Low Level Design (LLD)

*Adult Census Income Prediction*

Revision Number: 1.0

Last date of revision: 10/08/2021

# 

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 10th Aug 2021 | 1.1 | First Draft | Devendra pitaliya |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_gjdgxs)

[**Abstract** 4](#_30j0zll)

[1](#_1fob9te) Introduction 4

[1.1](#_3znysh7) Why this Low-Level Design Document? 4

[1.2](#_2et92p0) Scope 5

[1.3](#_tyjcwt) Constraints 5

[1.4](#_3dy6vkm) 1.3 Risks 5

[1.5](#_1t3h5sf) 1.4 Out of Scope 5

[2](#_4d34og8) Technical specifications 6

[2.1 Prediction dataset 6](#_2s8eyo1)

[2.1.1 dataset overviwe 6](#_3rdcrjn)

[2.1.1 Input schema 6](#_3rdcrjn)

[2.2 predicting salary](#_2s8eyo1) 7

[2.3 Diployment](#_2s8eyo1) 7

[3](#_26in1rg) Technology stack 8

[4](#_35nkun2) Model training/validation workflow 8

[5](#_1ksv4uv) User I/O workflow 9

6 test cases 9

**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 and data mining 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 50K Dollar’s category based on a certain set of attributes.

# Introduction

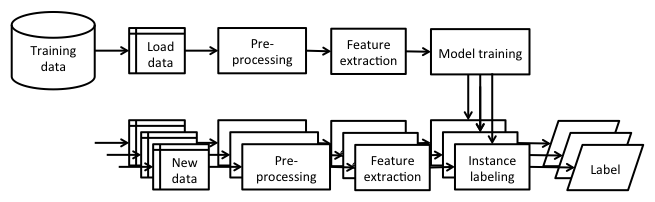
## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Deep EHR System. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The Goal is to predict whether a person has an income of more than 50K a year or not.

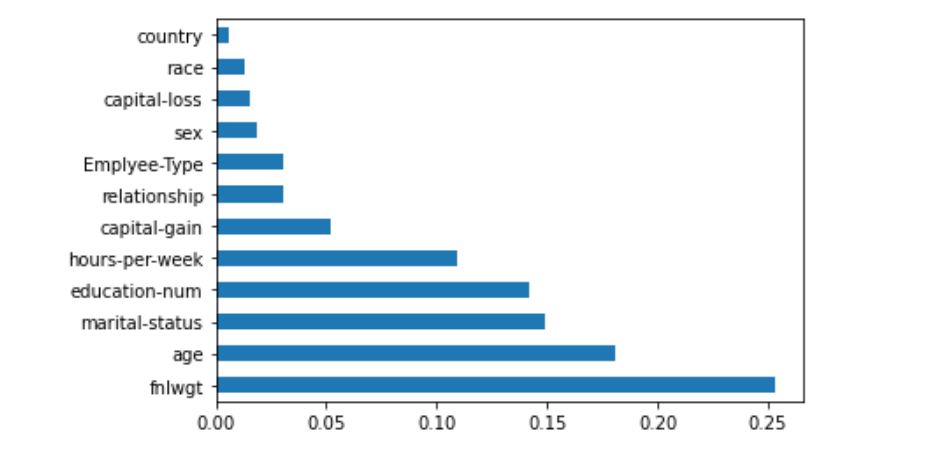
This is basically a binary classification problem where a person is classified into the

>50K group or <=50K group..



## Scope

This software system will be a Web application This system will be designed to predict the 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.The data set includes figures on 32,561 different records and 14 attributes for 42 nations. The 14 attributes consist of 8 categorical and 6 continuous attributes containing information on age, education, nationality, marital status, relationship status, occupation, work classification, gender, race, working hours per week, capital loss and capital gain.



## Constraints

We are selecting all the columns.

## Risks

Document specific risks that have been identified or that should be considered.

## Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

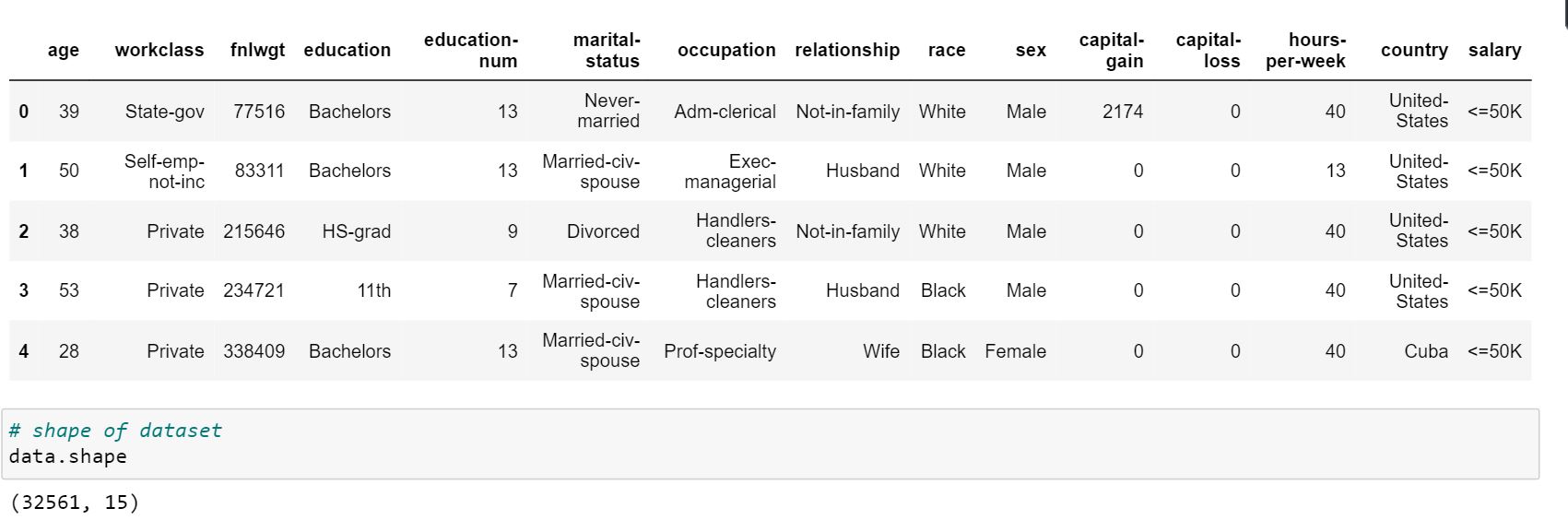
# Technical specifications

## 2.1 Dataset

|  |  |  |
| --- | --- | --- |
| **dataset** | **Finalized** | **Source** |
| Adult Census Income Prediction | yes | https://www.kaggle.com/overload10/adult-census-dataset |

## 2.1.1 dataset overview

* The data set includes figures on 32,561 different records and 14 attributes for 42 nations. The 14 attributes consist of 8 categorical and 6 continuous attributes containing information on age, education, nationality, marital status, relationship status, occupation, work classification, gender, race, working hours per week, capital loss and capital gain.



## 2.1.2 Input schema

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature name** | **Datatype** | **Size** | **Null/Required** |
| Age | int | 3 | Required |
| **Fnlwgt (Final weight)** | Int | Max (5) | Required |
| **education-num** | int | 2 | Required |
| **marital-status** | int | 1 (0 or 1) | Required |
| **relationship** | int | 1 | Required |
| **race** | int | 1 | Required |
| **sex** | int | 1 (0 or 1) | Required |
| **capital-gain** | int | 1 (0 or 1) | Required |
| **capital-loss** | int | 1 (0 or 1) | Required |
| **hours-per-week** | int | 2 | Required |
| **country** | int | 1 | Required |
| **Employee-Type** | int | 1 | Required |

## 2.2 Predicting salary

* The system displays the all the field.
* The system presents the set of inputs required from the user.
* The user gives required information.
* The system should be able to predict whether it is >50K or <50K

**2.3 Deployment**

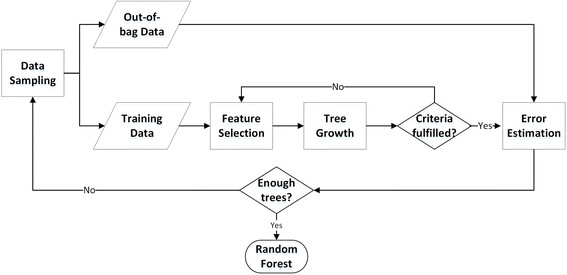
1. AWS



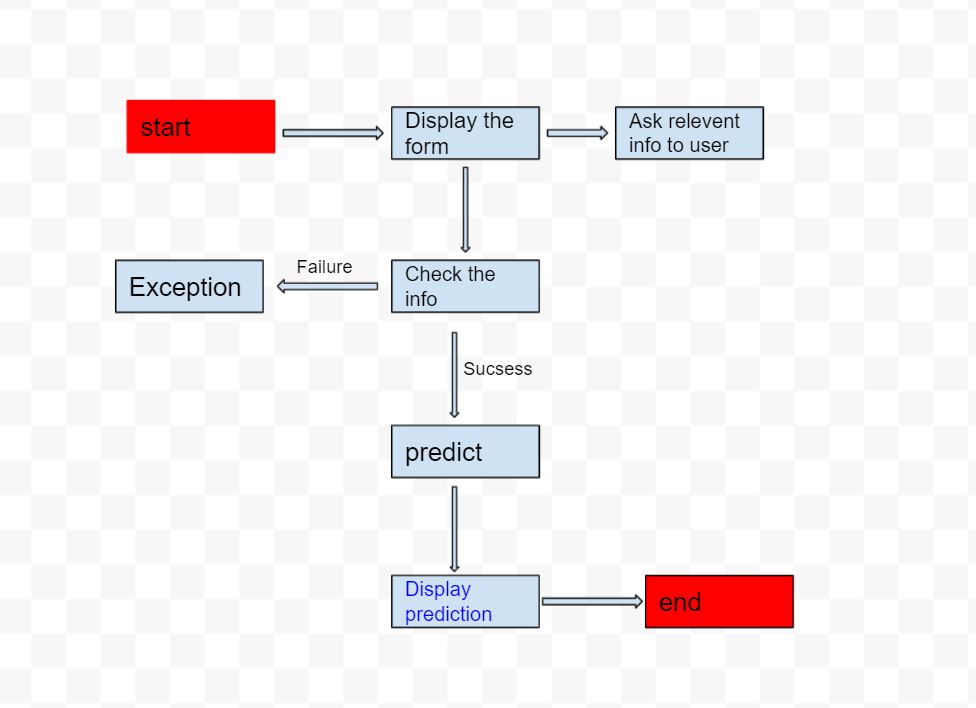
# Technology stack

|  |  |
| --- | --- |
| **Front End** | HTML/CSS |
| **Backend** | Python Django |
| **Deployment** | AWS |

# Model training/validation workflow



# User I/O workflow



# Test cases

|  |  |  |  |
| --- | --- | --- | --- |
| Test case | Steps to perform test case | Module | Pass/Fail |
| 1 | Enter the value | Should take appropriate values | pass |
| 2 | Null field | Should not accept null field | pass |
| 3 | Predict button | User should able to click on button | pass |