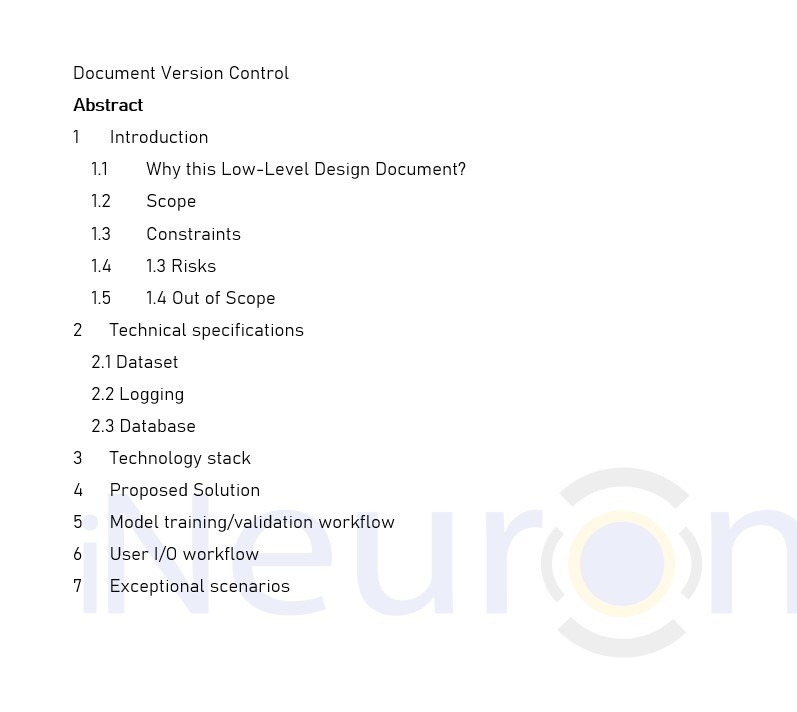
Low Level Design (LLD)

# Adult Census Income Prediction

Umang Tank

## Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 20th Aug 2021 | 1.1 | Initial HLD-V1.0 | Umang Tank |
| 12th Aug 2021 | 1.2 | update | Umang Tank |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |



## 1Introduction

1.1Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Adult Census income prediction. 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.



An

Adult Census Income

contains

the

information, such as:

▪

Age

▪

Cap

ital

-

gain

▪

C

apital

-

loss

▪

E

ducation

-

year

▪

W

orking

-

hours

▪

Sex

▪

R

ace

* Country
* Occupation
* Work-class
* Education
* Merital-status
* relation

This project shall be delivered in two phases:

Phase 1: All the functionalities with Scikit-learn packages.

Phase2: Integration of UI to all the functionalities.

### 1.2Scope

This software system will be a Web application This system will be designed to predict whether income of people is more than 50K or not. . The models can be applied to the data collected in coming years to predict the income. This system is designed to predict the income based on some information like age, working-hours, sex, race, occupation ,capital-gain/loss etc.

### 1.3Constraints

Adult Census Income prediction must be user friendly, as automated as possible and users should not be required to know any of the workings.

1.4Risks

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

|  |  |
| --- | --- |
| Name | Description |
| Age | Age |
| Work Class | : Working Class (Private, Self-emp-not-inc, Self-emp-inc, Federalgov, Local-gov, State-gov, Without-pay, Neverworked) |
| Education level | Level of Education (Bachelors, Some-college, 11th, HS-grad,  Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th,  Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool) |
| Education-num | : Number of educational years completed |
| Marital status | Marital status (Married-civ-spouse, Divorced, Never-married,  Separated, Widowed, Married-spouse-absent, Married-AFspouse) |
| Occupation | Work Occupation (Tech-support, Craft-repair, Other-service,  Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machineop-inspct, Adm-clerical, Farming-fishing, Transportmoving, Priv-houseserv, Protective-serv, Armed-Forces) |

### 1.5Out of Scope



Delineate specific activiti

es, capabilities, and items that are out of scope for the

project.

2

Technical specifications

2.1

Dataset

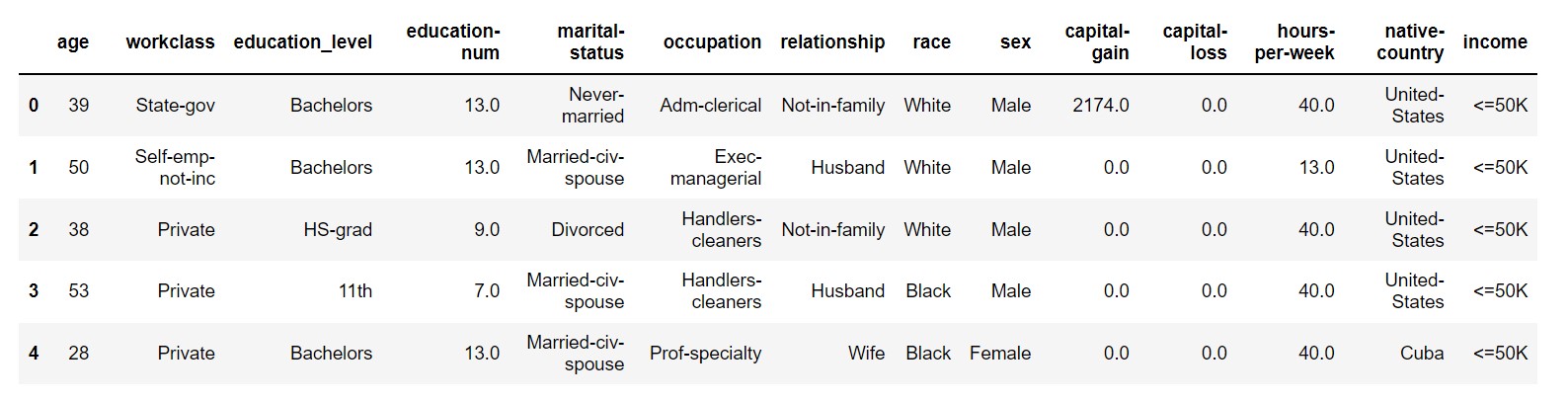
|  |  |
| --- | --- |
| Relationship | :    Relationship Status (Wife, Own    -    child,    Husband, Not    -    in    -    family,    Other    -    relative, Unmarried) |
| race | Race (White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black) |
| sex | Sex (Female, Male) |
| Capital-gain | Monetary Capital gains |
| Capital-loss | Monetary Capital Losse |
| Hours-per week | Average Hours Per Week Worked |
| Native-  Country | Native Country (United-States, Cambodia, England, PuertoRico,  Canada, Germany, Outlying-US(Guam-USVI-etc), India,  Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines,  Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland,  France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti,  Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand,  Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-  Netherlands) |



#### 2.1.1Insurance Premium dataset overview

To create the income prediction model, we obtained the data set through the Kaggle site. The data set includes 14 attributes, the data set is separated into two-part the first part called training data, and the second called test data; training data makes up about 80 percent of the total data used, and the rest for test data The training data set is applied to build a model as a predictor of Income. the test set will use to evaluate the Classification model.

Some of the records in the dataset are following



#### 2.1.2 Input schema



Feature name

Datatype

Size

Null/Requi

red

Age

int

2

Required

### 2.2 Predicting income

● The system displays the form where the all features are available if user can all features correctly then machine can able to predict the income of that person.

### 2.3 Logging

We should be able to log every activity done by the user.

* The System identifies at what step logging required
* The System should be able to log each and every system flow.
* Developers can choose logging methods. You can choose database logging/ File logging as well.
* System should not be hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

### 2.4 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User gives required information.
2. The system stores each and every data given by the user or received on request to the database. Database you can choose your own choice whether MongoDB/ MySQL. Here we use MySQL.

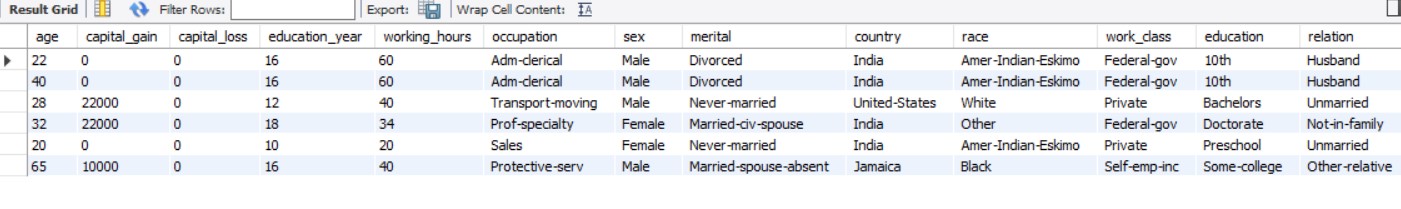


2.5

Deployment

1.

FLASK



## 3Technology stack

|  |  |
| --- | --- |
| Front End | HTML/CSS/JS/React |
| Backend | Python Flask |
| Database | MySql |
| Deployment | Heroku |

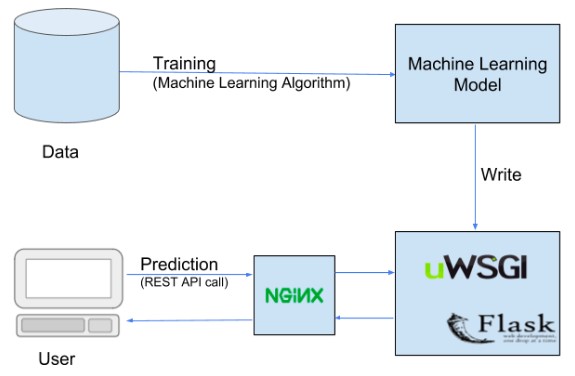


## 4Proposed Solution

AdaBoostClassifier gives better accuracy as compare to other so in this project we use AdaBoostClassifier algorithm to predict income. However, drawing a baseline in the form of some Machine Learning algorithm would be helpful.

1. Actual model: AdaBoostClassifier

## 5Model training/validation workflow



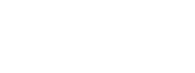
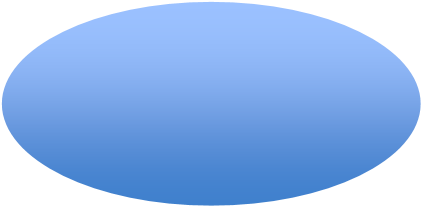
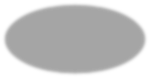
## 6User I/O workflow



Start



Prediction



Fill the all

information

