High Level Design (HLD)

Air Quality Index Prediction

Revision Number: 2.0

Last date of revision: 17 January 2023

MUHAMMED SHIBIL C V

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 27/12/2022 | 1.0 | Added Introduction and General description | Muhammed Shibil |
| 27/12/2022 | 1.1 | Added Design details and abstract | Muhammed Shibil |
| 29/12/2022 | 1.2 | Created and organized whole document | Muhammed Shibil |
| 02/01/2023 | 1.3 | Updated Design Details | Muhammed Shibil |
| 02/01/2023 | 1.4 | Checked and added Performance | Muhammed Shibil |
| 17/01/2023 | 1.5 | Added Conclusion and References | Muhammed Shibil |
| 17/01/2023 | 2.0 | Updated diagrams in Design Details | Muhammed Shibil |

**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 refers to the average number of years that a person is expected to live based on statistical data. Factors that can affect life expectancy include genetics, lifestyle choices, access to healthcare, and environmental factors. Life expectancy can vary greatly between different countries and populations.

There are several factors that can affect life expectancy rate, including:

1. Genetics: Certain genetic conditions can lead to a shorter life expectancy.
2. Lifestyle: Factors such as diet, exercise, smoking, and alcohol consumption can impact life expectancy.
3. Healthcare: Access to quality healthcare can improve life expectancy, as can preventative measures such as vaccinations and regular check-ups.
4. Environment: Exposure to pollutants and toxic substances can negatively impact life expectancy, as can living in areas with high crime rates or limited access to healthy food.
5. Socio-economic status: Lower socio-economic status is associated with lower life expectancy.
6. Geography: Life expectancy can vary significantly between different regions and countries, due to a combination of these factors.
7. Global events: Wars, pandemics, natural disasters and other events can also have an effect on 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:

* GDP - Gross domestic product
* BMI - Body mass index

**General Description**

**1. Product Perspective**

Life expectancy prediction is an analysis based on linear regression model which helps to predict Life expectancy of each country. Such a tool would need to be accurate, reliable and easy to use, and would require access to large amounts of data to make predictions, including demographic, lifestyle and medical data. It would also need to consider privacy and ethical implications

**2. Problem Statement**

To create the machine learning based solution to predict Life expectancy prediction based on the parameters.

**3. Problem Solution**

One problem that a life expectancy prediction tool could help solve is the lack of accurate and individualized information on life expectancy. This can make it difficult for individuals, healthcare providers, and financial planners to make informed decisions about long-term health and financial planning..

**4. Further Improvement**

1. Incorporate machine learning and artificial intelligence techniques to improve the accuracy of predictions, by analyzing more data and providing more personalized recommendations.
2. Incorporate more data sources, such as environmental and socio-economic data, to provide a more complete picture of an individual's life expectancy.
3. Develop a user-friendly interface that is accessible to a wide range of users, including individuals, healthcare providers, and financial planners.

**5. Data Required**

To make accurate predictions, a life expectancy prediction tool would require a wide range of data, including:

1. Demographic data: Information such as age, gender, race, and ethnicity could be used to make predictions about life expectancy.
2. Medical data: Information about an individual's current and past health conditions, such as chronic diseases, could be used to make predictions about life expectancy.
3. Lifestyle data: Information about an individual's habits, such as diet, exercise, smoking, and alcohol consumption, could be used to make predictions about life expectancy.

**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.
* Heroku is used for deployment of the model.
* GitHub is used as version control system.

**7. Constraints**

1. Data availability: A lack of data, particularly in certain demographic or geographic areas, could limit the accuracy of predictions.
2. Data quality: Incorrect or incomplete data could lead to inaccurate predictions.
3. Privacy and ethical concerns: Obtaining and using personal information for life expectancy predictions may raise privacy concerns and ethical considerations, such as how the data will be used and protected.
4. Model's limitations: The life expectancy prediction model may have some limitations in its ability to predict life expectancy, as it may not consider all factors that affect an individual's life expectancy.

**8. Assumptions**

1. Linearity: The model may assume that the relationship between the predictors and life expectancy is linear, when in reality it may be more complex.
2. Stationarity: The model may assume that the underlying data is stationary, meaning that the statistical properties of the data do not change over time.
3. Independence: The model may assume that the predictors are independent of one another,

**Design Details**

**1. Error Handling**

When developing a life expectancy prediction tool, it is important to consider how to handle errors and outliers in the data to ensure the accuracy of predictions. Here are a few strategies that can be used for error handling:

1. Data validation: Validate the data at the time of input to ensure that it is in the correct format and within acceptable ranges. This can help identify and correct errors before they are used in the prediction model.
2. Data cleaning: Clean and preprocess the data to remove outliers, duplicate data, and missing values that can affect the accuracy of predictions.

**Performance**

**1. Reusability**

Reusability is an important aspect to consider when developing a life expectancy prediction tool, as it allows the tool to be used in a variety of different applications. Here are a few strategies to increase reusability:

1. Modularity: Break down the tool into smaller, modular components that can be easily reused in different contexts.
2. API: Create an API (Application Programming Interface) that allows other developers to easily access and use the tool's functionality in their own applications.

**2. Application compatibility**

When developing a life expectancy prediction tool, it is important to consider the different applications in which the tool may be used, and to ensure that it is compatible with those applications.

**3. Resource utilization**

At the initial stage, we were using high space to create the model. Once the model is created, our system only needs at least of 2GB RAM and 1 GB of storage to run the application smoothly. Whenever user tries to predict the LE, system uses less than 10% of the processing power.

**4. Deployment**

The code is deployed in GitHub.

**Conclusion**

In conclusion, a life expectancy prediction tool could provide valuable insights and help individuals, healthcare providers, and organizations make more informed decisions about long-term planning. To improve the accuracy of predictions, it should incorporate machine learning, artificial intelligence techniques, more data sources and user-friendly interface. It should also consider the ethical implications of the tool, such as privacy concerns, and implement appropriate measures to protect users' personal information.

**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/>