

High Level Design (HLD)

CENSUS INCOME PREDICTION

Document Version Control

DATE	DESCRIPTION	AUTHOR
20-06-2023	HLD- Version 1.0	SHREYASH VIRENDRA CHAWDA

- Introduction

Machine Learning Project for predicting person income status. Based on different parameters model will be trained and will predict the salary status of the person which is greater or lesser than average salary.

- 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
- Maintainability
- Portability
- Reusability
- Application compatibility
- Resource utilization
- Serviceability
- Scope

The HLD documentation presents the structure of the system 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.

- **Product Perspective**

The census income prediction is machine-learning model which helps us to predict person income status based on different parameters.

- **Problem statement**

For good development of the country and growth of the society it is mandatory to have the information of the economic and social condition of the citizen. For that collecting census will give broad ideas of individual status. Different parameters like social condition economic condition like employment, revenue generation education etc. plays important role for knowing individuals satisfaction and lifestyle. If we can predict individual income based on census data and different parameters we will understand the needs of individual how much they can afford the decent lifestyle or helps us to provide necessary amenities to needy for developing good and healthy society.

- **PROPOSED SOLUTION**

To understand the social-economical condition of the society this model will help to understand the income status of the individual. Which will help to calculate GDP, GNP, PPP etc.

- **FURTHER IMPROVEMENTS**

Census income prediction project need to train the model by collecting each year data which will give the more reliable and accuracy for prediction income status. Economy condition like previous GDP, GNP, demand and supply chain also consider and use to train the model with more independent features will give better prediction.

- **Data**

Dataset containing csv file has information which are require for census. This data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry Becker (Data Mining and Visualization, Silicon Graphics). A set of reasonably clean records was extracted using the following conditions: ((AAGE>16) && (AGI>100) && (AFNLWGT>1) && (HRSWK>0)). *The prediction task is to determine whether a person makes over \$50K a year.*

- Tools used

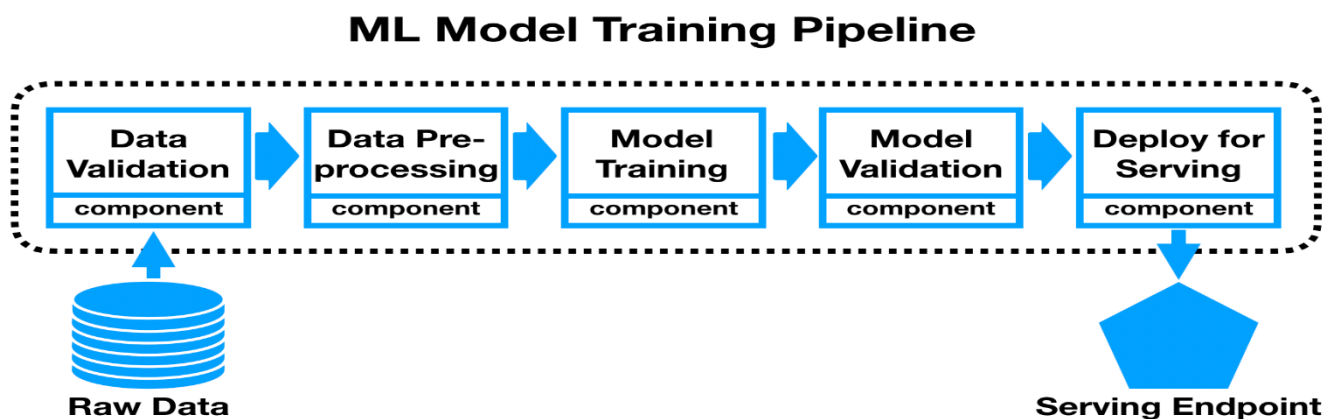
1. VS code is used as IDE.
2. Pandas, Numpy is used for data wrangling.
3. Scikit learn library is used for pipeline and model building.
4. For visualization of the plots, Matplotlib, Seaborn are used.
5. Docker is used for deployment.
6. Front end development is done using HTML.
7. GitHub is used as version control system.

- Design Details

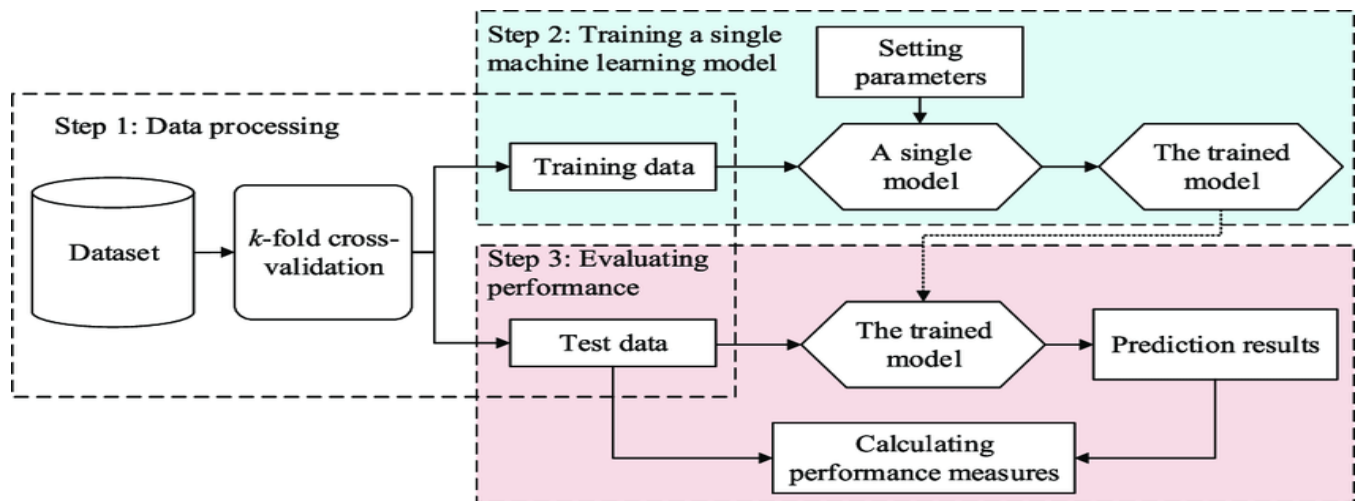
Process Flow

For identifying the different types of anomalies, we will use a machine learning base model.
Below is the process flow diagram is as shown below.

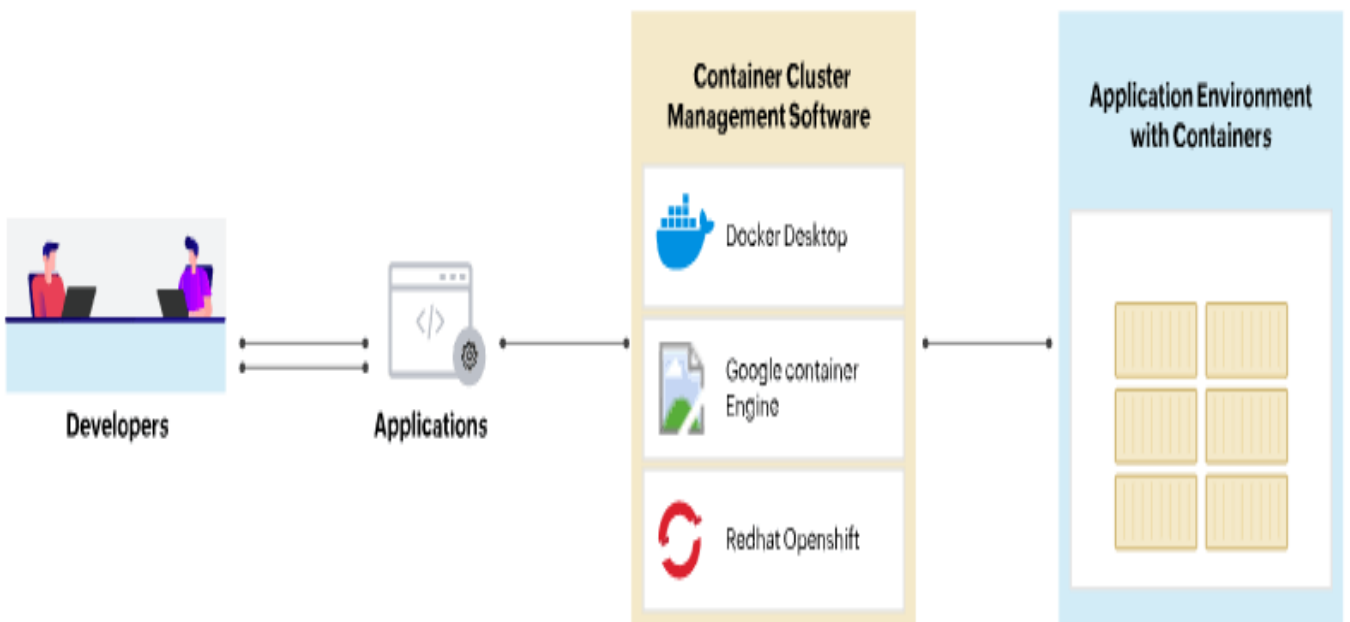
Proposed Model



- Model Training and Evaluation



- Model Deployment



- Event log

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

Initial Step-By-Step Description:

The System identifies at what step logging required

The System should be able to log each and every system flow.

Developer can choose logging method. You can choose database logging/ File logging as well.

System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

- Error Handling

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

- Performance

The census adult income prediction model contains classifier algorithm which is giving good accuracy. This model will helps to understand economic condition of the individuals.

- Reusability

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

- 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.

- Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

- KPIs (Key Performance Indicator)

1. Keep the track and predict that individual's income status.
2. Helps to analysis economic and social condition of the society.
3. Helps in to analysis the growth of individual.
4. Helps to boost the society development by providing facility to needy person.
5. Helps to give broad ideas regarding GDP, GNP, Per Capita Income, and PPP.

