# PROJECT TOPIC

## Life Expectancy Analysis – Machine Learning Project

**By: Sujeet Kumar**

## Project Submitted To

**Unified Mentor Private Limited**
**Gurugram,Haryana ,122002**

# Content

# 1. Introduction

Life expectancy is a critical indicator of a nation's health and socio-economic development. Various factors such as immunization, mortality rate, GDP, alcohol consumption, and healthcare expenditure impact life expectancy. This project aims to analyze life expectancy across 193 countries over the years 2000 to 2015, using data collected from the World Health Organization (WHO) and United Nations (UN).

The objective is to explore how socio-economic and health-related variables affect life expectancy and build predictive models using machine learning techniques to estimate life expectancy based on these factors.

# 2. Project Statement

The goal of this project is:

- To determine which variables significantly influence life expectancy.

- To predict the life expectancy of a country using relevant health and economic indicators.

- To identify actionable insights that countries with low life expectancy (<65 years) can adopt to improve the overall lifespan of their citizens.

Key questions addressed:

- What are the major influencing factors of life expectancy?
- How do mortality, alcohol use, immunization, and education correlate with lifespan?
- Can machine learning models accurately predict life expectancy using historical data?

# 3. Methodology

## Data Source

- The dataset includes life expectancy-related data from 193 countries from 2000–2015.
- Data was sourced from the WHO Global Health Observatory and the UN.

## Data Preprocessing

- Missing values handled using mean imputation.
- Outliers treated using interquartile range (IQR)-based methods.
- Categorical variables were encoded using Label Encoding.
- Data was normalized using StandardScaler.

# Exploratory Data Analysis (EDA)

- Distribution of life expectancy and correlation matrix were analyzed.
- Time series plots showed increasing life expectancy over the years.
- Relationships between life expectancy and features like GDP, alcohol consumption, schooling, and immunization were visualized.
- Developed vs. developing country analysis was performed.

# Model Building

Models used:

- Random Forest Regressor
- Extra Trees Regressor
- Gradient Boosting Regressor
- XGBoost Regressor

Train-test split: 80% training and 20% testing

### Evaluation Metrics

- $R^2$ Score
- RMSE (Root Mean Squared Error)
- Cross-validation (K-Fold = 20)

## 4. Results

| **Model** | **R² Score** | **RMSE** |
|---|---|---|
| **XGBoost Regressor** | 0.959 | 1.98 |
| Extra Trees Regressor | 0.958 | 1.99 |
| Random Forest Regressor | 0.957 | 2.03 |
| Gradient Boost Regressor | 0.939 | 2.41 |

- **XGBoost** performed best with an R² of **0.959**.
- **Cross-validation mean R²**: **0.9624**, indicating model stability across folds.
- Important positive predictors: **Schooling**, **Healthcare Expenditure**, **GDP**, **Immunization**.
- Negative impact factors: **Adult Mortality**, **HIV/AIDS prevalence**, **Alcohol consumption (in excess)**.

# 5. Conclusion

This project successfully analyzed the impact of various socio-economic and health factors on life expectancy. Key takeaways:

- **Education (schooling)** and **health expenditure** significantly enhance life expectancy.
- **Adult and infant mortality rates** are strong negative indicators.
- Countries aiming to increase life expectancy should prioritize **universal immunization**, **education**, and **access to healthcare**.

By applying machine learning models, especially XGBoost, we can reliably predict a country's life expectancy and guide policy decisions to improve public health outcomes.