HEALTH EXPECTANCY PREDICTION DEPENDING ON DATA

Mr. Syed Mujeeb Ul-Hassan HOD, Mr. Mahmood Assistant Professor,
Mail id : syedmujeebul@gmail.com, Mail id : shaheensarmast@gmail.com,
Mohd Abdul Umair uddin,Mohammed Feroz, Abdul Bari Sarfaraz, UG Scholar,
Mail Id : uddinumair770@gmail.com, Mail id : mohdferoz6098@gmail.com ,Mail id : bari.sarfaraz2002@gmail.com,
Department of Information Technology,
ISL Engineering College, Bandlaguda, Hyderabad.

Abstract

The amount of data available about patients, medical equipment, doctors, hospitals, and health insurers is constantly expanding. As a consequence, more and more healthcare companies are relying on cutting-edge IT systems to sift through this data and draw useful conclusions. Major policy and managerial implications for healthcare systems worldwide emerge from projected healthy life expectancy. Health expectancy studies have broadened to include aspects from the fields of education, health, economics, and social welfare as a result of advances in predicting technology and years of hard labor. While the prior model used a neural network, the new one will make use of a machine learning method called Random Forest. These models provide a wealth of opportunity for enhancing the social mechanism of ACP and treatment planning. It is possible to make an educated guess as to the average lifetime of a population by taking into account a number of factors in the present.

INTRODUCTION

The term "health expectancy" may refer to either an analytical or statistical assessment of the typical lifespan in a given population. The government and corporate sector alike have relied heavily on health expectancy data for medical, healthcare planning, and retirement reasons for decades. Improved prediction models are now feasible thanks to developments in forecasting, predictive analytic methods, and data science tools. In many nations, determining the appropriate retirement age and planning for the associated public expenditures is a hotly contested political issue. Industrialization and modernization are largely responsible for the worldwide rise in life expectancy witnessed in the last century. People today are healthier and more prosperous than ever before because of scientific progress, improved medical care, and easier access to higher education. Which enhances well-being and adds years to one's life. Recent years have seen significant advancements in health care and public distribution systems in developed nations. Because of the wealth disparity, life expectancy varies considerably between developed and developing nations.

It may be used for a variety of financial purposes, including as retirement preparation and life insurance. It is used for a variety of life insurance-related tasks, including underwriting and pricing tasks. The health of the policyholder is the most important consideration for life insurance companies when setting rates reduce the likelihood of legal action. Predictions of the chosen site's Health Expectancy are rising to prominence as a consequence of recent advancements in the area of Data Science. Institutionalized power over citizens and businesses via the use of rules and laws.

OBJECTIVE

We achieved our aim of producing predictions in the chosen population by combining the aforementioned components into a single set of datasets. Choosing the most appropriate and accurate ML algorithms, however, is the most time-consuming component of constructing this prediction model. The accuracy and reliability of the outcomes are influenced both by the technique used and by the quality of the dataset (which is based on the choice of independent variables and their sources).To improve accuracy, we will not utilize a neural network as the previous model suggested but rather a machine learning approach called Random Forest. Since Random Forest may be used to both classification and regression issues, it is versatile.

SYSTEMIC CONCEPTS

One's "health expectancy" is a statistical projection of how long they have left to live. Statistical analysis makes use of age, birth year, and other demographic information. This study sheds light on the "mortality characteristics" of a group by dissecting yearly deaths according to various demographic variables. This seems to be the result of some retrospective thinking. The purpose of this research is to calculate life expectancy by examining medical histories. We suggest doing a preliminary evaluation of environmental factors and local conveniences in order to more accurately forecast a person's healthy lifetime. We utilize random forest, a

supervised machine learning technique that may be used to both regression and classification issues. making advantage of the myriad of statistical features included in today's machine learning software.

Added benefits: -

Predictions from both regression and classification tasks are precise and straightforward.

Effectively handles massive datasets.

LITERATURE SURVEY

It was created with help from N. Noor Hannah Bodhi, M. Jaya Balan, and S. N. Ramanujan. In order to accurately estimate risks, the life insurance industry uses supervised learning algorithms. In the fourth edition (2018) issue of the Journal of Artificial Intelligence Research, pages 145–154. https://doi.org/10.1007/s40747-018-0072-1

Title:

In order to accurately estimate risks, the life insurance industry uses supervised learning algorithms. This book was written by Noor Hannah Bodhi and Manoj Jaya Balan.

FACTS FOR THE YEAR 2018:

Applicants for life insurance must be assessed for their risk level. Insurance companies utilize the underwriting process to assess applications and determine rates. Automated underwriting has the potential to streamline the application process as data availability increases and data analytics matures. The primary focus of this research is on enhancing the predictive analytics used by life insurance firms throughout the risk assessment process. The study relied on a real-world dataset with over a hundred (anonymized) features. We improved the precision of our models by using dimensionality reduction to focus on the most important elements. Feature selection and feature extraction approaches have been used to reduce the dimensionality of data, with examples being Correlation-Based Feature Selection (CFS) and Principal Components Analysis (PCA). The dataset was used to train machine learning systems for risk prediction. Multiple Linear Regression, an Artificial Neural Network, a Random Tree, and other classifiers were used. When using the PCA approach with the Multiple Linear Regression model, the mean absolute error (MAE) and root-mean-squared error (RMSE) were 1.6396 and 2.0659, respectively, whereas the CFS method with the Restreet algorithm yielded the best results (MAE and RMSE values of 1.5285, respectively).
The "life prediction equation for humans" was presented by A.A. Bhosale and K.K. Sundaram at the 2019 IEEE International Conference on Bioinformatics and Biomedical Technology.

Title:

Calculating an approximation of human longevity
The writers' names are A.A. Bhosale and K.K. Sundaram.

Writing Descriptively in 2019

Models of mortality trends, such as mortality forecasting models or parametric life expectancy models, are often used by researchers in this area. This scientifically-based model of human lifespan accounts for body mass index (BMI), respiratory rate (RR), heart rate (HR), and blood pressure (BP). The many different simulations suggest that advancements in medicine, hygiene, and technology have contributed to a longer average human life duration. Particularly relevant to poor countries are the study's conclusions that governments and insurance companies may utilize easily accessible data to predict average life expectancy.

PROJECT DESCRIPTION
GENERAL:

Using regression models like Linear Regression, Decision Tree Regressor, and Random Forest Regressor, it is possible to predict a person's Health Expectancy. The most common linear approach for regression is ordinary least squares (OLS), sometimes known as linear regression. Linear regression describes the process through which the value of a dependent variable is compared to the value of one or more independent variables. While this method has the benefit of not needing any input parameters, it lacks the flexibility to regulate how complicated the model becomes. Decision tree-based models are often used in the fields of classification and regression. These models employ a decision-making process based on the "if-else" question learning approach. Overfitting the training data is a problem with these models. This issue is solved by using random forests. Overfitting is possible in random forests because of their structure (multiple decision trees). By averaging the trees' results, we can preserve the trees' predictive power while preventing overfitting. Mean squared error, mean absolute error, and the R2 score were used to choose the best model.

MODULES IN METHODOLOGY

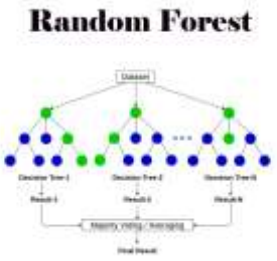
Filtering Data

The gathering of pertinent data is the first necessary step in creating a learning model. As more and better data is added, the overall quality of the model will rise steadily. Web scraping, human involvement, and other methods may be used to collect this information. Prediction of Death Based on Medical Records.

Dataset:

There are a total of 2939 rows of data and 22 distinct columns to examine.

1. Country
2. Year
3. Status
4. Life expectancy
5. Adult Mortality
6. infant deaths
7. Alcohol
8. percentage expenditure
9. Hepatitis B
10. Measles
11. BMI
12. under-five deaths
13. Polio
14. Total expenditure
15. Diphtheria
16. HIV/AIDS
17. GDP
18. Population
19. thinness 1-19 years
20. thinness 5-9 years
21. Income composition of resources
22. Schooling



datasets

Clean and organize data in preparation for machine learning. Remove duplicates, fix mistakes, deal with missing data, normalize, convert data types, etc. to clean up the data. When data is mixed together, its original sequence of collection and/or processing is irrelevant. Data visualization is a powerful method for uncovering biases, outliers, and other anomalies in large datasets.

Model Choosing:

Our Health Expectancy Prediction was developed using a Random Forest model, and its 97.6% accuracy in both testing and forecasts was more than enough to enable its widespread use.

Plan ahead and assess:

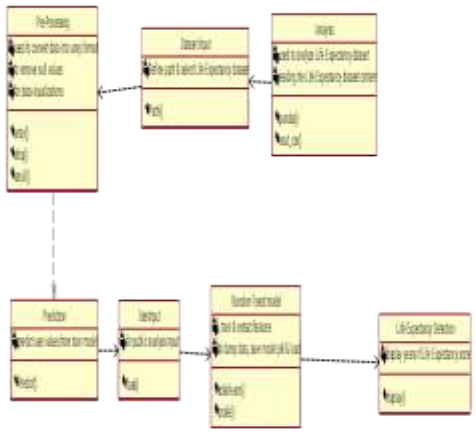
The real dataset had these necessary components:
One's own medical history, broken down into its component fields.
Use of any intoxicating drug, such as alcohol or cannabis.
The third number is an estimate of the person's remaining life expectancy given their current state of health and lifestyle factors.

Information Saved for Model Instruction:

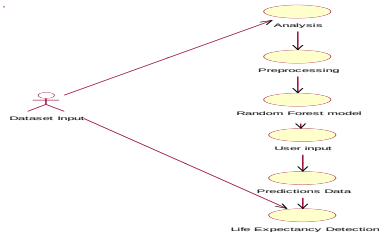
When your model is ready for production, you may save it as a.h5 or.pkl file using a library like pickle.
Verify your PC has the pickle program installed.

General Issues in Engineering and Design

Project Completion as an Endpoint In Design Engineering, UML [Unified Modeling Language] diagrams play a central role. Anybody may follow the steps in a design to create the item. Modeling the ultimate product based on the requirements is an integral part of the software design process. Quality is completed at the design phase of software development. Design is the process of creating a product after hearing feedback from potential buyers.



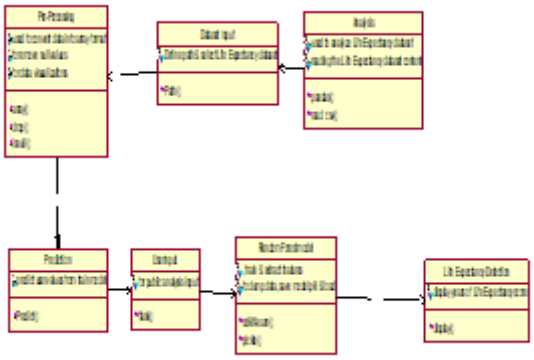
UML DIAGRAMS
USE CASE DIAGRAM



EXPLANATION:

A use case diagram's principal purpose is to highlight which actors are supported by which aspects of the system. The system's actors and their roles may be visually portrayed. The above instance demonstrates how significant the user is. The concept's success will rely on the ideas of everybody involved.

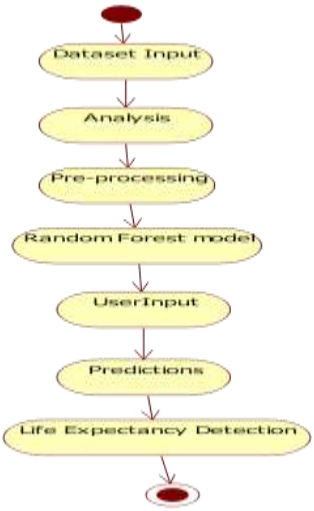
CLASS DIAGRAM



EXPLANATION:

This class diagram shows the relationships between classes, their properties, and their methods, which is necessary for doing the verification safely. The numerous groups working on various aspects of our project are shown in the following diagram..

STATE DIAGRAM



EXPLANATION:

Unspecific diagram meant to depict cycles when certain parts are repeated but others aren't. Having a limited number of states that can be represented by a state diagram is not always the case, but it is usually a suitable abstraction for describing a system. The syntax and meaning of state diagrams vary depending on the kind.

ACTIVITY DIAGRAM

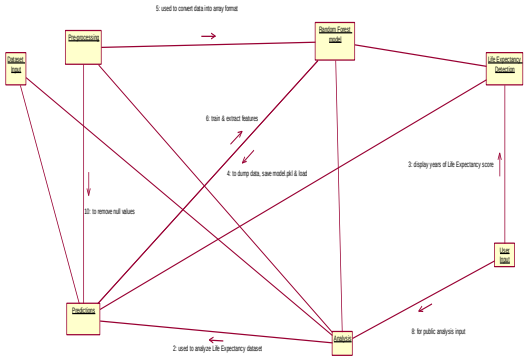
EXPLANATION:

Activity diagrams are visual representations of processes that reveal decision-making, task repetition, and possible task overlap. Activity diagrams are a powerful tool in the Unified Modeling Language for depicting the business and operational stages of system components. A control flow is shown in an activity diagram.

Reading a Sequence Diagram:

In the Unified Modeling Language (UML), sequence diagrams are used to show the sequential execution of related operations. A Message Sequence Chart was used to create this. A sequence diagram is a graphical representation of the temporal order of object interactions. All of the classes and objects involved in the scenario are described, along with the messages used to exchange information between them.

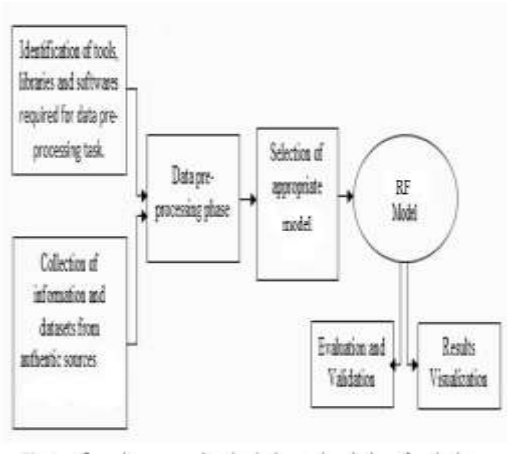
COLLABORATION DIAGRAM



EXPLANATION:

Collaboration diagrams are a subtype of UML diagrams intended to illustrate the interdependencies and interactions between different parts of a program. They are also known as interaction diagrams and communication diagrams. The concept was developed about a decade before the current modeling paradigm and has undergone steady refinement ever since.

SYSTEM ARCHITECTURE:



Proposed Methodology

DEVELOPMENT TOOLS

Python

Python's powerful scripting capabilities stem from the language's high level, interpreted, interactive, and object-oriented nature. Python was created with readability as a primary goal. It has simpler syntactical structures and employs English terms where punctuation is typical in other languages.

Python: A Quick Rundown

Guido van Rossum, Python's developer, created the language in the late '80s and '90s while employed at the Netherlands' National Research Institute for Mathematics and Computer Science.

Aside from the Unix shell and other scripting languages, Python's roots may be traced back to ABC, Modula-3, C, C++, Algol-68, and Smalltalk.

The Python programming language has strict copy protections. The GNU General Public License (GPL) has made Python's source code freely accessible to the public, much as it has for Perl.

While Guido van Rossum is still a key figure in Python's evolution, the language is currently supported by a team of developers at the school.

Essential Python Features

Since Python code is processed by an interpreter at runtime, we use the word "interpreted." Your code will run as anticipated without the requirement for pre-compilation. That reminds me of PHP or PERL.

At a Python prompt, code may be put in directly to the interpreter.

Thanks to Python's Object-Oriented features, we may write code in which individual functions are abstracted away into their own objects.

Python is an excellent first programming language because it can be used to create so many various sorts of programs, from text editors to web browsers to games.

CONCLUSION

The main emphasis was on investigating the impact and variability of the whole via the examination of its constituent parts. The first order of business was to choose the optimal model. The test set MAE for random forest is just 1.27, and its R2 score of 96% is the highest of any model. Life expectancy is heavily influenced by adult mortality, HIV/AIDS prevalence, literacy, and obesity rates. Higher levels of education, socioeconomic diversity, and lower body mass index all contribute favorably to the result. The Gross Domestic Product, total expenditures, and infant mortality rates all had surprisingly little roles in the grand scheme of things. It now seems that our original interpretation of these characteristics was incorrect.

REFERENCES

[1] Noor Hannah Boodhun, Manoj Jayabalan, "Risk prediction in life insurance industry using supervised learning algorithms," Complex & Intelligent Systems, vol. 4, no. 2, pp. 145-154, 2018.

[2] Mahumud, R.A., Hossain, G., Hossain,R., Islam, N. And Rawal, L.,;, "Impact of Life Expectancy on Economic Growth and Health Care Expenditures in Bangladesh.," Universal Journal of Public Health, vol. 1, no. 4, pp. 180-186, 2013.

[3] Bhosale, A.A. and Sundaram, K.K., "Life prediction equation for humanbeings," International Conference on Bioinformatics and biomedicaltechnology, vol. IEEE, pp. 266-268, 2019.

[4] Aggarwal, D., Mittal, S., Bali V., "Prediction Model for Classifying Students Based on Performance using Machine Learning Techniques," International Journal of Recent Technology and Engineering, vol.8 p.2S7, 496-503, 2019.

[5] Chen, Tianqi; Guestrin, Carlos, "xgboost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[6] Aggarwal, D., Mittal, S. And Bali, V., "Significance of Non- Academic Parameters for Predicting Student Performance Using Ensemble Learning Techniques", International Journal of System Dynamics Applications (IJSDA), Vol. 10, Issue 3, Article 3, pp. 38-49, 2020.

[7] Kerdprasop, N. And Foreman, K. J., "Association of economic and environmental factors to life expectancy of people in the Mekong basin," IEEE Conference on industrialelectronics and Applications (ICIEA), pp. 1984-1989, 2017.

[8] Beekshma., "A neural-networkanalyzer for mortality forecast," astinbulletin: The Journal of the IAA, vol. 48, no. 2, pp. 481- 508, 2018.

[9] Deb, C., Zhang, F., Yang, J., Lee, S.E. and Shah, K.W., "A review on time series forecasting techniques for building energy consumption," Renewable and Sustainable Energy Reviews, vol.
74, pp. 902-924, 2017.

[10] Sindhwani, N., Verma, S., Bajaj, T., & Anand, R. (2021). Comparative Analysis of Intelligent Driving and Safety Assistance Systems Using YOLO and SSD Model of Deep
Learning. International Journal of Information System Modeling and Design (IJISMD), 12(1), 131-146

[11] Aggarwal, D., Bali, V., Agarwal, A., Poswal, K., Gupta, M., Gupta, A. "Sentiment Analysis of Tweets Using Supervised Machine Learning Techniques Based on Term Frequency," Journal of Information Technology Management, vol.13 no.1,pp. 119-141, 2021

[12] .M. R. Hebb, "The Organization of Behaviour," New York, Wiley,, p.437,1949.

[13] Rosenblatt, F., "The perceptron: a probabilistic model for informationstorage and organization in the brain," Psychological review, vol. 65, no. 6, p.386, 1958.

[14] Sormin, M.K.Z., Sihombing, P., Amalia, A., Wanto, A., Hartama, D. And Chan, D.M., "Predictions of World Population Life Expectancy Using Cyclical Order Weight/Bias," Physics:Conference Series (IOP Publishing), vol. 1255, no. 1, p. 012017, 2019.

[15] Nath, B., Dhakre, D.S. and Bhattacharya, D., "Forecasting Wheat production in India: An ARIMA modelling approach," Journal of Pharmacognosy and Phytochemistry, vol. 8, no. 1,Pp. 2158-2165, 2019.

[16] Bali, V. , Kumar, A. And Gangwar, S., "A Novel Approach for Wind Speed Forecasting Using LSTM-ARIMA Deep Learning Models", International Journal of Agricultural and Environmental Information Systems (IJAEIS), Volume 11, Issue 3, pp. 13-30, ISSN: 1947-3192, EISSN: 1947-3206, 2020.

[17] Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., Time series analysis: forecasting and control, John Wiley & Sons., 2015.

[18] Foreman, K.J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., mcgaughey, M., Pletcher, M.A., Smith, A.E., Tang, K., Yuan, C.W. andbrown, J.C., "Forecasting life expectancy, years of life lost, and all- cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016– 40 for 195 countries and territories," The Lancet, vol. 392, no. 10159, pp. 2052- 2090, 2018.