Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

Machine Learning-Based Life Expectancy Prediction in Developed and Developing Regions

Battina Reshma
Registration Number: 324228860013
Department of Master of Computer Applications (MCA)
Dr. Lankapalli Bullayya PG College, Visakhapatnam, India
Internal Guide: **B.V.P Latha**,
Assistant Professor, Department of MCA

Abstract - Life expectancy is a core demographic and public health indicator reflecting the overall quality of healthcare, living conditions, nutrition, socioeconomic progress, and disease control in a country. However, manually analyzing how dozens of health, social, and economic indicators influence lifespan is complex and time-consuming. This paper presents an interactive, web-based intelligence application developed using the Django framework, SQLite database, and machine learning regression algorithms to forecast life expectancy in developed and developing regions. The system utilizes real-world health and development indicators from the World Health Organization (WHO) and United Nations. We implement and evaluate three regression models: Linear Regression, Random Forest Regressor, and XGBoost Regressor. Preprocessing steps handle missing data, normalize features, and split records into training and test sets. XGBoost achieves superior predictive performance due to its robust gradient-boosting trees, yielding the lowest Mean Absolute Error (MAE) and highest R-squared (R2) value. The system is integrated into a Django web platform, allowing users to view data tables, train models, compare performance metrics via interactive visualizations (correlation matrices, scatter plots, violin plots, and radar charts), and predict life expectancy through user-friendly forms.

Keywords— *Life Expectancy Forecasting, Machine Learning, Django Framework, XGBoost Regressor, Random Forest, Public Health Analytics, Predictive Modeling.*

I. Introduction

Life expectancy serves as a primary benchmark for a country's development. While global life expectancy has increased over the last century, significant disparities remain between developed and developing nations. These variations are driven by complex, interrelated factors such as immunization rates (Hepatitis B, Polio, Diphtheria), disease prevalence (HIV/AIDS, Measles), economic strength (GDP, healthcare expenditure), and social measures (schooling, income composition of resources).

Traditional public health analysis relies heavily on static reports and retrospectively compiled census spreadsheets. These tools do not offer immediate predictive forecasting for hypothetical or newly observed health indicator sets. Moreover, non-technical users find it difficult to explore high-dimensional data or compare machine learning models on a single interface.

To address this challenge, we introduce a Django-based Life Expectancy Prediction and Forecasting System. The platform integrates machine learning workflows with web technologies, allowing students, researchers, and public health officials to train regression models, compare evaluation metrics, and perform real-time lifespan estimations.

II. Literature Review & Existing System

A. The Traditional Approach

Traditionally, actuarial and public health analyses have used basic linear and log-linear statistical models to forecast mortality rates and life expectancy.
1. Static Reporting: Standard reports do not support interactive scenarios where users can input custom health metrics to instantly forecast life expectancy.
2. Limited Feature Interaction: Classical regression models struggle to capture non-linear relationships and high-degree interactions between multiple features, such as the compound effect of HIV/AIDS and low GDP in developing regions.
3. No Integrated User Interface: Research is often limited to standalone scripts (R or Python notebooks) without user-friendly interfaces, making it inaccessible to policy makers.

B. Machine Learning Solutions

Recent studies have successfully applied machine learning to public health datasets:
• Random Forest: Ensemble methods mitigate overfitting and model non-linear boundaries.

• XGBoost: Gradient boosting algorithms provide high accuracy on structured tabular datasets.

III. Proposed System and Workflow

The proposed system is an interactive Django web application integrated with scikit-learn and XGBoost regression engines. The system allows users to register and log in, view the WHO life expectancy dataset in tabular form, train and compare multiple regression models, view evaluation visualization graphs, and enter indicators to predict life expectancy. The admin module manages user accounts and activation statuses.

IV. System Implementation and Technology Stack

The software environment comprises Python runtime, Django MVC framework, and SQLite database. Machine learning workflows use pandas, numpy, scikit-learn, xgboost, matplotlib, seaborn, and joblib. The dataset contains country-wise and year-wise health records. Ingestes indicators include Adult Mortality, Infant Deaths, Alcohol consumption, Percentage Expenditure on Health, Hepatitis B, Measles, BMI, Under-five Deaths, Polio, Total Expenditure on Health, Diphtheria, HIV/AIDS, GDP, Population, Schooling, and Income Composition of Resources. The target variable is Life Expectancy.

V. Algorithms and Evaluation Metrics

A. Regressors

1. Linear Regression: A baseline regression model establishing linear coefficients.
2. Random Forest Regressor: An ensemble model fitting multiple decision trees on sub-samples, using averaging to control overfitting.
3. XGBoost Regressor: A gradient-boosted decision tree algorithm designed for speed and performance, implementing regularization to prevent overfitting.

B. Evaluation Metrics

Models are compared using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2). In testing, XGBoost achieved superior performance with an R2 score of 0.945, outperforming Random Forest (0.912) and Linear Regression (0.825).

VI. Conclusion & Future Scope

The Django-based Life Expectancy Prediction System successfully automates the training and comparison of machine learning models for public health forecasting. By embedding XGBoost regression weights behind a clean Django interface, the system provides accurate predictions without exposing raw code complexity to users.

Future enhancements include: (1) Setting up automatic weekly dataset scraping from the World Health Organization (WHO) database to keep predictions aligned with recent trends; (2) Migrating from SQLite to PostgreSQL for concurrent multi-user database write operations; and (3) Deploying the application to cloud platforms (AWS/Azure) for global public health accessibility.

VII. References

[1] Django Web Framework Documentation. URL: https://docs.djangoproject.com/

[2] Scikit-learn: Machine Learning in Python. URL: https://scikit-learn.org/

[3] XGBoost: Scalable and Flexible Gradient Boosting. URL: https://xgboost.readthedocs.io/

[4] World Health Organization (WHO) Life Expectancy Dataset. URL: https://www.who.int/data/gho

[5] McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.

[6] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. ACM SIGKDD, 785-794.