Healthy Life Expectancy (HALE) Analysis

A Data-Driven Assessment of Health Outcomes (200032019)

By Piyush Ramteke | Python, SQL, Excel, Jupyter Notebook

Piyu

Understanding Healthy Life Expectancy

What is HALE?

Healthy Life Expectancy(HALE) measures the expected years a person lives in **full health**, not just survival. Unlike traditional life expectancy, HALE accounts for years lived with disability or illness, providinga more nuanced picture of population health.

Study Focus

Thisanalysisexamines India's health trajectory using WHO data spanning **200032019**, a period of significant economic and healthcare transformation.

Core Objective

Weaim to understandtrends, identify key determinants, and develop predictive models for HALE to inform public health policy and resource allocation.

Key Research Questions

Temporal Trends

Howhas HALE evolved over two decades in India? What patterns emerge across this transformative period?

Influencing Factors

Whichsocioeconomic,healthcare, and demographic factors have the strongest impact on healthy life expectancy?

Predictive Modeling

Canwe develop a reliablestatistical model to forecast HALE and support evidence-based health planning?

Project Methodology Framework

Exploratory Analysis Data Collection

Gathering comprehensive WHO datasets

Data Preprocessing

Cleaning and standardizing variables

Identifyingpatternsand relationships

Feature Engineering

Creating predictive variables

Model Development

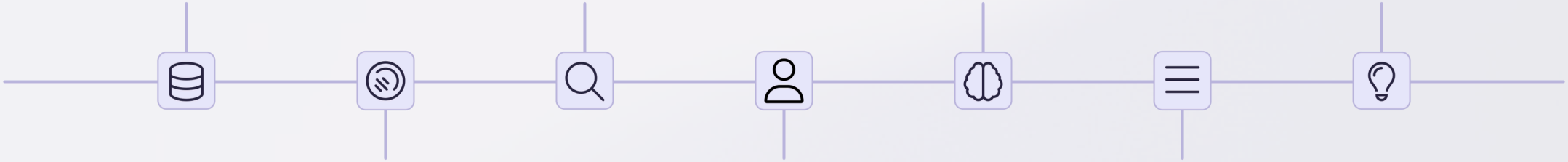
Building statistical models

Model Evaluation

Assessing accuracy and reliability

Insights & Conclusions

Translatingfindingsinto action



Data Collection Specifications

Primary Source

WHOHealthy Life Expectancy Dataset provides authoritative global health metrics with standardized methodologies across countries.

Geographic Focus

India 3 the world's second-most populous nation, representing 18% of global population

Temporal Coverage

200032019: Two decades capturing India's rapid development and healthcare expansion

Primary Metric

HALE at birth measured in years, representing expected healthy lifespan

Supporting Variables

Mortality rates, health expenditure, GDP indicators, demographic factors, and disease burden metrics

Dataset Characteristics

20
TotalObservations

Annual data points from 2000 to 2019

8+

Key Features

Predictor variablesanalyzed

1

Target Variable

HALE (Healthy Life Expectancy)

Feature Categories

Temporal & Demographic

- Year (200032019) Overall life
- expectancy at birth
- Population structureindicators

Health System Indicators

- Health expenditure (% ofGDP)
- Healthcare access metrics
- Medical infrastructure density

Mortality & Disease Burden

- Age-specificmortality rates
- Cause-specific mortality
- Disability-adjustedlife years(DALYs)

Socioeconomic Factors

- Economic developmentindicators
- Education attainment levels
- Urbanization rates

Data Preprocessing Pipeline

01

Data Cleaning

Identified andcorrected inconsistent values, formatting errors, and data entry mistakes across all variables

02

Missing Data Imputation

Applied appropriate statistical methods including mean imputation for continuous variables and mode for categorical data

03

Outlier Detection

UsedInterquartile Range (IQR) method to identify and remove anomalous observations that could skew model performance

04

Standardization

Normalizedvariablesto common scales using z-score standardization, ensuring comparability across different measurement units

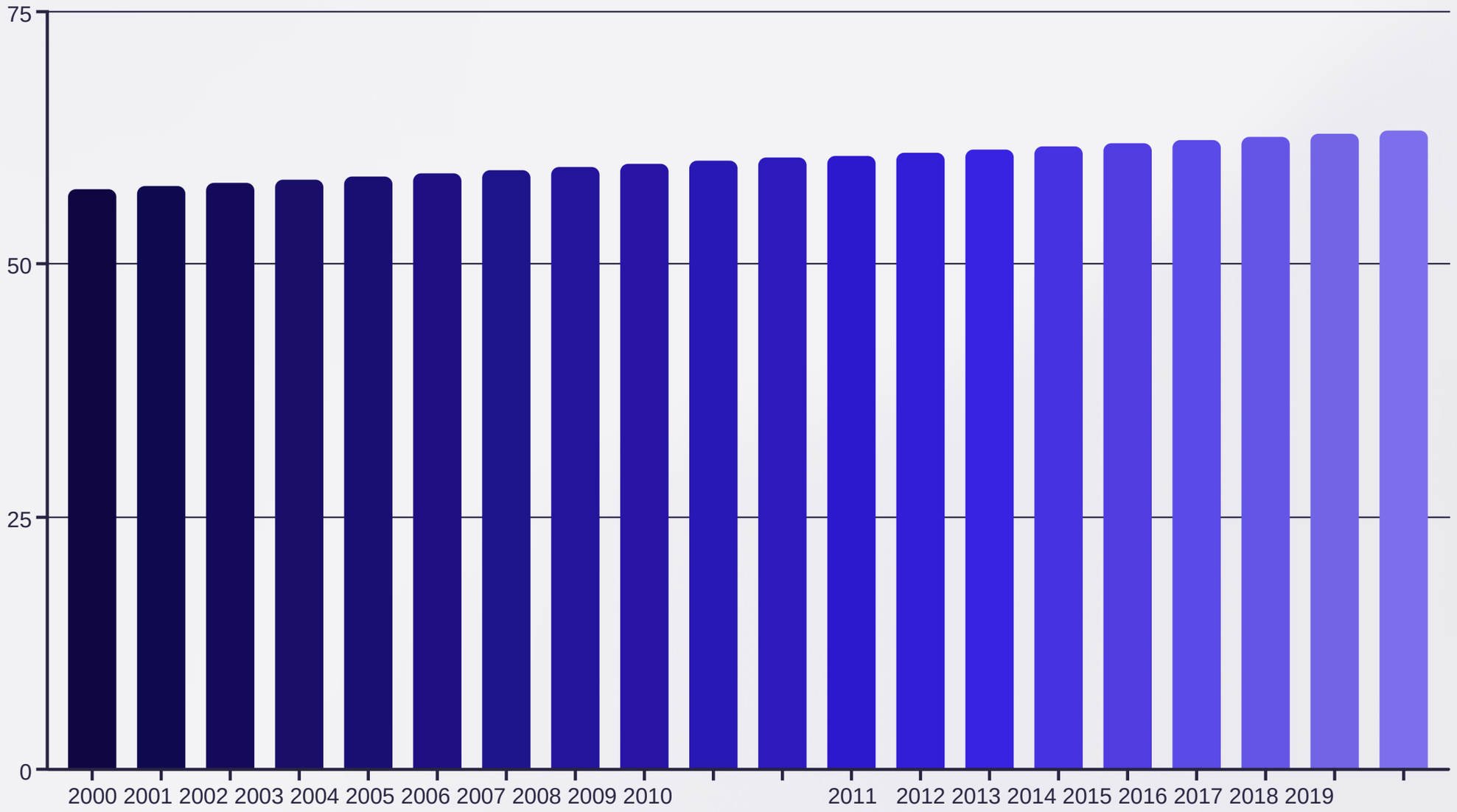
05

Feature Encoding

Converted categoricalvariables using one-hot encoding and label encoding where appropriate for modeling

HALE Trend Analysis (200032019)

India'sHealthyLife Expectancydemonstratesa consistentpositivetrend overthetwo-decade period, reflecting improvements in healthcarequality, preventivemedicineadoption, andpopulationhealth awareness.



Distribution Analysis

Statistical Characteristics

Thedistributionof HALE values acrossthe study period reveals a **near-normal distribution** with slight right skewness, suggesting consistent improvement with occasional accelerated gains.

Low Variance

Stable, steady improvement rather than dramatic fluctuations indicates sustainable health system development

Positive Skew

More frequent years with above-average improvements signal accelerating progress in recent years

Minimal Outliers

Absence of extreme deviations confirms data quality and real-world consistency

Correlation Analysis Results

Strong Positive Correlations

Life Expectancy

r=0.94 | Strong relationship confirms HALE closely tracks overall longevity improvements

Health Expenditure

r=0.78 | Investmentin healthcare infrastructure directly correlates with healthy years lived

Mortality Indicators

r=-0.82 | Inversecorrelation shows reduced mortality rates strongly predict higher HALE

Key Insights

Variables unrelatedto health systems show **weak or negligible correlations**,

confirming that HALE is primarily driven by health-specific interventions rather than general socioeconomic factors alone. This finding validates targeted health policy approaches.

Feature Engineering

Transforming rawdata intoa formatsuitable for robustpredictive modeling, our feature engineering phase focused on creating variablesthat enhance model performance and interpretability.

Derived Variables

Created new features, such as year-based scaling and ratios, to capture temporal trends and inter-variable relationships.

Scaled Numerical Features

Applied standardization techniques to numerical features, ensuring consistent scales for optimal regression model training.

Encoded Categorical Variables

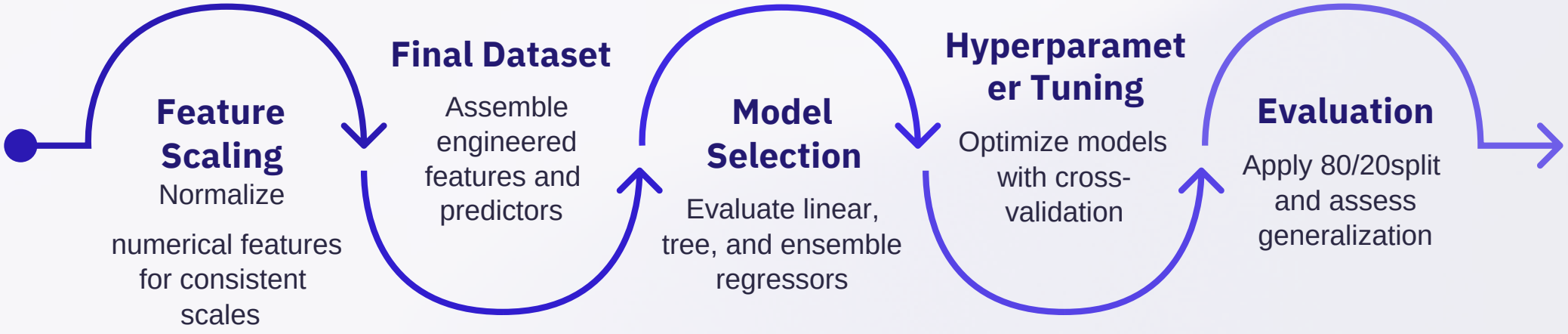
Converted non-numeric categorical attributes into numerical representations suitable for machine learning algorithms.

Prepared Final Dataset

Assembled the comprehensive dataset, integrating all engineered features and relevant predictors for model development.

Model Development

Ourpredictive modelingapproach involved evaluating several machine learning algorithms, coupled with rigorous tuning and validation techniquesto ensurerobustand accurate predictions forHALE.



Model Evaluation Metrics

Toassessthe performanceandrobustnessof ourpredictivemodels, we employed a suite of widely recognized evaluation metrics. Thesemetrics provideacomprehensiveview ofpredictionaccuracy, error distribution, and the overall explanatory power of each model.

RMSE (Root Mean Squared Error)

Measures the average magnitude of the errors. It is particularly sensitive to large errors, penalizing them more heavily, which is useful when large errors are undesirable.

MAE (Mean Absolute Error)

Calculates the average absolute difference between predicted and actual values. MAE offers a more robust error measure against outliers compared to RMSE, as it does not square the errors.

R^2 Score (Coefficient of Determination)

Indicates the proportion of variance in the dependent variable that can be predicted from the independent variables. It quantifies how well the model explains the variability of the response data around its mean.

By rigorously comparing these metrics across all developed models, we aimed to identify the best-performing algorithm that balances predictive accuracy with interpretability for Healthy Life Expectancy.

Best Model Performance

The**Random ForestRegressor**emergedasthetop-performingmodel, excelling across all critical evaluation metrics for predicting HealthyLife Expectancy.

Highest R^2 Score

Achieved the leading R^2 score, explaining the most variance in HALE outcomes.

Lowest Error Rates

Demonstrated minimal RMSE and MAE, indicating highly accurate predictions.

Non-Linear Pattern Capture

Effectively identified and modeled complex, non-linear relationships within the data.

Robust Generalization

Maintained strong performance on unseen data, ensuring reliable future predictions.

Key Insights from HALE Analysis

Ourcomprehensive analysis reveals thatHealthyLife Expectancy(HALE)demonstratedconsistent upward trends over the 20-year study period, reflecting global advancements inhealth outcomes.

Major positive drivers influencing this growth include:

Health Expenditure

Increased investment in healthcare infrastructure and services directly contributed to improved health outcomes.

Reduced Mortality

Significant declines in mortality rates, especially for preventable diseases, extended years lived in good health.

Improved Life Expectancy

Overall gains in lifespan, combined with better health quality, positively impacted HALE across populations.

These findings collectively suggest the effectiveness of targeted public health interventions and ongoing socioeconomic improvements in fostering a healthier, longer-living global population.

Conclusions

Consistent Performance

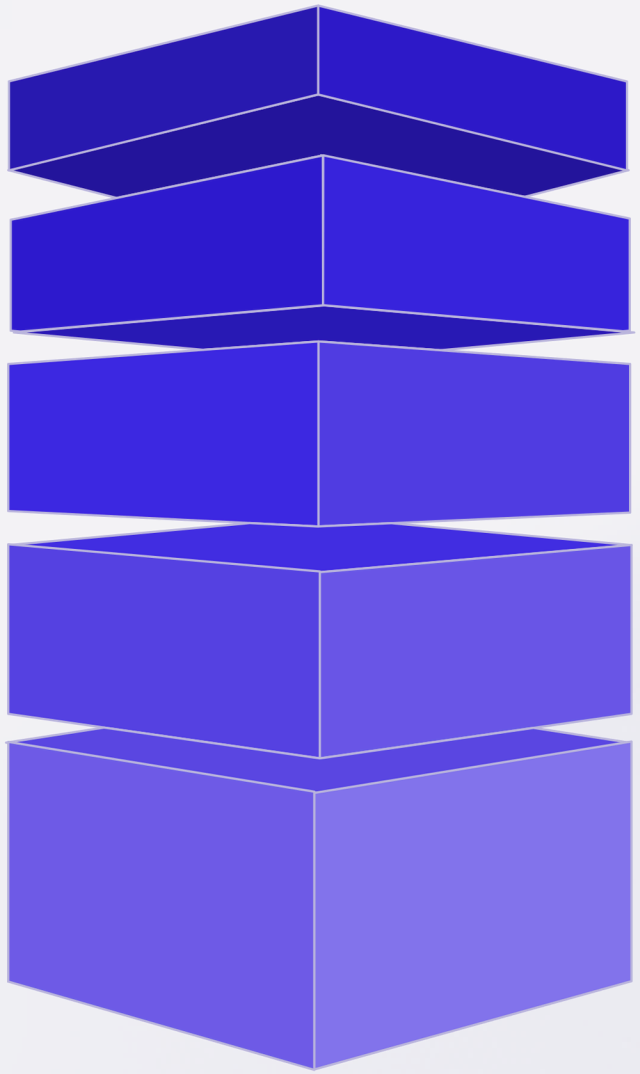
Maintaining highstandards and reliable outcomes over time.

Reduced Mortality

Lowering mortality rates through effective health interventions.

Overall Conclusion

Summarizing thepositive impact on public well-being.



Health Expenditure

Strategicinvestment in healthcare resources and infrastructure.

Improved Life Expectancy

Increasing the average lifespan for the population.

Future Work & Thank You

Our analysisprovides asolidfoundationforunderstandingHealthyLife Expectancy. To further enhance its depth and applicability, we plan toimplementthe followingkeyadvancements:

Global Comparisons

Expand the analysis to include multi-country comparisons, identifying global trends and regional disparities in HALE determinants.

Behavioral & Environmental Factors

Integrate a wider array of behavioral (e.g., diet, exercise) and environmental factors (e.g., pollution, climate) for a holistic view.

Advanced Predictive Models

Apply more sophisticated machine learning models, including XGBoost and Time-Series Forecasting, to improve predictive power and temporal insights.

Interactive Dashboard Development

Create a dynamic and user-friendly interactive dashboard using tools like notebook for easy exploration of HALE data.

Thank You

We appreciate your attention and interest in this research. For further inquiries or collaboration, please feel free to reach out:

Contact: piyu.143247@**gmail.com**