Low Level Design

LLD

Low Level Design (LLD)

Finance (Adult Census Income Prediction)

**Contents**

[Document Version Control 2](file:///C:\Project\Documents\LLD_UGV.docx#_gjdgxs)

[**Abstract** 4](file:///C:\Project\Documents\LLD_UGV.docx#_30j0zll)

[1](file:///C:\Project\Documents\LLD_UGV.docx#_1fob9te) Introduction 5

[1.1](file:///C:\Project\Documents\LLD_UGV.docx#_3znysh7) Why this Low-Level Design Document? 5

[1.2](file:///C:\Project\Documents\LLD_UGV.docx#_2et92p0) Scope 6

[1.3](file:///C:\Project\Documents\LLD_UGV.docx#_tyjcwt) Constraints 6

[1.4](file:///C:\Project\Documents\LLD_UGV.docx#_3dy6vkm) 1.3 Risks 6

[1.5](file:///C:\Project\Documents\LLD_UGV.docx#_1t3h5sf) 1.4 Out of Scope 7

[2](file:///C:\Project\Documents\LLD_UGV.docx#_4d34og8) Technical specifications 7

[2.1 Predicting Disease 7](file:///C:\Project\Documents\LLD_UGV.docx#_2s8eyo1)

[2.2 Logging 7](file:///C:\Project\Documents\LLD_UGV.docx#_17dp8vu)

[2.3 Database 7](file:///C:\Project\Documents\LLD_UGV.docx#_3rdcrjn)

[3](file:///C:\Project\Documents\LLD_UGV.docx#_26in1rg) Technology stack 8

[4](file:///C:\Project\Documents\LLD_UGV.docx#_lnxbz9) Proposed Solution 9

[5](file:///C:\Project\Documents\LLD_UGV.docx#_35nkun2) Model training/validation workflow 10

[6](file:///C:\Project\Documents\LLD_UGV.docx#_1ksv4uv) User I/O workflow 12

[7](file:///C:\Project\Documents\LLD_UGV.docx#_44sinio) Exceptional scenarios 13

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 22nd Oct 2022 | 1.1 | First Draft | Rutuja Bankar |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Abstract**

**We study how researchers can apply machine learning (ML) methods in finance. We first establish that the two major categories of ML (supervised and unsupervised learning) address fundamentally different problems than traditional econometric approaches. Then, we the current state of research on ML in finance and identify three approaches types of applications: i) the construction of superior and novel measures, ii) the reduction of prediction error, and iii) the extension of the standard econometric toolset. With this taxonomy, we give an outlook on potential future directions for both researchers and practitioners. Our results suggest large benefits of ML methods compared to traditional approaches and indicate that ML holds great potential for future research in finance.**

# Introduction

## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of internal design of the actual program Coad. 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 main objective of the project is to predict if a person has an income more than $ 50K a year or not.

Receiving loans at any age for a particular reason has become a cumbersome task:

* Dataset contains personal information about individuals with their country and income.
* Automate and streamline provider workflow

An [adult census record](https://www.healthit.gov/providers-professionals/learn-ehr-basics) contains adults personal as wells as financial information, such as:

* Work Class
* fnlwgt
* Education
* Education in number of years
* Marital Status
* Occupation
* Relationship
* Race
* Sex
* Capital Gain
* Capital Loss
* Working Hours per Week
* Country
* Salary

This project shall be delivered in two phases:

Phase 1: All the functionalities with PyPi packages.

Phase2: Integration of UI to all the functionalities.

## **Scope**

This software system will be a model evaluation and keeping the best model in the system to predict the outcomes.

* 1. Constraints

We will only be predicting the model on the limited number of features, additional number of features can increase model’s robustness.

## **Risks**

 The possibility of losing prospective client if system is unable to predict the right candidate who is eligible for the loans.

## **Out of Scope**

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

**2.Technical specifications**

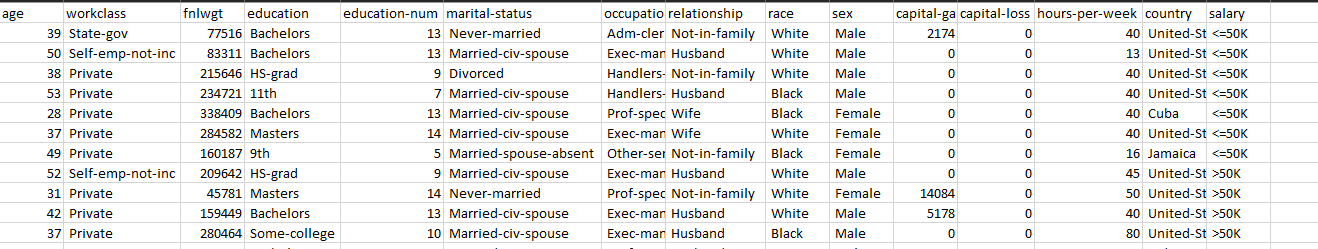
## **2.1 Dataset**



## **2.1.1 Adul Census dataset overview**

## Consists of single rectangular dataset with 32562 records, 14 feature columns and one target column. In which Age is a Numerical Discrete Variable, Work Class is a Categorical Nominal Variable, fnlwgt is a Numerical Continuous Variable, Education is a Categorical Nominal Variable, Education Num is Numerical Discrete Variable, Marital Status is a Categorical Nominal Variable, Occupation is a Categorical Nominal Variable, Race is a Categorical Nominal Variable, Sex is a Categorial Binary Variable, Capital Gain and Loss is a Numerical Continuous Variable, Hours per week is a Numerical Discrete Variable, Country is a Categorical Nominal Data and Target Variable is a Categorical Binary Variable.

* Adult Data Table



**1.2 Input schema**

## **2.2 Predicting**

* The system displays a binary classification if an individual has a salary more than $ 50K or not.

## **2.3 Logging**

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

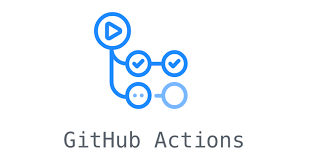
* The System is an immutable , time stamped record of discrete event that happened over time
* The System should be able to log each and every system flow.
* This system is easy to generate ;it’s great whenit come to providing valuable insight , it’s focus on specific event.
* System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do

## **2.4 Database**

financial datasets, a machine learning model might be able to predict the behavior of a given asset. That’s why the financial sector is doing everything in its power to create an effective ML model, as anything that can predict even reasonably well has the potential to generate millions of dollars. Machine learning is already predicting the behavior of citizens, which is impacting the way policy makers are doing their jobs.

* 1. **Deployment**

1. Git Action 2. AWS Lambda 3. AWS



**3. Technology stack**

|  |  |
| --- | --- |
| **Front End** | HTML/CSS/JS/React |
| **Backend** | Python Django |
| **Database** | MongoDB/MySql/Cassandra |
| **Deployment** | AWS |

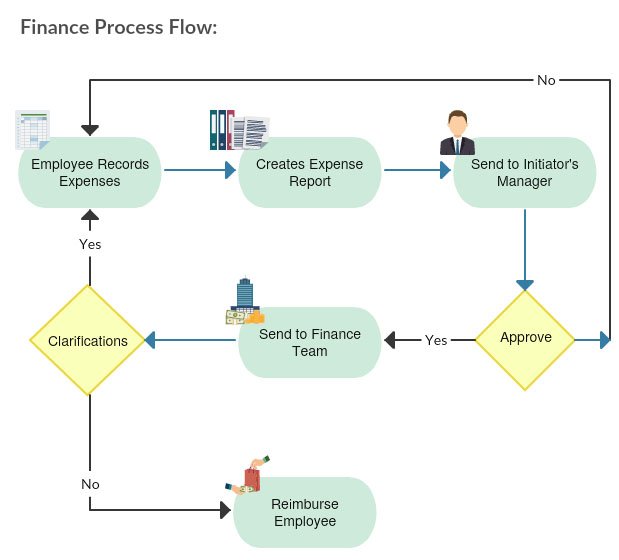
# 4.Proposed Solution

Based on the actual research paper, considering Logistic Regression as base model will help to identify the accuracy and further cross verifying with Random Forest will help to made the robust model if necessary.

1. Baseline Model: Logistic Regression, since this is a classification problem.

2. Random Forest

# Model training/validation workflow



# 6.User I/O workflow

# 

# Exceptional scenarios

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Exception | Mitigation | Module |
| 18th May 2020 | 1.1 | First Draft | Amit K Gupta |
| 20th May 2020 | 1.2 | Added Workflow chart | Amit K Gupta |

# Test cases

# Key performance indicators (KPI)

* Determine whether you’re on track to reach your financial goals
* Evaluate the success of your strategy
* Pinpoint areas in your business that may need improvement
* Identify any opportunities and challenges
* Assess whether your customers are happy or not