PROJECT REPORT

INTRODUCTION:

Overview:

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features.

Purpose:

This problem statement provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

LITERATURE SURVEY:

Existing problem:

The right to healthcare still isn’t a reality for millions of children around the globe, and more particularly in developing countries. Every year, over 13 million children under the age of 5 years die from illnesses which could have been avoided or treated. In many countries the populations do not benefit from health coverage. Thus, they do not have access to healthcare that remains beyond their means. Some must save up for years before being able to see a doctor or a specialist.

Proposed solution:

The solution is aimed at predicting Life Expectancy rate of a country given various features, in order to develop appropriate strategies to improve the quality and performance of health care systems and thereby increase life expectancy. The proposed machine learning model provides insight to how the following factors affect the life of an individual in their nation.

THEORETICAL ANALYSIS:

Software Designing:

IBM Watson studio platform for creating the machine learning model and deploying it. A Random Forest regression model based on 21 features considering data from a period of 2000 to 2015 for all the countries.

Node Red UI for providing a user friendly interface for the machine learning model.

EXPERIMENTAL INVESTIGATIONS:

Investigations in the form of data analysis were conducted to interpret how various factors like adult mortality, infant deaths, alcohol consumption, total expenditure and other continuous numerical factors affected Life expectancy in the form of Histograms, bar graphs, box plots and so on.

Missing values were identified and corrected with suitable values. Relationships among analysed features and life expectancy were mapped giving insights into the features our prediction mainly depended.

FLOWCHART:

The project revolves around mainly these 6 milestones:

|  |
| --- |
| 1. **Project Planning & Kickoff** |
| 1. **Explore IBM Cloud Platform** |
| 1. **Explore IBM Watson Services** |
| 1. **Introduction to Watson Studio** |
| 1. **Predicting Life Expectancy with python** |
| 1. **Predicting Life Expectancy without python** |

RESULT:

A machine learning model for predicting the Life Expectancy rate of a country is developed. The model can be used to predict the life expectancy of almost 193 nations around the globe.

ADVANTAGES & DISADVANTAGES:

Advantages:

* The proposed machine learning model provides insight to how factors like education, alcohol intake of people in the country, expenditure on healthcare systems and some specific disease related deaths affect the life of an individual in their nation.
* It was found that the effect of immunization and human development index was not taken into account in early prediction models. But the proposed model covers important immunizations like Hepatitis B, Polio and Diphtheria.

Disadvantages:

* The model contains data only from the years 2000 to 2015.Since 1900 the global average life expectancy has more than doubled and is now above 70 years. Therefore in order to predict such a volatile quantity we need more recent data.
* No prediction is perfect. So there is a need to perform further efficient feature engineering techniques to increase the accuracy of the model.

APPLICATIONS:

The proposed model can be used to develop appropriate strategies to improve the quality and performance of health care systems and thereby increase life expectancy.

CONCLUSION:

Many factors like schooling, immunization, alcohol consumption, gdp, adult mortality and population play a major role in determining life expectancy of any nation. Efficient approaches to correct these factors to the required extent can improve the quality and performance of health care systems and thereby increase life expectancy.

FUTURE SCOPE:

* Further addition of data to our model over the coming years can help increase the accuracy of prediction. Thus the machine learning model becomes more reliable to help shape a better future for a nation.
* More factors that affect the life expectancy of a country have to be discovered and data on those factors has to be collected. Later incorporating these to our model will help build better strategies to improve healthcare systems.

BIBLIOGRAPHY:

* <https://www.kaggle.com/kumarajarshi/life-expectancy-who>
* <https://www.youtube.com/user/krishnaik06/playlists>
* <https://www.youtube.com/channel/UCvB8PgOZdb2y7lgToPE-Dfw>
* <https://www.w3schools.com/>
* <https://stackoverflow.com/>

APPENDIX:

* Life\_Exp- Data Analysis.ipynb
* Life\_Exp- Feature Engineering.ipynb
* Life\_Exp- Modelling & Deployment.ipynb
* Life\_exp AutoAI node red flow.json
* Life\_exp ML node red flow.json