PROJECT REPORT FOR PREDICTING LIFE EXPECTANCY

ON SMARTINTERNZ PLATFORM

IT IS A 30 DAYS SUMMER INTERNSHIP PROGRAM

SUBMITTED BY: JYOTI VERMA

STUDY : BTech (2nd year)

BRANCH : CSE

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Overview

1.2 Purpose

2 LITERATURE SURVEY

2.1 Existing problem

2.2 Proposed solution

3 THEORITICAL ANALYSIS

3.1 Block diagram

3.2 Hardware / Software designing

4 EXPERIMENTAL INVESTIGATIONS

5 FLOWCHART

6 RESULT

7 ADVANTAGES & DISADVANTAGES

8 APPLICATIONS

9 CONCLUSION

10 FUTURE SCOPE

`

1.INTRODUCTION

1.1 OVERVIEW

The overall project is to make a model which is used for predicting the life expectancy of people within a particular country. As for building the model the data of a thousands of people is used.

The various factors on which the life expectancy of people is depend is used for building the model.

Life expectancy is the key metric for assessing population health. Broader than the narrow metric of the [infant and child mortality](https://ourworldindata.org/child-mortality), which focus solely at mortality at a young age, life expectancy captures the mortality along the entire life course. It tells us the average age of death in a population.

Life expectancy is a measure of premature death and it shows large differences in health across the world. In the countries with the worst health life expectancy is between 50 and 60 years. The decline of [child mortality](https://ourworldindata.org/child-mortality) was important for the increase of life expectancy, but as we explain in our [entry on life expectancy](https://ourworldindata.org/entries/life-expectancy) increasing life expectancy was certainly not only about falling child mortality – [life expectancy increased at all ages](https://ourworldindata.org/life-expectancy#it-is-not-only-about-child-mortality-life-expectancy-by-age).

The United Nations estimate a [global average life expectancy of 72.6 years](https://ourworldindata.org/grapher/life-expectancy?time=1770..2019&country=OWID_WRL) for 2019

Life expectancy is affected by many factors such as: socioeconomic status, including employment, income, education and economic wellbeing; the quality of the health system and the ability of people to access it; health behaviours such as tobacco and excessive alcohol consumption, poor nutrition and lack of exercise.

For creating a model which learn from the previous data of people and predict the life expectancy of people.

We optimized the model architecture and feature set, and tested the performance of several models. The following sections describe:

* the dataset;
* the train-validation-test split;
* our methods for creating the input data for the model;
* our methods for determining the model architecture;
* our methods for feature selection;
* the evaluation protocol

1.2 PURPOSE

* The purpose for predicting the life expectancy of people because life expectancy is a measure that is often used to gauge the overall health of a community. Life expectancy at birth measures health status across all age groups. Shifts in life expectancy are often used to describe trends in mortality.
* This model tests the potential of using machine learning and natural language processing techniques for predicting life expectancy from different factors affect the life expectancy of people.
* Estimates of life expectancy are useful in assessing whether different ... models that use Medicare claims data to predict life expectancy and risk of death at ... rate included metastatic cancer, renal failure, alcohol abuse, drug.

2 LITERATURE SURVEY

2.1 Existing problem

Previous applications of extreme value theory to the study of lifespans addressed the question of whether there are limits to human lifespan. Studies in the biodemography of human longevity indicate a late-life mortality deceleration law: that death rates level off at advanced ages to a late-life mortality plateau. That is, there is no fixed upper limit to human longevity, or fixed maximal human lifespan.

2.2 Propsed solution

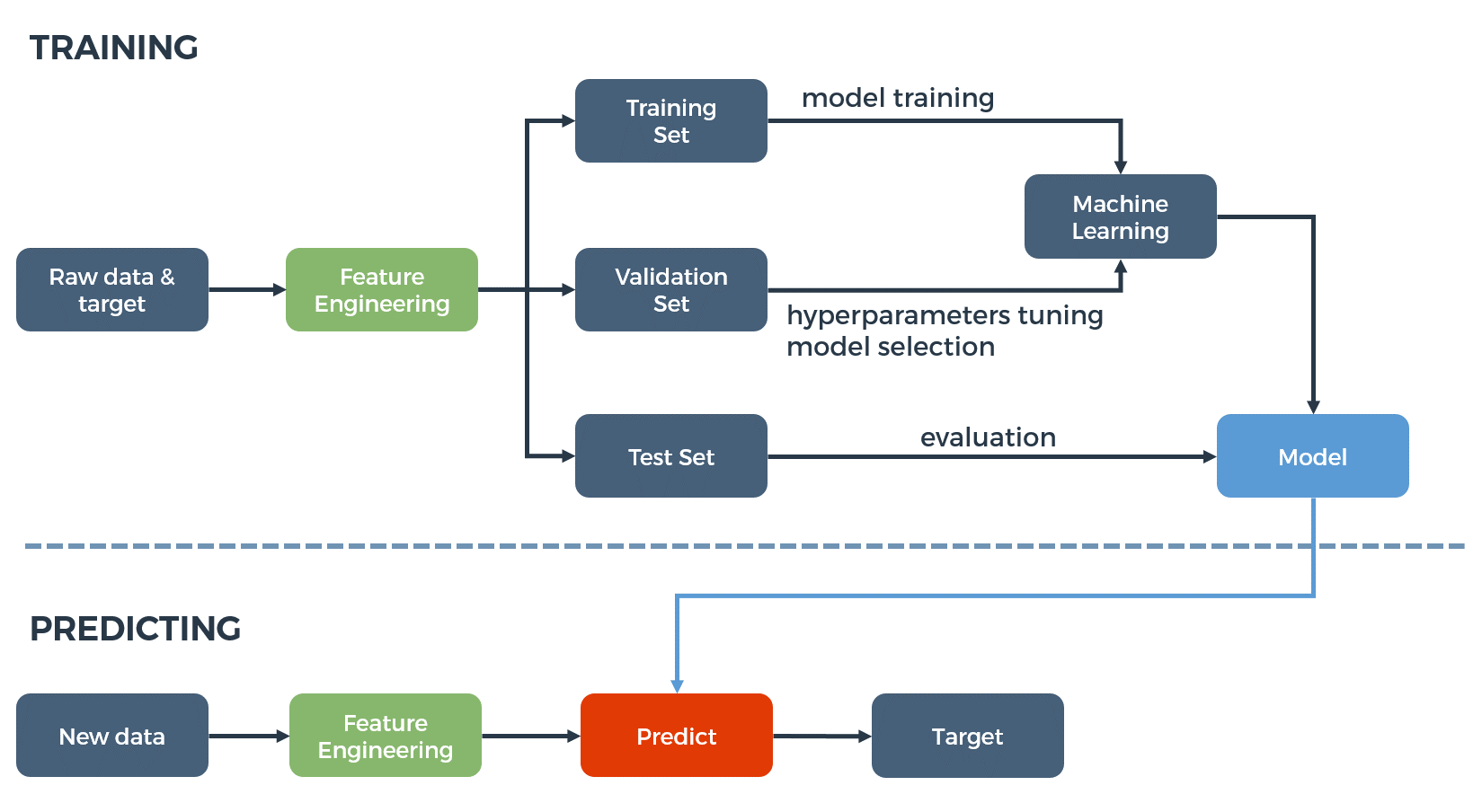
We approached the task of predicting life expectancy as a supervised machine learning task. For predicting the life expectancy of people a model is build using regression algorithm of machine learning which learned from the previous data and predict the life expectancy of people of a country based on the given data.The model is trained and tested on the data of the people of various countries.With the help node-red application a UI is made for getting easy access of predicting the life expectancy of people.

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they haven’t been trained. two algorithms have been used:

* Linear Regression
* Random Forest

3. THEORITICAL ANALYSIS

3.1 Block Diagram of for building the model using machine learning



3.2 Software Designing

Linear regression is one of the simplest and easiest algorithm to understand and apply. It is a predictive machine learning algorithm which belongs to both, statistics, as well as machine learning. Linear Regression is used to find the linear relationship between the label and its one or more features. Thus, the features play the role to help predict the label [6]. Data points are plotted which generates a regression line. We get the best fit line which is the most estimated value with minimum distance between the predicted value and the actual value. It is used mainly for determining the strength of predictors, forecasting an effect and trend forecasting. Linear regression is categorized into two parts: Simple Linear regression and Multiple Linear regression.

Using Machine Learning to Predict life expectancy

1. Define the problem.
2. Gather the data.
3. Clean & Explore the data.
4. Model the data.
5. Evaluate the model.
6. Answer the problem

In this project the IBM cloud services are used such as Watson studio, node-red, Machine-learning, Watson assistant etc.

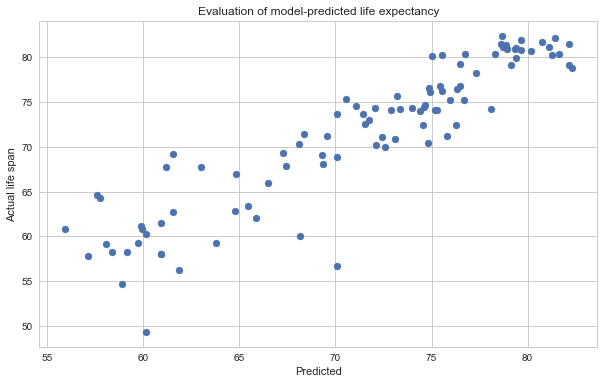
The GUI of the model is built using node-red. A flow is made using the predicting model. The code for building the project is written in jupyter notebook of Watson studio of IBM cloud. The model is connected to the UI of node red with the help of scoring endpoint which we get after deploying the model in notebook.

4 EXPERIMENTAL INVESTIGATION

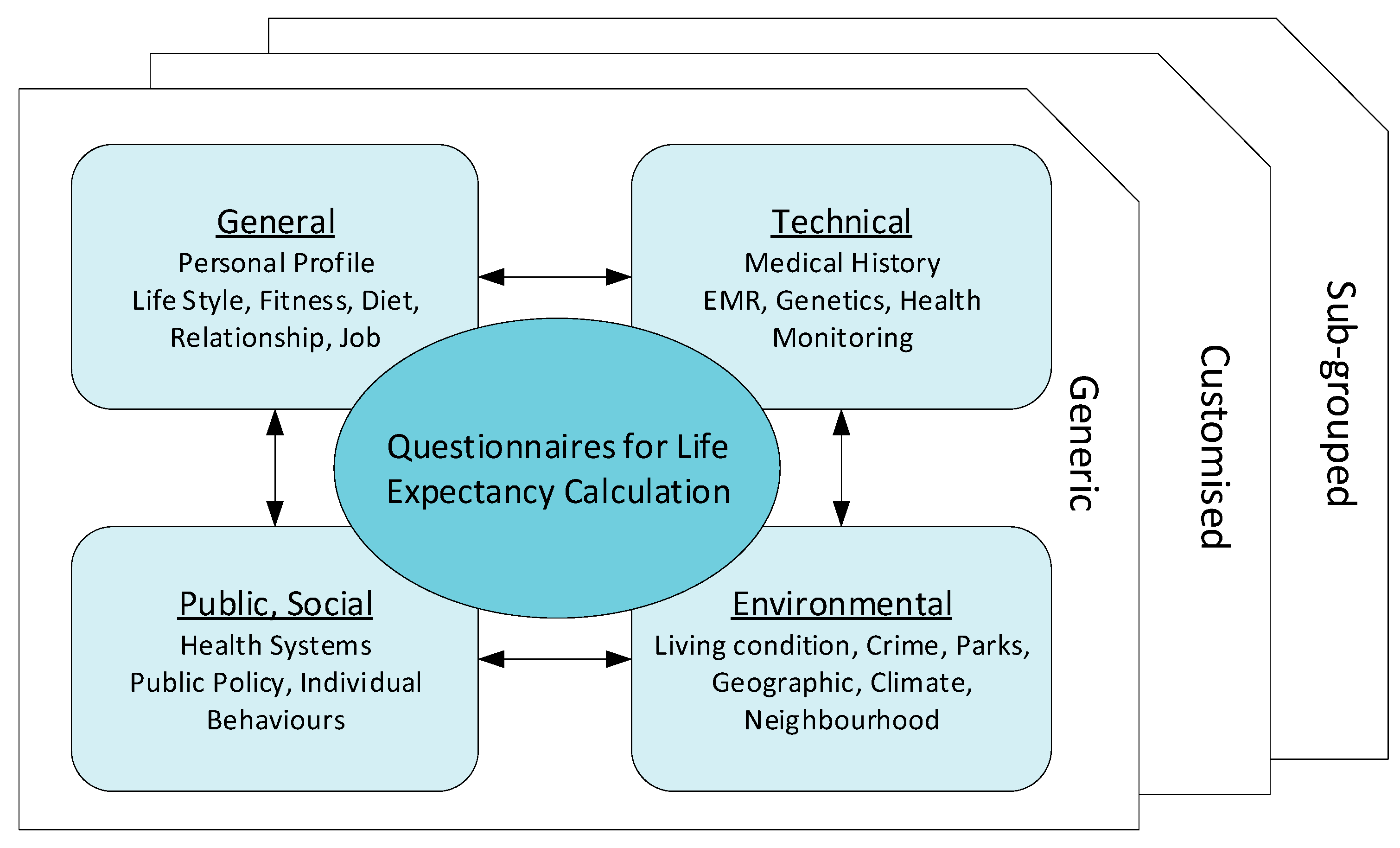
For building the model different predictions and judgment are used for knowing which features affect the life expectancy of people. And how much this features affect the average life expectancy.

Life expectancy is calculated by constructing a life table. A life table incorporates data on age-specific death rates for the population in question, which requires enumeration data for the number of people, and the number of deaths at each age for that population. Significant factors in life expectancy include gender, genetics, access to health care, hygiene, diet and nutrition, exercise, lifestyle, and crime rates. Evidence-based studies indicate that longevity is based on two major factors, genetics and lifestyle choices.

Various graphs and scatter plots are used for investigating or checking which factor affect the life expectancy.



5. FLOW CHART



6.RESULTS

Rising income and the subsequent improved standards of living have long been thought to be the most important factors contributing to a long and healthy life. However, new research has shown that instead, the level of education a person has is a much better predictor of life expectancy. For those with <5 years of predicted life expectancy, 69.2% died within 5 years. For those with <1 year of predicted life expectancy, 47.6% died within 1 year. The results were almost identical for subjects in the training sample

7.ADVANTAGE AND DISADVANTAGES

Life expectancy predictions have the potential to be beneficial to individuals, health service providers and governments. For instance, they would make people more aware of their general health, and its improvement or deterioration over time

Life expectancy can be estimated at any age, e.g. life expectancy at 65 years. Gives more weight to deaths at younger ages. Life expectancy has been used nationally to monitor health inequalities. Disadvantages: At smaller geographies may be influenced by nursing homes in the area. Generally, increased life expectancy has increased the risk of disease, disability, dementia and advanced ageing prior to death.

8.APPLICATION

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning.

9.CONCLUSION

By the application of linear regression, we conclude that according to the given dataset, the features that affect the life expectancy the most are Adult mortality rate, Percentage expenditure and total expenditure on healthcare and treatments, Hepatitis B, Polio, Under 5 death rate, Measles, Population, GDP, HIV/AIDS, Schooling, Income composition, BMI, and Alcohol consumption rate. Application of multiple linear regression to the dataset has produced successful results in predicting the life expectancy of the population in the world as well as the continents in the world. The accuracy score has been good which indicates that the predicted result is almost accurate.

10. FUTURE SCOPE

To plan for pensions and health and social services, future mortality and life expectancy need to be forecast. Consistent forecasts for all subnational units within a country are very rare. Our aim was to forecast mortality and life expectancy for England and Wales' districts. mortality and life expectancy vary substantially within countries, both geographically and in relation to social class. Planning and priority setting at the subnational level need local mortality forecasts. Local information is especially important in countries like the UK that are devolving health and social care responsibilities to local governments to ensure that such decentralisation, which might particularly reduce services in deprived areas, does not worsen health inequalities.