language and frame works like numpy, pandas, sklearn and alternative modules for building the model.
* Visual Studio is employed as IDE.
* For visualizations seaborn and components of matplotlib are getting used.

6

High Level Design (HLD)

* For information assortment prophetess info is getting used.
* Front end development is completed.
* Streamlit is employed for each information and backend readying.
* GitHub is employed for version management.
* Azure is employed for deployment.

**1.9 Constraints**

The objective of this project is to develop a predictive model that can accurately estimate the income level of adults based on various demographic, educational, and occupational features. The model should analyze a given individual's characteristics and provide a prediction of whether their income exceeds a certain threshold, such as $50,000 per year.

**1.10 Assumptions**

The most objective of the project is to implement the utilization cases as for the new dataset that user provides through the programme. Machine learning model is employed for process the on top of computer file. It's additionally assumed that each one aspects of this project have the flexibility to figure along within the approach as the designer is expecting.



7

High Level Design (HLD)

**2.1 and 2.2 Design Flow and Deployment Process**



|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
|  |  | |  |
|  |  |



Web page using Streamlit

Deployed on Local Host by Streamlit

Hosted on Azure

**2.3 Logging**

Each step is being logged within the system that runs internally, that shows the date time and therefore the processed that has been performed, work is completed in several layers as information, DEBUG, ERROR, WARNINGS. this provides US the perceive of the logged info.

**2.4 Error Handling**

Once ever a slip is occurred, the reason are logged in its several log file, in order that the developer will rectify the error.

High Level Design (HLD)

**3 Performance analysis**

**3.1 Reusability**

Elements of the code written is accustomed different applications and therefore the rest is changed and be reused.

**3.2 Application Compatibility**

The various parts for this project are exploitation python as associate interface between them. every element can have its own tasks to perform, and it's the work of the python to make sure correct transfer of data.

**3.3 Resource Utilization**

Once any task is performed, it'll doubtless; use all the process power offered till that performs is finished.

**3.4 Deployment**

The model is being deployed on Azure.

**Conclusion**

The flight fare prediction will predict the worth supported the trained knowledge set within the rule. therefore, the user will recognize the approximate value for his or her journey.

9