VIEH GROUP

Low Level Document (LLD)
Life Expectancy Prediction

Version number: 1.0
Last date of revision: 30 May 2023

Sathvik N G

DECLARATION

I declare that this written submission represents my own ideas in my words. Where others' ideas or words have been included, I have properly cited and referenced the original sources.

VIEH GROUP

I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Revision History

Version	Date	Author	Reviewer	Approver	Comments

VIEH GROUP

0.1	26-05-2023	Sathvik N G	Sathvik N G		Draft version
0.2	27-05-2023	Sathvik N G	Sathvik N G		Suggested some selections like key notes, screen validations and attributes to be added
0.3	28-05-2023	Sathvik N G	Sathvik N G		Suggested document format related comments like correction of version, adding one sections for open issues etc
0.4	29-05-2023	Sathvik N G	Sathvik N G		Suggested some changes like correct sequence diagram, changes in data design sections etc
1.0	30-05-2023	Sathvik N G	Sathvik N G		Baseline version

Table of Contents

1. Introduction:

1.1 Scope of the Document

- This section will cover details regarding scope of the document
- Low level design document will be at component level i.e., for website portal there will be one LLD

1.2 Intended Audience
- This section will cover categories of audiences who will be referring/reviewing this document

1.3 System Overview
- This section will capture overview of system application i.e for what system is being developed
- Who are the stake holders of system?
- What are other external Systems through which this will be interacting

2. Project Briefing:
In the proposed life expectancy prediction project, a machine learning model is developed to estimate the life expectancy of individuals based on various factors. The model takes into account factors such as demographics, lifestyle choices, and healthcare indicators to predict life expectancy. It does not rely on traditional methods such as medical history or genetic information. By analysing these factors, the model can provide insights into a person's life

expectancy and contribute to healthcare decision-making. The model's ability to consider multiple factors and provide accurate predictions sets it apart from conventional life expectancy estimation methods. Additionally, the project aims to raise awareness about factors influencing life expectancy and promote healthy living practices.

3. Problem Statement:

To create the machine learning based solution to predict the life expectancy of the person.

4. Problem Solution:

Develop a web application to predict life expectancy based on various factors, providing individuals and healthcare professionals with valuable insights into potential life expectancy. This application can assist individuals in making informed decisions about their lifestyle choices, healthcare planning, and overall well-being. By leveraging machine learning techniques, the application aims to provide accurate and personalized predictions of life expectancy, empowering users to take proactive steps towards a healthier and longer life.

5. Objective of the Project:

The objective of this project is to develop a machine learning-based model to predict life expectancy based on various factors. The model aims to provide individuals and healthcare professionals with insights into life expectancy and promote informed decision-making regarding health and lifestyle choices.

6. Scope of Project:

The project's scope is to create a predictive model that takes into account factors such as demographics, lifestyle choices, and healthcare indicators to estimate life expectancy. It focuses on providing accurate predictions and raising awareness about the factors that influence life expectancy. The model can be applied to different populations and individuals to assess their estimated life expectancy.

7. Requirements Gathering:

- Window 10 Operating system
- Visual studio software
- 2 Team members for the research part
- Few Github Non copyrighted source codes

8. Analysis:

In the life expectancy prediction project, the analysis focuses on exploring the relationships between various factors and life expectancy. These factors can include demographic information such as age, gender, and socioeconomic status, as well as lifestyle choices like smoking, exercise habits, and diet. Healthcare indicators such as access

to healthcare services, vaccination rates, and disease prevalence can also be considered.

Through data analysis and machine learning techniques, the project aims to uncover patterns, correlations, and dependencies among these factors and life expectancy. This involves using statistical methods and algorithms to identify the most significant variables and their impact on life expectancy. The model is trained on a dataset of historical data that includes information about individuals' characteristics and their corresponding life expectancies.

By leveraging this trained model, the project can make predictions about life expectancy for new individuals or populations based on their specific set of factors. These predictions can provide insights into the potential impacts of certain variables on life expectancy and help individuals, policymakers, and healthcare professionals make informed decisions regarding healthcare, lifestyle interventions, and public health initiatives.

Overall, the analysis in the life expectancy prediction project involves a comprehensive exploration of factors influencing life expectancy, leveraging machine learning techniques to generate accurate predictions and valuable insights for various stakeholders.

9. Final Screenshot of Project Output

VIEH GROUP

Life Expectancy Prediction

Year

2015

Status

0

Adult Mortality

263

Alcohol

0.01

Hepatitis B

65

Measles

1154

BMI

19.1

Under-Five Deaths

83

Polio

6

83

Polio

6

Total Expenditure

8.16

Diphtheria

65

HIV/AIDS

0.1

GDP
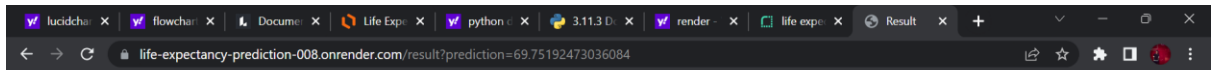
584.259210

Population

33736494

Thinness 1-19 Years

17.2

Income Composition of Resources

0.479

Schooling

10.1

Predict

life-expectancy-prediction-008.onrender.com/result?prediction=69.75192473036084

Prediction: The Estimated Life Expectancy is 69.0

Prediction: The Estimated Life Expectancy is Moderate

localhost:8888/notebooks/model/Life%20Expectancy.ipynb

Jupyter **Life Expectancy** Last Checkpoint: Last Wednesday at 9:48 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

▶ Run ■ C ▶▶ Code

```python
In [82]: import pandas as pd

# Assuming your preprocessed dataset is stored in a DataFrame called "df"

# Separate the feature matrix (X)
X = df.drop(columns=['Life expectancy '])

# Extract the target variable (y)
y = df['Life expectancy ']
```

Train Test Split

Now let's split the data into a training set and a testing set. We will train out model on the training set and then use the test set to evaluate the model.

```python
In [83]: from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Split the data into training, validation, and testing sets
X_train, X_temp, y_train, y_temp = train_test_split(X, y, test_size=0.3, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)

# Choose a regression model (e.g., Linear Regression)
model = LinearRegression()

# Train the model on the training set
model.fit(X_train, y_train)

# Predict the target variable for the validation set
y_val_pred = model.predict(X_val)
```

localhost:8888/notebooks/model/Life%20Expectancy.ipynb

Jupyter **Life Expectancy** Last Checkpoint: Last Wednesday at 9:48 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel) O

Run | Code

```python
In [85]:  # Combine the training and validation sets
          X_trainval = np.concatenate((X_train, X_val), axis=0)
          y_trainval = np.concatenate((y_train, y_val), axis=0)

          # Create the normal linear regression model
          final_model = LinearRegression()

          # Train the final model on the combined training and validation sets
          final_model.fit(X_trainval, y_trainval)

          # Predict the target variable for the test set
          y_pred = final_model.predict(X_test)

          # Calculate the residuals
          residuals = y_test - y_pred

          # Plotting the residuals
          plt.scatter(y_test, residuals, color='blue', label='Residuals')
          plt.axhline(y=0, color='red', linestyle='--', linewidth=2, label='Zero Residuals')
          plt.xlabel('Actual Values')
          plt.ylabel('Residuals')
          plt.legend()
          plt.show()
```
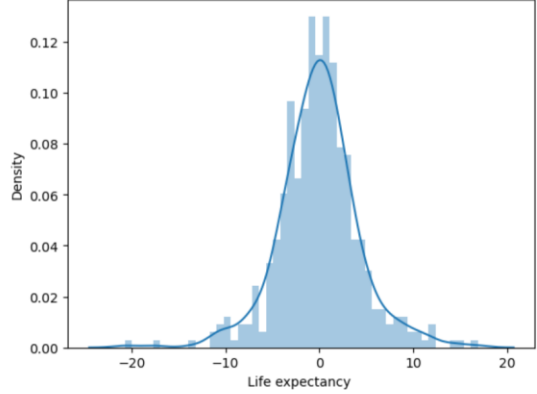
C:\Users\sathv\anaconda3\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but LinearRegression was fitted without feature names
  warnings.warn(

localhost:8888/notebooks/model/Life%20Expectancy.ipynb

Jupyter **Life Expectancy** Last Checkpoint: Last Wednesday at 9:48 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel) O

Run | Code

xibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)


```python
In [89]:  coeff_df = pd.DataFrame(model.coef_,X.columns,columns=['Coefficient'])
          coeff_df
```

Out[89]:

Coefficient

Life expectancy

```
In [89]: coeff_df = pd.DataFrame(model.coef_,X.columns,columns=['Coefficient'])
         coeff_df
```

Out[89]:

	Coefficient
Year	-9.059371e-03
Status	1.567110e+00
Adult Mortality	-2.010868e-02
Alcohol	2.562622e-02
Hepatitis B	-2.040685e-02
Measles	-3.309585e-05
BMI	4.282824e-02
under-five deaths	-1.383187e-03
Polio	2.641039e-02
Total expenditure	4.047598e-02
Diphtheria	4.508620e-02
HIV/AIDS	-4.921031e-01
GDP	4.375682e-05
Population	2.091451e-09
thinness 1-19 years	-4.808128e-02
Income composition of resources	6.568379e+00
Schooling	7.022890e-01