

Meta Scifor Mini Project

Presented by Aryan Malviya

Project 5.0 - World Happiness Report

Problem Statement

- The World Happiness Report is a landmark survey of the state of global happiness.
- The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.



About the Dataset

▶ data

→

	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
0	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	2.51738
1	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630	2.70201
2	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	2.49204
3	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699	2.46531
4	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	2.45176
...
153	154	3.465	0.03464	0.22208	0.77370	0.42864	0.59201	0.55191	0.22628	0.67042
154	155	3.340	0.03656	0.28665	0.35386	0.31910	0.48450	0.08010	0.18260	1.63328
155	156	3.006	0.05015	0.66320	0.47489	0.72193	0.15684	0.18906	0.47179	0.32858
156	157	2.905	0.08658	0.01530	0.41587	0.22396	0.11850	0.10062	0.19727	1.83302
157	158	2.839	0.06727	0.20868	0.13995	0.28443	0.36453	0.10731	0.16681	1.56726

About the Columns

- The following columns: GDP per Capita, Family, Life Expectancy, Freedom, Generosity, Trust Government Corruption describe the extent to which these factors contribute in evaluating the happiness in each country.
- The Dystopia Residual metric actually is the Dystopia Happiness Score(1.85) + the Residual value or the unexplained value for each country.



Happiness Score Prediction process

Data Preprocessing

- Loaded the dataset using pandas and conducted an initial exploration to understand the data structure and identify relevant features.
- Handled data cleaning, verified there were no missing values, and confirmed data types for consistency.

Feature Selection

- Selected key features that contribute to the Happiness Score: Economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), Generosity, and Dystopia Residual.
- Dropped irrelevant columns like Country and Region to focus only on numerical predictors for the model.

Model Training

- Split the dataset into training and testing sets (80% training, 20% testing) to validate the model's performance.
- Trained a Linear Regression model using the selected features and evaluated its performance using Mean Squared Error (MSE) and R-squared (R^2) metrics.

Evaluation

- Achieved an R-squared score close to 1, indicating a near-perfect fit.



Linear Regression

- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It aims to fit a straight line (best-fit line) through the data points that minimizes the difference between the predicted and actual values.
- The model's performance is typically evaluated using metrics like Mean Squared Error (MSE) and R-squared (R^2).

Mean Squared Error (MSE)

Mean Squared Error (MSE) measures the average squared difference between predicted and actual values. A lower MSE indicates better model performance, as it means the predictions are closer to the actual values.

R-squared (R^2)

R-squared (R^2) indicates how well the independent variables explain the variation in the dependent variable. An R^2 value closer to 1 means the model fits the data well.

Evaluation

```
[12] y_pred = model.predict(x_test)

[13] mse = mean_squared_error(y_test, y_pred)
     r2 = r2_score(y_test, y_pred)

[14] print(f'Mean Squared Error (MSE): {mse}')
     print(f'R-squared Score (R2): {r2}')

→ Mean Squared Error (MSE): 7.427424037464989e-08
   R-squared Score (R2): 0.9999999476481373

[15] print('Model Coefficients:', model.coef_)
     print('Intercept:', model.intercept_)

→ Model Coefficients: [-4.58542330e-06 -1.03770291e-03  9.99934281e-01  9.99802806e-01
   9.99686551e-01  9.99643097e-01  9.99836310e-01  9.99736954e-01
   9.99851761e-01]
   Intercept: 0.0013999091298630972
```

Evaluation Results

- MSE Value:- This very low error value suggests that the model's predictions are extremely close to the actual values.
- An R^2 value near 1 means that almost all the variability in the Happiness Score is explained by the model, indicating a near-perfect fit.
- Model Coefficients:- The coefficients are close to 1 for most features, indicating that these variables have a strong linear relationship with the Happiness Score. The specific values suggest how much the Happiness Score would change with a unit change in each feature.
- Intercept Value — This small value represents the baseline Happiness Score when all features are zero.



**Thank you
very much!**