LIFE EXPECTANCY PREDICTION ANALYSIS

"Data-Driven Insights for Longer, Healthier Lives"

OUTLINE

PROJECT OBJECTIVE

SDG PROBLEM & OBJECTIVES

DATA

MODELLING

KEY INSIGHTS

ETHICAL CONSIDERATIONS

RECOMMENDATIONS

PROJECT OBJECTIVE

- Build a high-accuracy predictive model for life expectancy estimation.
- Identify and rank the most critical health indicators that influence population longevity.
- Provide actionable insights to help optimize healthcare resource allocation.
- Replace guesswork-based policy decisions with evidence-based recommendations.

SDG PROBLEM & OBJECTIVES

UN SDG 3 TARGETS:

- Project supports global efforts to improve life expectancy.
- Uses AI/ML to uncover critical health-related predictors

OUR FOCUS:

- Enhance decision-making in health sectors by providing insights.
- Predict life expectancy using demographic/health data.

DATA

- 20+ health indicators (e.g., Adult Mortality, GDP, Immunization, Schooling) from WHO/World Bank.
- Preprocessing:
- Handled missing data (median imputation).
- Normalized numerical features.
- Encoded categorical variables (e.g., Region)

MODELLING

Algorithms Implemented:

- Random Forest Regressor
- Decision Tree Regressor
- Logistic Regression
- Multiple Linear Regression

Evaluation Methodology:

- **Cross-validation** for reliable performance estimation

Final Rankings:

Model	Accuracy
Random Forest	0.933447
Tuned Random Forest	0.931741
Tuned Decision Tree	0.904437
Decision Tree	0.880546
Tuned Logistic Regression	0.841297
Logistic Regression	0.617747
Multilinear (Logistic)	0.617747

KEY INSIGHTS

Feature Importance:

❖ The Random Forest model identified key determinants of life expectancy including:

✓ Adult Mortality (strong negative correlation)

✓ Income Composition of Resources (strong positive correlation)

✓ HIV/AIDS (strong negative correlation)

✓ Schooling (positive correlation)

✓ BMI (positive correlation)

✓ Developed vs Developing Nations:

❖ The binary "Status" feature showed significant impact, with developed countries generally having higher life expectancy.

ETHICAL CONSIDERATIONS

✓ Data Collection Bias:

- The dataset may underrepresent certain populations leading to skewed predictions for these groups.
- Health indicators like healthcare spending or immunization rates might be missing or less accurate for marginalized regions, affecting model reliability.

✓ Measurement Bias:

- Variables like GDP or education levels might not capture the full context introducing inaccuracies.

✓ Algorithmic Bias:

- Features like "Status" (Developing/Developed) might oversimplify complex socioeconomic factors, leading to stereotyped predictions.

This solution helps promote:

✓Fairness:

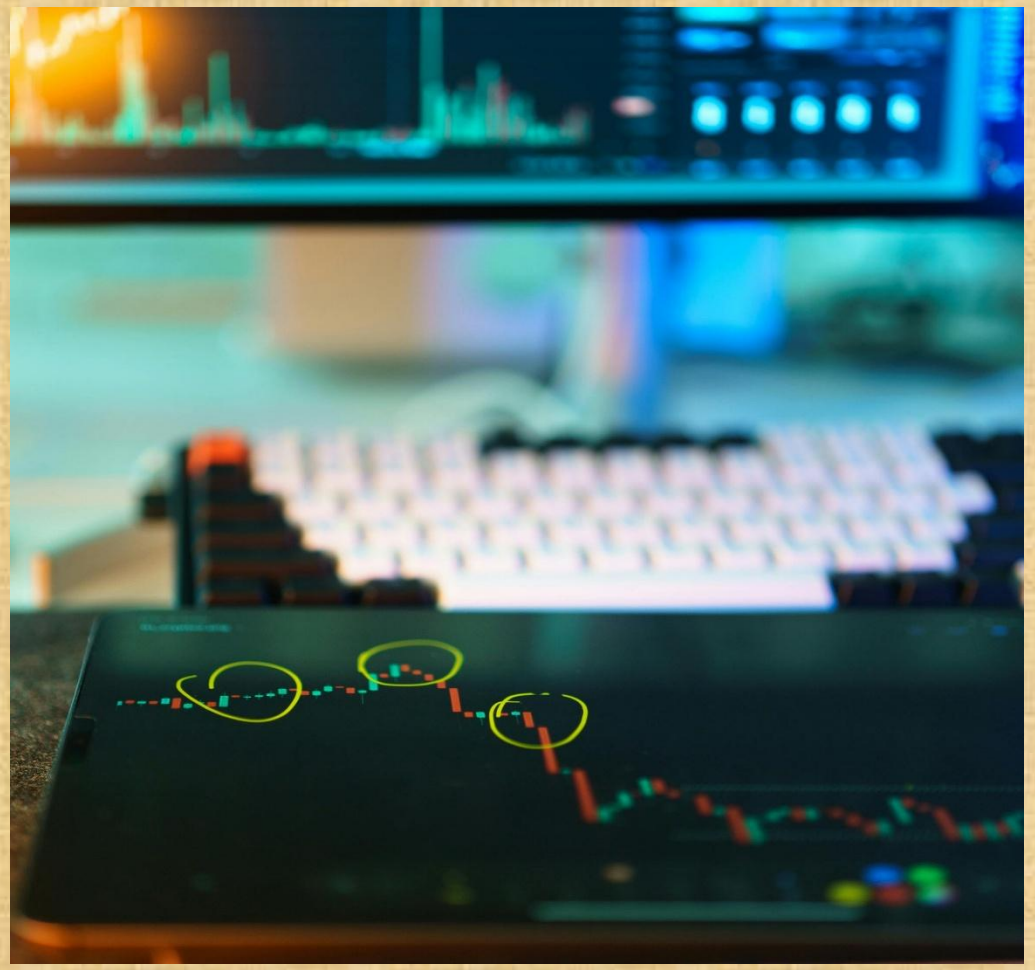
Ensures diverse representation in the dataset.

Clearly communicate model certainties to policymakers, especially for marginalized populations.

✓Sustainability:

By identifying key health factors (e.g., immunization, education), the model helps direct resources efficiently, aligning with SDG 3's goal of equitable healthcare.

Encouraging policies based on data-driven insights to reduce future healthcare costs and improve systemic resilience.



RECOMMENDATIONS

❖ **For Policymakers:**

- Prioritize reducing adult mortality rates through disease prevention programs.
- Invest in education (Schooling) as it showed strong secondary health benefits.
- Focus on comprehensive healthcare rather than just increasing expenditure.
- Target HIV/AIDS prevention and treatment in high-risk regions.

❖ **For Healthcare Organizations:**

- Develop integrated programs addressing multiple health indicators simultaneously.
- Monitor and improve immunization coverage, especially in developing nations.
- Address malnutrition (thinness indicators) as it showed significant impact.

THANK YOU

Questions?

Let's predict better health for a longer life

danielpriscilla61@gmail.com