**High Level Design(HLD)**

Adult Census Income Prediction

Revision Number :

Last Date of Revision :

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
|  | 1 | Abstract,  Introduction,  General Description | B. Sreenath Reddy |
|  | 2 | Design Detail ,KPI,  Deployment | B. Sreenath Reddy |
|  | 3 | Final Revision | B. Sreenath Reddy |

Contents

Document Version Control…………………………………………………………

Abstract…………………………………………………………………………………………………-

1.Introduction ………………………………………………………………………..-4

1.1- Why this High Level Design Document ?.............................-4

1.2- Scope……………………………………………………………………………….-4

1.3- Definations……………………………………………………………-5

2.General Description………………………………………………………………..-5

2.1-Product perspective……………………………………………………………………..-5

2.2-Problem Statement……………………………………………………………………….-5

2.3-Proposed Solution…………………………………………………………………………-5

2.4-Tools Used……………………………………………………………………………………-6

3.Design Details…………………………………………………………………………..-7

3.1-Process Flow ……………………………………………………………………………………-7

3.1.1-Model Training and Evaluation………………………………………………………..-8

3.2- Event Log……………………………………………………………………………………..-9

3.3-Error Handling……………………………………………………………………………….-9

4.Performance………………………………………………………………………….-9

4.1-Reusabilty………………………………………………………………………………………-9

4.2- Application Compatibility…………………………………………………………….-9

4.3- Deployment………………………………………………………………………………..-10

5.KPIs(Key Performance Indicators)…………………………………………-10

6.Conclusion…………………………………………………………………-11

Abstract:

The prominent inequality of wealth and income is a huge concern especially in the United States. The likelihood of diminishing poverty is one valid reason to reduce the world's surging level of economic inequality. The principle of universal moral equality ensures sustainable development and improve the economic stability of a nation. Governments in different countries have been trying their best to address this problem and provide an optimal solution. This study aims to show the usage of machine learning techniques in providing a solution to the income equality problem. The UCI Adult Dataset has been used for the purpose. Classification has been done to predict whether a person's yearly income in US falls in the income category of either greater than 50K Dollars or less equal to 50KDollars category based on a certain set of attributes

Introduction

* 1. Why this High Level Design Document :

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

* + Present all of the design aspects and define them in detail
  + Describe the user interface being implemented
  + Describe the hardware and software interfaces
  + Describe the performance requirements
  + Include design features and the architecture of the project
  + List and describe the non-functional attributes like:

Security

Reliability o Maintainability

* + - Portability
    - Reusability
    - Security
    - Maintainability

1.2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

* 1. Definitions :

|  |  |
| --- | --- |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |
| AWS | Amazon Web Services |

2.General Description:

2.1 Product Perspective:

Adult Census Income prediction (ACI) is a Machine Learning classification model-based modelling where we try to predict the income group of a person and take necessary actions.

2.2 Problem Statement :

* The Goal is to predict whether a person has an income of more than 50K a year or not.
* This is basically a binary classification problem where a person is classified into the >50K group or <=50K group.

2.3 PROPOSED SOLUTION:

The solution proposed here is ACI (Adult Census Income prediction) is used to perform above use cases, if there any person with less than 50k income it categorizes into less category and vice versa. This study helps in knowing the economy of the country and it helps in solving income equality problem

2.4. Tools Used

    

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib and Seaborn are used to build the whole model.

* + Jupyter Notebook is used as IDE.
  + For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
  + AWS is used for deployment of the model.

3.Design Details

3.1-Process Flow :

To predict whether a person has an income of more than 50K a year or not.

Proposed Methodology :

ACIP data set from Kaggle

Exploratory Data Analysis

Splitting Dataset into

Train and Test Datasets

Selecting Best Suitable Algorithm

Prediction

3.1.1- Modal Training and Evaluation :

Data Splitting Process

Data Pre-processing

Dataset

Evaluation Process

Deployment Process

Prediction

3.2- Event Log :

The system should log every event so that the user will know what process is running internally

Initial Step-By-Step Description:

1. The System identifies at what step logging required
2. The System should be able to each and every system flow
3. Developer can choose logging method. You choose database logging / File logging as well
4. System should not hang even after using so many logging. Logging just because we can easily debug issues so logging is mandatory to do

3.3-Error Handling :

Should errors be encountered, an explanation as to what went wrong . An error as anything that falls outside normal and intended usage.

4.Performance :

The adult census income prediction should be accurate in classification of salary of a person. So that it doesn’t mislead the authorities and users. And also Model retraining approach used t improve performance.

4.1-Reusabilty :

The code written and the components used should have the ability to be reused with no problems

4.2- Application Compatibility :

The different components for this project will be using python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3- Deployment :



5.KPIs(Key Performance Indicators) :

1. Key Indicators displaying a summary of Adult Census Income Prediction
2. To predict whether a person has an income of more than 50K a year or not.
3. It gives persons Year Income

6.Conclusion :

The solution proposed here is ACI (Adult Census Income prediction) will classify persons, and if there any person with less than 50k income it categorizes into low category and vice versa. This study helps in knowing the economy of the country and it helps in solving income equality problem