

Regression Model Results and HPV Feature Overview

Sainath

June 5, 2025

Low_cin_combined Results - Regression Models

Model	Train R^2	Test R^2	Train MSE	Test MSE
Ridge Regression	0.1527	-2.4809	101.7097	95.7018
Lasso Regression	0.1511	-2.2133	101.9043	88.3462
Random Forest	0.8621	-4.0147	16.5584	137.8725
Support Vector	-0.0392	0.1046	124.7549	24.6178

High_cin_combined Results - Regression Models

Model	Train R^2	Test R^2	Train MSE	Test MSE
Random Forest	0.8294	-0.3623	18.3394	129.7479
XGBoost Regressor	1.0000	-1.3537	0.0000	224.1666
Ridge Regression	0.7455	-1.0004	27.3616	190.5146

Selected Features Related to HPV

The following 27 features were selected for HPV-related analysis:

Country

Mean targeted age

Male circumcision (WHO 2007)

Condom Use

Mortality rates

Gross National Income (GNI) per capita

Population estimate

Total Fertility rate (2017)

Contraception use (updated 2019)

HIV Prevalence (in adults)

Sexual Initiation age

Reported STIs (in%)

Mean marital age

Multiple pregnancies (%age)

More Selected Features

Smoking Prevalence (females, 2016)

Anemia prevalence among women 15–49

Anemia prevalence among pregnant women

Physicians per 1,000 people

Incidence of TB

Diabetes Prevalence

Hypertension

Coverage ever screened (30–49 years)

Coverage last 5 years (30–49 years)

Coverage last 3 years (30–49 years)

Coverage ever screened (25–65 years)

Coverage last 5 years (25–65 years)

Coverage last 3 years (25–65 years)

Columns have high Missing Values

The following features had a high proportion of missing values and were dropped:

- Reported STIs (in%)

- Sexual Initiation age

- Multiple pregnancies (%age)

- Smoking Prevalence (females, 2016)

- Physicians per 1,000 people

Additional Dropped Columns

Anemia prevalence among non-pregnant women

Continent

Sample size studied

Start of Screening coverage (year)

HPV vaccination introduction

Age-adjusted incidence (standardized rates)

Number of deaths (all ages, 2021)

HDI, Life expectancy, Schooling years

HPV Vaccine