Introduction

OVERVIEW

This project was made as an internship project with Smartinternz under the Machine Learning field. The project was completed in one month from 15 May 2020 – 15 June 2020.

PURPOSE

The aim of the project was to predict the life expectancy of a person in a given country.

LITERATURE SURVEY

PROBLEM STATEMENT

Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. 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.

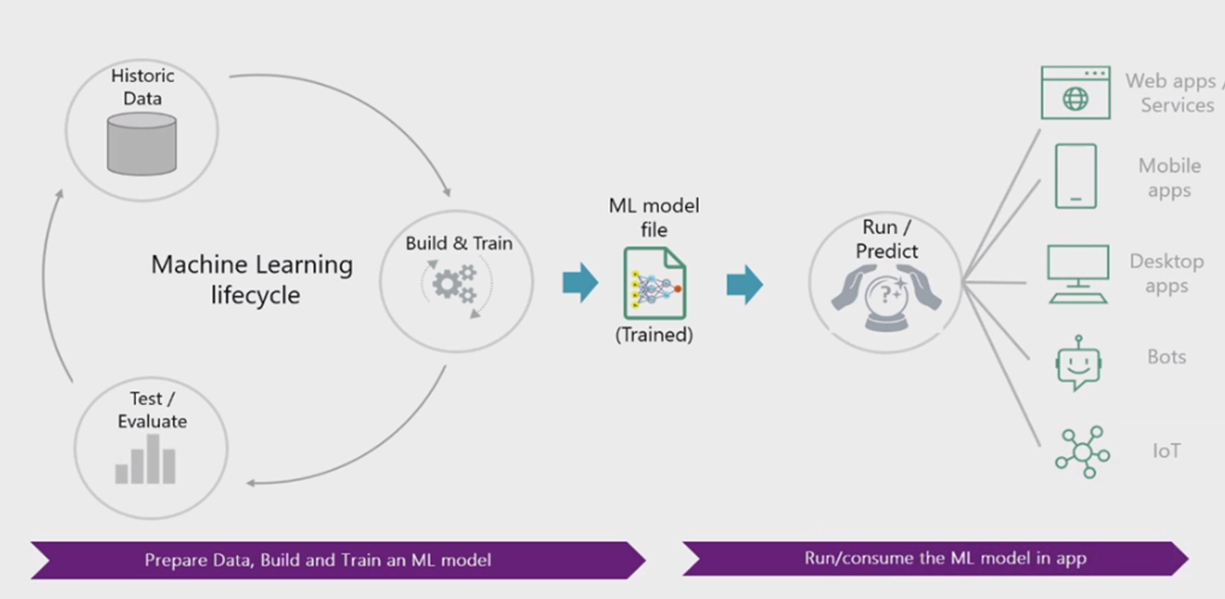
PROPOSED SOLUTION

The factors enlisted above in the problem are collected and a dataset is maintained for the same in every country. The countries compile all this data and send it to WHO , which then combines the data from all the countries and a final dataset is prepared.

To predict the life expectancy, we build a regression machine learning project that 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.

THEORITICAL ANALYSIS

DIAGRAM



EXPERIMENTAL INVESTIGATIONS

DATASET

The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries The datasets are made available to public for the purpose of health data analysis. The dataset related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single dataset. On initial visual inspection of the data showed some missing values. As the datasets were from WHO, we found no evident errors. Missing data was handled in R software by using Missmap command. The result indicated that most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo,Cabo Verde etc. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model dataset. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

The factors collected in the dataset are as follows :-

* Country
* Year
* Status
* Life Expectancy
* Adult Mortality
* Infant deaths
* Alcohol
* Percentage expenditure
* Hepatitis B
* Measles
* BMI
* Under-five deaths
* Polio
* Total expenditure
* Diphtheria
* HIV/AIDS
* GDP
* Population
* Thinness 1-19 year
* Thinness 5-9 year
* Income composition of resources
* Schooling

IBM CLOUD

The IBM cloud platform combines platform as a service (PaaS) with infrastructure as a service (IaaS) to provide an integrated experience. The platform scales and supports both small development teams and organizations, and large enterprise businesses. Globally deployed across data centers around the world, the solution you build on IBM Cloud™ spins up fast and performs reliably in a tested and supported environment you can trust.

The platform is built to support your needs whether it's working only in the public cloud or taking advantage of a multicloud deployment model. With our open-source technologies, such as Kubernetes, Red Hat OpenShift, and a full range of compute options, including virtual machines, containers, bare metal, and serverless, you have as much control and flexibility as you need to support workloads in your hybrid environment. You can deploy cloud-native apps while also ensuring workload portability.

Whether you need to migrate apps to the cloud, modernize your existing apps by using cloud services, ensure data resiliency against regional failure, or leverage new paradigms and deployment topologies to innovate and build your cloud-native apps, the platform's open architecture is built to accommodate your use case.

IBM Watson Studio

IBM Watson Studio helps data scientists and analysts prepare data and [build models](https://dataplatform.cloud.ibm.com/docs/content/wsj/getting-started/welcome-main.html?audience=wdp&context=wdp) at scale across any cloud. With its open, flexible multicloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI, such as:

* Automate AI lifecycle management with [AutoAI](https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html)
* Visually prepare and build models with [IBM SPSS Modeler](https://dataplatform.cloud.ibm.com/docs/content/wsd/spss-modeler.html?audience=wdp&context=wdp)
* Build models using images with [IBM Watson Visual Recognition](https://www.ibm.com/cloud/watson-visual-recognition) and texts with [IBM Watson Natural Language Classifier](https://www.ibm.com/cloud/watson-natural-language-classifier)
* Deploy and run models through one-click integration with [IBM Watson Machine Learning](https://www.ibm.com/cloud/machine-learning)
* Manage and monitor models through integration with [IBM Watson OpenScale](https://www.ibm.com/cloud/watson-openscale)

IBM WATSON MACHINE LEARNING

IBM Watson Machine Learning helps data scientists and developers accelerate AI and machine-learning [deployment](https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-deploy_new.html). With its open, extensible model operation, Watson Machine Learning helps businesses simplify and harness AI at scale across any cloud. Watson Machine Learning provides capabilities to help you:

* Deploy models built with IBM Watson Studio and open source tools. [Watch the video (02:49)](https://mediacenter.ibm.com/media/0_5knhb8fu)
* Dynamically retrain models
* Automatically generate APIs to build AI-powered applications
* Manage models through integration with [IBM Watson Openscale™](https://www.ibm.com/cloud/watson-openscale)
* Streamline model management and deployment end-to-end with an easy-to-use interface

[IBM Cloud Pak™ for Data](https://www.ibm.com/products/cloud-pak-for-data) is a unified data and AI platform that [automates](https://www.ibm.com/cloud/watson-studio/autoai) the AI lifecycle management integrating Watson AI technology.

FLOWCHART

RESULT

The final result of the project was a UI that interacts with the user and inputs the various attributes and predicts the expected age.

BIBILOGRAPHY

* <https://www.allbusinesstemplates.com/download/?filecode=2KBA4&lang=en&iuid=9f9faa69-9fab-40ee-8457-ea0e5df8c8de>
* <https://www.youtube.com/watch?v=LOCkV-mENq8&feature=youtu.be>
* <https://github.com/>
* <https://my15.digitalexperience.ibm.com/b73a5759-c6a6-4033-ab6b-d9d4f9a6d65b/dxsites/151914d1-03d2-48fe-97d9-d21166848e65/>
* <https://www.ibm.com/watson/products-services>
* <https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2>
* <https://www.youtube.com/watch?v=Jtej3Y6uUng>