**High level design**

**HLD**

**High Level Design (HLD)**

Finance (Adult Census Income Prediction)

**Contents:**

Document Version Control......................................................................................................04

Abstract……………………………………………………………………………………………………………………………..05

1 Introduction……………………………………………………………………………………………………………………06

1.1 Why this High-Level Design Document?.......................................................................06

1.2 Scope..............................................................................................................................06

1.3. Definitions……………………………………………………………………………………………………………….06

2 General Description.............................................................................................................07

2.1 Problem Perspective………………………………………………………………………………..…………………07

2.2 Problem statement………………………………………………………………………………………………….…07

2.3 Proposed Solution……………………………………………………………………………………………………..07

2.5 Further Requirement…………………………………………………………………………………………………07

2.5 Technical Requirements…………………………………………………………………………………………….07

2.6 Data Requirements……………………………………………………………………………………………………08

2.7 Tools Used…………………………………………………………………………………………………………………08

2.8 Constraints………….……………………………………………………………………………..…………….………09

2.9 Assumptions…….………………………………………………………………………………………………...……09

3 Design Details……………………………………………………………………………………………………….……….10

3.1 Process Flow....................................................................................................................10

3.2 Deployment Process………………………………………………………………………………………………….10

3.3 Event log……………………..…………………………………………………………………………………………….11

3.4Error Handling…………………………………………………………………………………………………………….11

4 Performance.........................................................................................................................11

4.1 Reusability........................................................................................................................11

4.2 Application Compatibility…………………………………………………………………………………………...11

4.3 Resource Utilization……………………………………………………………………………………………………11

4.4 Deployment......................................................................................................................11

5 Dashboards............................................................................................................................12

5.1 KPIs (Key Performance Indicators) …………………………………………………………………………..…12

6 Conclusion………………………………………………………………………………………………………………………13

**Document Version Control :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 27 Oct 2022 | 1 | First Draft | Rutuja Bankar |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Abstract:**

Finance, of financing, is the process of raising funds or capital for any kind of expenditure. It is the process of channelling various funds in the form of credit, loans, or invested capital to those economic entities that most need them or can put them to the most productive use. Machine learning is a branch of artificial intelligence that uses statistical models to make predictions. In finance, machine learning algorithms are used to detect fraud, automate trading activities, and provide financial advisory services to investors.

1. **Introduction:**
   1. **Why this High-Level Design Document?**

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.

* 1. **Scope:**

The HLD documentation presents the structure of the system, such as the database Estimating Financial Requirements, Deciding Capital Structure, Selecting Source of Finance, selecting a Pattern of Investment, Proper Cash Management, Implementing Financial Control, and Proper use of Surpluses.

* 1. **Definitions:**

In finance, machine learning algorithms are used to detect fraud, automate trading activities, and provide financial advisory services to investors.

**2 General Description:**

**2.1 Product Perspective**:

supervised learning, unsupervised learning, semi-supervised learning & reinforcement learning. The type of algorithm data scientists chooses to use depends on what type of data they want to predict. This application is a machine learning model which will help us to predict whether the person income is more than 50k or not.

**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 groups.

**2.3 Problem Solution:**

Machine learning can help businesses improve products or speed up manual and time-consuming processes. Managing an increasing number of online customer interactions has become a pain point for most businesses. The concept of demand forecasting is used in multiple industries, from retail and e-commerce to manufacturing and transportation.

* 1. **FURTHER IMPROVEMENTS:**

1. Assessment
2. Redesign
3. Implementation
4. Optimization
5. Strategic planning

* 1. **Technical Requirements:**

In this project, the requirements are to get personal income through various platforms. For that, in this project, we are going to use different technologies. Here is some requirements for this project.

* Model should be exposed through API or user interface, so that anyone can test the model.
* Cassandra database should be integrated in this project for any kind of user input.
* Model should be deployed on cloud (Azure, AWS, GCP).
  1. **Data Requirements:**

Data Requirement completely depends on our problem.

* For training and testing the model, we are using the adult census income dataset from the UCI machine learning repository.
* From the user we are taking the following input :
* Age
* Work class
* Education
* Marital-Status
* Occupation
* Relationship
* Race
* Sex
* Country
* Capital Gain
* Capital Loss
* Hours-per-week
  1. **Tools used:**

#### Python programming language and frameworks such as NumPy, Pandas, TensorFlow and Jupyter Notebook are used to build the whole model.

 

 

* 1. **Constraints:**

The income prediction system must be user-friendly, and error-free and users should not be required to know any of the back-end work.

* 1. **Assumptions:**

The main aim of this application is to predict whether the person’s income is more than 50k or not. So it is also assumed that whenever new data will be loaded the model will give an accurate result.

1. **Design Details:**

**3.1 Process Flow**

Feature Selection

Feature Engineering

Data Cleaning

Data Collection

Model Training

Feature Scaling

Train Test Split

Handling Categorical

Variable

Model Deployment

Model Testing

Hyperparameter

Tuning

Model Testing

**3.2 Deployment Process:**

.



**:**

**3.3 Event log**

Systems are the source of data for the discovery of the business processes that make up a company. The system should log every event so that the user will know what process is running internally.  
The system should be able to log each and every system flow. The developer can choose a logging method. You can choose database logging/ File logging as well. The system should not hang even after using so many loggings. Logging is just because we can easily debug issues so logging is mandatory to do.

**3.4 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.

**4. Performance**

Financial performance is a company's ability to use assets from its primary mode of business to generate revenues. It is also used as a general measure of a firm's overall financial health over a given period.

**4.1 Reusability**

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

**4.2 Application Compatibility**

Each component will have its own task to perform, and it is the job of Python to ensure the proper transfer of information between them.

**4.3 Resource Utilization**

When a computer is being used for an important task, it will be using all the processing power available to it at that time

**4.4 Deployment**



**5. Dashboards:**

As and when the system starts to capture the historical/periodic data for a user, dashboards will be included to display charts over time with progress on various indicators or factors. Dashboards will also indicate certain KPIs and relevant indicators for the unveiled problems that if not addressed in time could cause catastrophes.



* 1. **KPIs (Key Performance Indicators)**

### Marketing

### **Website Traffic**

### [Social Media](https://www.investopedia.com/terms/s/social-media.asp)**Traffic**

### **Conversation Rate on Call-To-Action Content**

### **Blog Articles Published Per Month**

### **Clickthrough Rates**

### IT

### **Total System Downtime**

### **Number of Tickets/Resolutions**

### **Number of Developed Features**

### **Count of Critical Bugs**

### **Back-up Frequency**

### Sales

### **Customer Lifetime Value (CLV)**

### [Customer Acquisition Cost](https://www.investopedia.com/terms/a/acquisition-cost.asp)**(CAC)**

### **Average Dollar Value for New Contracts**

### **Average Conversion Time**

### **Number of Engaged Leads**

### Staffing

### [**Absenteeism**](https://www.investopedia.com/terms/a/absenteeism.asp) **Rate**

### **of Overtime Hours Worked**

### **Employee Satisfaction Number**

### **Employee Turnover Rate**

### **Number of Applicants**

### 6.Conclusion:

I hope you have now gained some knowledge of what is Data Analysis, how to build Machine learning models, and how python helps us to build those models. This article may be the first step of your Machine Learning journey but, remember, "The secret of getting ahead is getting started".