Life Expectancy task:

Introduction

The dataset contains country-wise features (e.g., GDP, schooling, health indicators), with the target variable being life expectancy. The objective is to predict life expectancy using these features.

Methodology

We began with exploratory data analysis. Scatterplots revealed that most variables followed linear patterns. A correlation heatmap was then used to identify how strongly each variable was associated with life expectancy. Based on this, we selected the top 10 variables, ensuring that no two highly interdependent variables were chosen (e.g., thinness 1–5 vs. thinness 5–19) to avoid multicollinearity.

Data preprocessing involved imputing missing values with the numerical mean of each column. Categorical variables (Country and Status) were converted into numeric format using one-hot encoding.

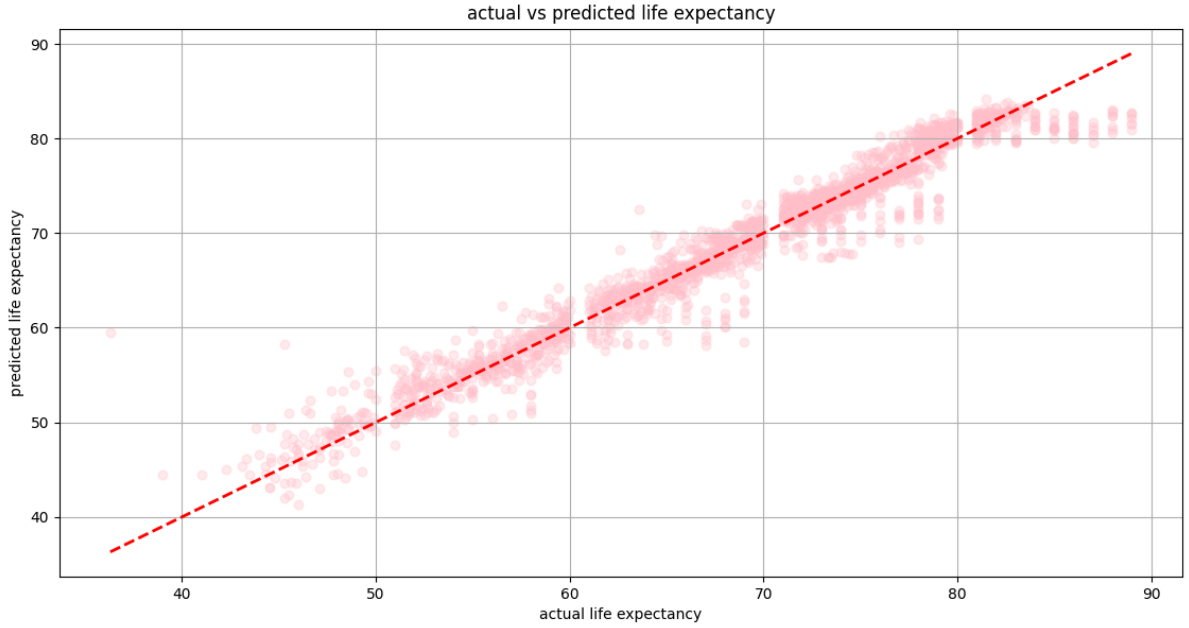
Experimentation

We tested different model variations by altering the feature set. One-hot encoding Country significantly improved performance. One-hot encoding Status also helped, but less so, as it was likely correlated with other variables.

Results

Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2

Model Name	MSE	RMSE	R^2
Model 1	16.81	4.1	0.86
Model 2	3.61	1.9	0.96
Model 3	3.61	1.9	0.96



Explanation

Regularization helped identify the most influential predictors (e.g., schooling, income, healthcare factors). Less relevant predictors were suppressed, clarifying which variables drive life expectancy.

Challenges

The main issues faced were missing data and multicollinearity. Missing values were handled by mean imputation, and interdependent variables were carefully excluded. Another challenge was that categorical data could not be used directly, so we applied one-hot encoding. This allowed the model to learn meaningful weights for each country.

Conclusion

We achieved a strong linear regression model. With an r^2 score of 0.96, the error tolerance was only 0.04, indicating limited multicollinearity. The model's RMSE of 1.9 years shows it could predict life expectancy with fairly high accuracy.

References

https://www.youtube.com/watch?v=VmbA0pi2cRQ

https://www.youtube.com/watch?v=i3IadpjctWg

https://www.youtube.com/watch?v=IHZwWFHWa-w