High Level Design (HLD)

Life Expectancy Prediction

Revision Number: 2.0

Last date of revision: 31 March 2023

Nijgururaj Ashtagi

Vaishnavi Patange

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 16/03/2023 | 1.0 | Added Introduction and General description | Nijgururaj Ashtagi |
| 18/03/2023 | 1.1 | Added Design details and abstract | Vaishnavi Patange |
| 23/03/2023 | 1.2 | Created and organized whole document | Nijgururaj Ashtagi |
| 25/03/2023 | 1.3 | Updated Design Details | Nijgururaj Ashtagi |
| 28/03/2023 | 1.4 | Checked and added Performance | Vaishnavi Patange |
| 29/03/2023 | 1.5 | Added Conclusion and References | Vaishnavi Patange |
| 31/03/2023 | 2.0 | Updated diagrams in Design Details | Nijgururaj Ashtagi |

**Table of Content**

Document Version Control ……………………………………………………………. 2

Abstract ………………………………………………………………………………………… 4

Introduction ………………………………………………………………………………….. 5

1. Why this High-Level Document? ……………………………………………. 5

2. Scope …………………………….………………………………………………………. 5

3. Definition ………………………………………………………………………………. 5

General Description ………………………………………………………………………. 6

1. Product Perspective ……………………………………………………….………. 6

2. Problem Statement ………………………………………………………..………. 6

3. Problem Solution ……………………………………………………………………. 6

4. Proposed Methodology …………………………………………………….……. 6

5. Further Improvements ……………………………………………………………. 6

6. Data Required ……………………………………………………...…………………. 6

7. Tools Used ………………………………………………………………………………. 7

8. Constraints …………………………………………………………………..…………. 7

9. Assumptions ……………………………………………………………………………. 7

Design Details ……………………………………………………………..…………………. 8

1. Process Workflow ………………………………………………………..…………. 8

2. Error Handling ………………………………………………………………..………. 9

Performance …………………………………………………………………………….……. 10

1. Reusability ……………………………………………………………..………………. 10

2. Application compatibility ………………………………………..………………. 10

3. Resources Utilization ………………………………………………………….……. 10

4. Deployment ……………………………………………………………………..…..…. 10

Conclusion ………………………………………………………………………….……..……. 11

Reference ………………………………………………………………………………..………. 12

**Abstract**

Life expectancy prediction is a project that uses machine learning algorithms to predict the life expectancy of individuals based on various factors like age, lifestyle, and medical history. The project aims to provide insights into the factors that affect life expectancy and help individuals make informed decisions to improve their life expectancy.

**Introduction**

**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 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

**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.

**3. Definition**

The terms used in the projects are:

* Life expectancy: The average number of years a person is expected to live based on various factors such as health, lifestyle, and environment.
* Adult Mortality: The number of deaths among adults in a population. Alcohol: The amount of alcohol consumption per capita in a population.
* Percentage expenditure: The percentage of GDP spent on healthcare.
* Hepatitis B: The percentage of the population vaccinated against hepatitis B.
* Measles: The number of reported cases of measles in a population.
* BMI: Body Mass Index, a measure of body fat based on height and weight.
* Under-five deaths: The number of deaths among children under the age of five in a population.
* Polio: The percentage of the population vaccinated against polio.
* Total expenditure: The total amount spent on healthcare per capita in a population.
* HIV/AIDS: The prevalence of HIV/AIDS in a population.
* GDP: Gross Domestic Product, the total value of goods and services produced by a country in a year.
* Population: The total number of people living in a country or region.
* Thinness 1-19 years: The percentage of the population with a body mass index below normal for ages 1-19.
* Thinness 5-9 years: The percentage of the population with a body mass index below normal for ages 5-9.
* Income composition of resources: A measure of income inequality based on the distribution of resources in a population.
* Schooling: The average number of years of education completed by adults in a population.
* Developed: A binary indicator variable indicating whether a country is considered developed or not.

**General Description**

**1. Product Perspective**

The life expectancy prediction project is a machine learning-based solution that helps predict the life expectancy of individuals based on various factors. The project can be used by individuals, healthcare professionals, and insurance companies to gain insights into the factors that affect life expectancy.

**2. Problem Statement**

The life expectancy prediction project aims to address the problem of predicting the life expectancy of individuals based on various factors. The problem is challenging due to the complex nature of the factors that affect life expectancy.

**3. Problem Solution**

The life expectancy prediction project provides a solution to the problem of predicting the life expectancy of individuals by using machine learning algorithms to analyse various factors like age, lifestyle, and medical history.

**4. Further Improvement**

The life expectancy prediction project can be improved by incorporating more data sources to improve the accuracy of the predictions. Additionally, the project can be expanded to include more factors that affect life expectancy, such as genetic predisposition.

**5. Data Required**

The life expectancy prediction project requires data on various factors that affect life expectancy. These include 'Adult Mortality', 'Alcohol', 'percentage expenditure', 'Hepatitis B', 'Measles ', 'BMI', 'under-five deaths', 'Polio', 'Total expenditure', 'HIV/AIDS', 'GDP', 'Population', 'thinness 1-19 years', 'thinness 5-9 years', 'Income composition of resources', 'Schooling', and 'Developed'.

**6. Tools Used**

* Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn are used to build the whole model.
* PyCharm and Visual Studio Code is used as IDE.
* For visualization of the plots, Matplotlib and Seaborn are used.
* Google Cloud is used for deployment of the model.
* Front end development is done using HTML5 and CSS3.
* GitHub is used as version control system.

**7. Constraints**

The life expectancy prediction website should be user friendly. Different model to be created for different cities.

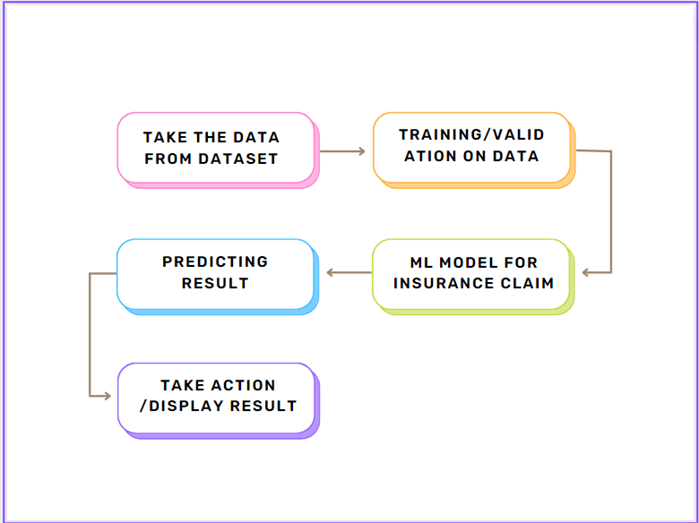
**8. Assumptions**

The life expectancy prediction project assumes that the data used to train the machine learning models is accurate and representative of the population. Additionally, the project assumes that the factors that affect life expectancy remain constant over time and across different geographical regions.

**Design Details**

**1. Process Workflow**

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



Model Training and Evaluation

Diagram

Description automatically generated

Deployment Process

**2. Error Handling**

We got an error while trying to host the website on platforms like pythonanywhere and heroku, but it was solved when we used Google Cloud Platform for model deployment.

**Performance**

**1. Reusability**

Reusability is an important aspect of software development that refers to the ability to reuse code or components across multiple projects or applications. In the context of life expectancy prediction, reusability can help improve performance by reducing the amount of code that needs to be written from scratch, which can save time and reduce errors. By building reusable components, the project can be more efficient and scalable, making it easier to maintain and modify in the future.

**2. Application compatibility**

The project should be designed to work seamlessly on different platforms, which can help improve performance by reaching a larger audience and ensuring that the application works as intended on different devices.

**3. Resource utilization**

The life expectancy prediction project should be designed to use resources efficiently, including CPU, memory, and storage. By optimizing resource utilization, the project can run faster and more efficiently, reducing the chances of crashes or slowdowns.

**4. Deployment**

The code is deployed in GitHub. The whole system is live and is hosted on Google Cloud.

**Conclusion**

The life expectancy prediction project aims to predict the life expectancy of a population based on various factors such as healthcare, education, environment, and lifestyle. By analysing these factors, we can identify the most significant predictors of life expectancy and use them to make accurate predictions.

**References**

1. <https://scikit-learn.org/stable/user_guide.html>
2. <https://numpy.org/doc/>
3. <https://seaborn.pydata.org/examples/regression_marginals.html>
4. <https://seaborn.pydata.org/examples/scatterplot_matrix.html>
5. <https://matplotlib.org/>
6. <https://pandas.pydata.org/docs/>