**HIGH -LEVEL DESIGN (HLD)**

**US -VISA- APPROVAL PREDICTION**

Revision Number - 1.0

Last Date of Revision :21-03-2024

ANNA ELSA LUIZ

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 21-03-2024 | 1.0 | First Draft | ANNA ELSA LUIZ |

# **Contents**

Document Version Control

Abstract

1. Introduction
   1. Why this HLD Document ?
   2. Scope
   3. Definitions
2. General Description
   1. Problem Perspective
   2. Problem Statement
   3. Proposed Solution
   4. Tools Used
3. Design Flow
   1. Model Creation and Evaluation
   2. Event Logging
   3. Error Handling
   4. Database
   5. Key Performance Indicators
   6. Deployment Process
4. Performance Evaluation
   1. Reusability
   2. Application Compatibility
   3. Resource Utilization
5. Conclusion

**Abstract**

This study proposes a predictive modeling framework to enhance the accuracy and efficiency of U.S. visa approval outcomes, leveraging advanced machine learning algorithms. Utilizing a comprehensive dataset encompassing historical visa application records, demographic, socio-economic, and geopolitical factors, the model aims to identify patterns associated with successful approvals. Through feature engineering and validation, algorithms like logistic regression and random forests optimize prediction performance, fostering transparency and accountability in the adjudication process. By empowering stakeholders with real-time insights through a user-friendly interface, our framework facilitates evidence-based decision-making, streamlining the visa application process while promoting procedural fairness. Ultimately, the adoption of predictive modeling offers significant potential to revolutionize the visa approval landscape, enhancing the integrity and accessibility of the U.S. visa system for global mobility and economic exchange.

# **1 Introduction**

## **1.1 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

## **1.2 Scope**

The HLD documentation presents the structure of the system, such as the database architecture, 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.

### **1.3 Definitions**

|  |  |
| --- | --- |
| Term | Description |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |
| AWS | Amazon Web Services |

# 

# **2 General Description**

## **2.1 Product Perspective**

Visa Approval Prediction utilizes advanced Machine Learning techniques to offer stakeholders valuable insights into visa approval outcomes. By analyzing historical data and key factors influencing approval decisions, it empowers immigration authorities and applicants to make informed decisions. This model enhances the efficiency of the visa application process, ensuring transparency and fairness while optimizing resource allocation for improved visa approval rates.

### **2.2 Problem statement**

OFLC gives job certification applications for employers seeking to bring foreign workers into the United States and grants certifications. As In last year the count of employees were huge so OFLC needs Machine learning models to shortlist visa applicants based on their previous data.

* To check if Visa get approved or not based on the given dataset.
* This can be used to Recommend a suitable profile for the applicants for whom the visa should be certified or denied based on the certain criteria which influences the decision.

### **2.3 Proposed Solution**

This can be used on real life by US Visa applicants so that they can improve their Resume and criteria for the approval process.

**Solution approach**

1. Machine learning: ML classification Algorithms
2. Deep Learning: Custom ANN with sigmoid activation Function

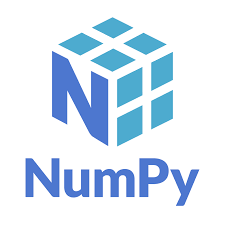
**Here we are going to take the first approach using ML Algorithm**

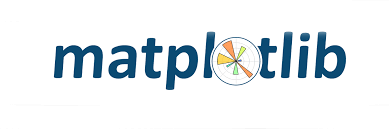
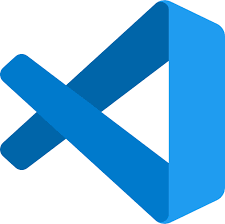
**Solution Proposed**

1. Load the data from DB
2. Perform EDA and feature engineering to select the desirable features.
3. Fit the ML classification Algorithm and find out which one performs better.
4. Select top few and tune hyperparameters.
5. Select the best model based on desired metrics

#### **2.4 Tools used**

Python programming language and other tools such as NumPy, Pandas, Scikit-learn, Fast API, and MongoDB , AWS , MLOps tool : Evidently AI are used to build the whole model.

Vscode is used as IDE

Numpy , numerical computing library for Python.

For visualization Seaborn , Matplotlib , Plotly are used

Pandas , data manipulation and analysis library.

Sci-kit Learn , Machine learning library in Python

AWS , Amazon Web Services cloud platform.

MongoDB , NoSQL document-oriented database.

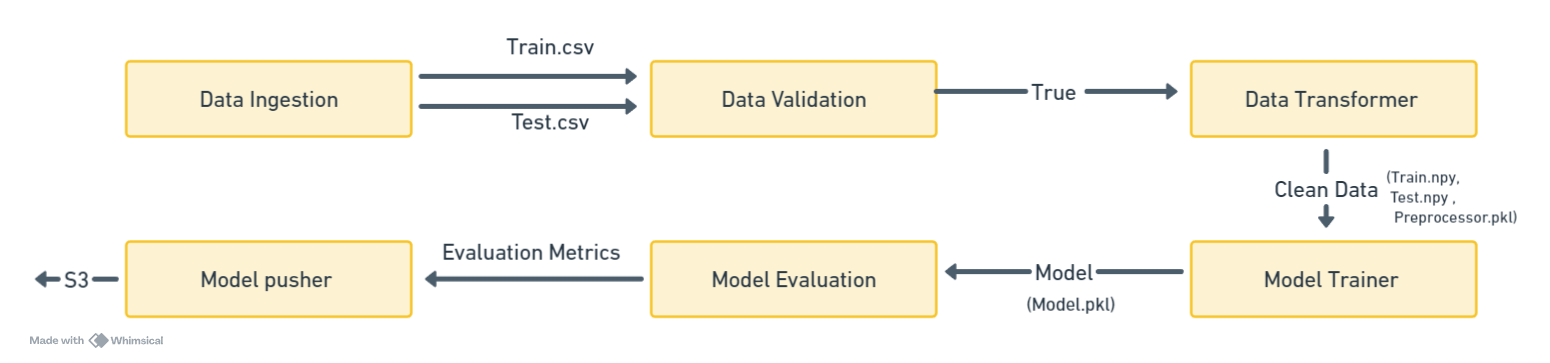
Evidently AI , MLOps tool for data drift

FastAPI , Fast web framework for APIs.

GitHub , Version control platform for collaboration.

#### **Design Details**

#### **3.1Process Flow**



**3.2 Event Log**

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

Initial Step-By-Step Description :

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

**3.3 Errors 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 use.

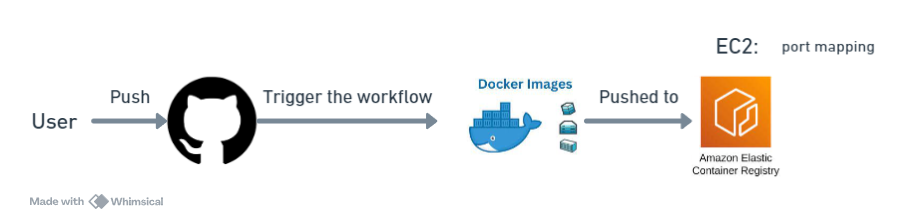
**3.4 Database**

MongoDB is used as Database. It is NoSql Database. MongoDB stores data in flexible JSON-like documents.

****

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

1. Continent (form.continent): The continent from which the applicant is applying.
2. Education of Employee: The level of education attained by the employee applying for the visa.
3. Has Job Experience: Whether the employee has previous job experience.
4. Requires Job Training: Whether the job position requires specific training.
5. Number of Employees (form.no\_of\_employees): The number of employees in the company or organization.
6. Company Age: The age or duration of existence of the company.
7. Region of Employment: The region where the employment will take place.
8. Prevailing Wage: The wage prevailing for similar positions in the relevant region.
9. Unit of Wage: The unit in which the prevailing wage is measured (e.g., hourly, monthly).
10. Full-Time Position: Whether the position offered is full-time or not.
    1. **Deployment**



1. **Performance**

**4.1 Reusability**

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

**4.2 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 .

**4.3 Resource Utilization**

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

1. **Conclusion**

Predicting US visa approval involves analyzing key performance indicators such as the applicant's education level, job experience, and prevailing wage, alongside company-related factors like size and age. Utilizing machine learning techniques on historical visa application data, a predictive model can be developed to forecast visa approval outcomes, aiding both applicants and employers in understanding their likelihood of success. However, it's essential to ensure fairness, transparency, and compliance with legal regulations throughout the model development and application process.