

WIE2003 INTRODUCTION TO DATA SCIENCE

SLEEP QUALITY PREDICTION

GROUP 11

PRESENTED BY:

ZHI WEI, JIN SZE, PANG LANG, SILVIA, ELAINE



Project Objectives



To train and evaluate multiple machine learning models on the provided sleep dataset in order to determine the optimal model

To conduct comprehensive feature importance analysis and interpretation to identify the most significant factors and variables that influence sleep quality.

To develop a data product that leverages the trained predictive model to provide personalized sleep quality predictions and recommendations for users

DATA MODELLING

4 different regression models

- Polynomial Regression
- Support Vector Machine
- Decision Tree Regression
- Random Forest Regression

3 different evaluation metrics

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score (Coefficient of Determination)

POLYNOMIAL RIGDE REGRESSION

- regression technique that extends traditional linear regression by adding polynomial terms to the model and incorporating a regularization term known as ridge regression.
- IMPROVED PREDICTIVE PERFORMANCE
- VERSATILITY
- INTERPRETABILITY

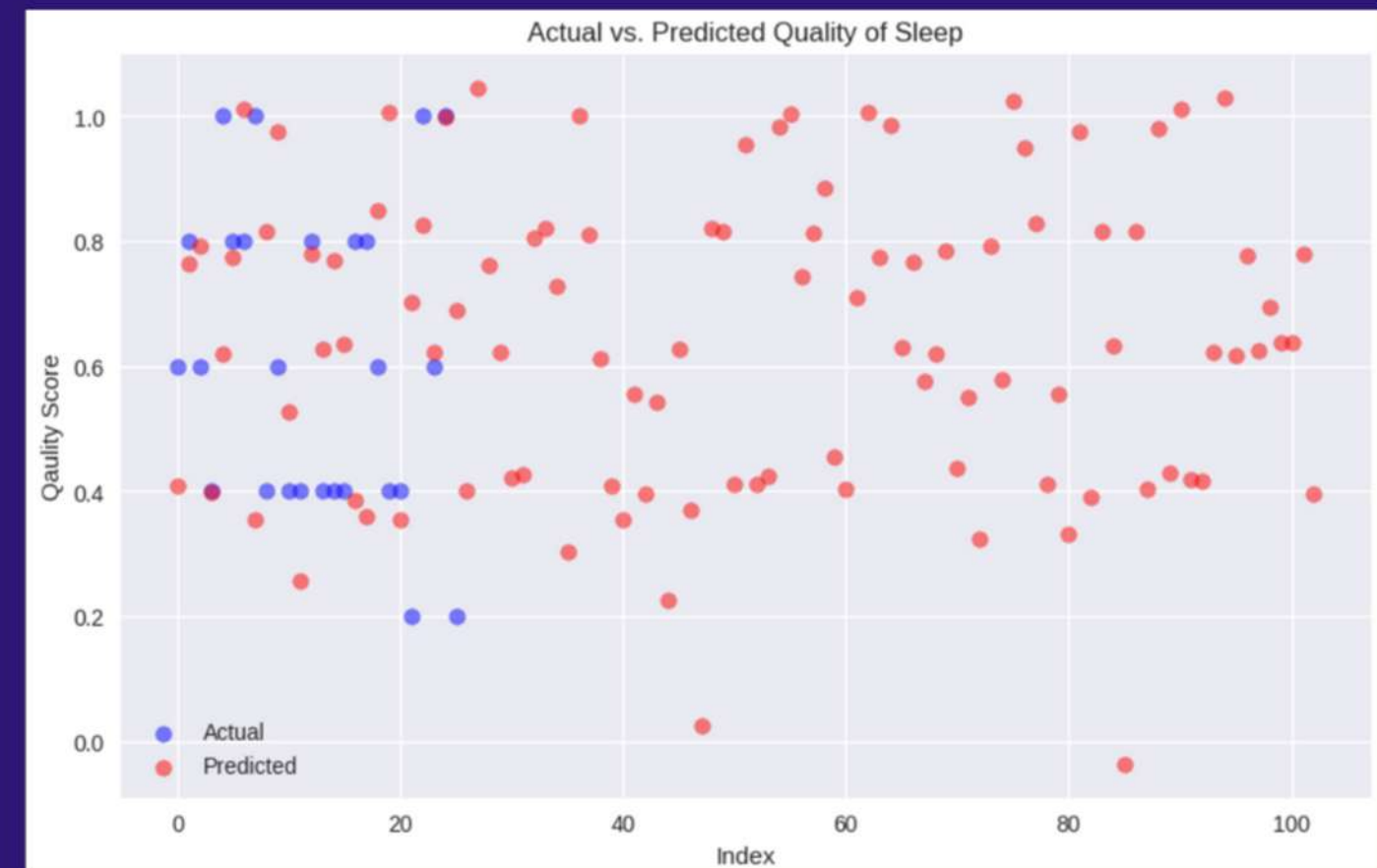
1. Model Building:

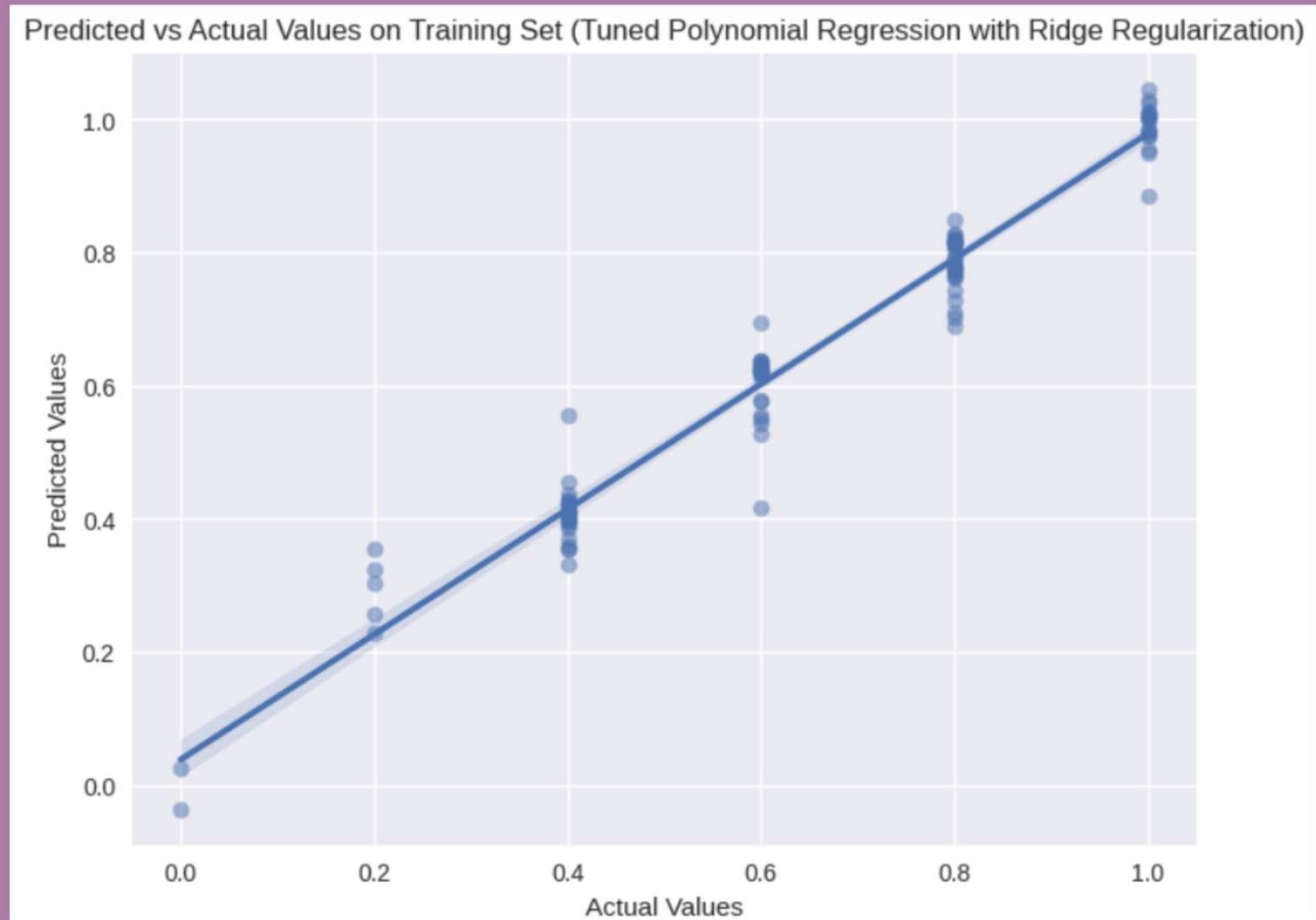
Mean Squared Error (Training): 0.00236

Root Mean Squared Error (Training): 0.04858

R² Score (Training): 0.96199

2. Visualisation Prediction:





Best Hyperparameter:

Mean Squared Error: 0.00236

Root Mean Squared Error: 0.04858

R^2 Score: 0.96199

Support Vector Regression

- Aims to find the hyperplane that passes through as many data points as possible within a certain distance, called the margin.
- Reduce the prediction error
- Allow SVR to handle non-linear relationships between input variables and the continuous target variable.

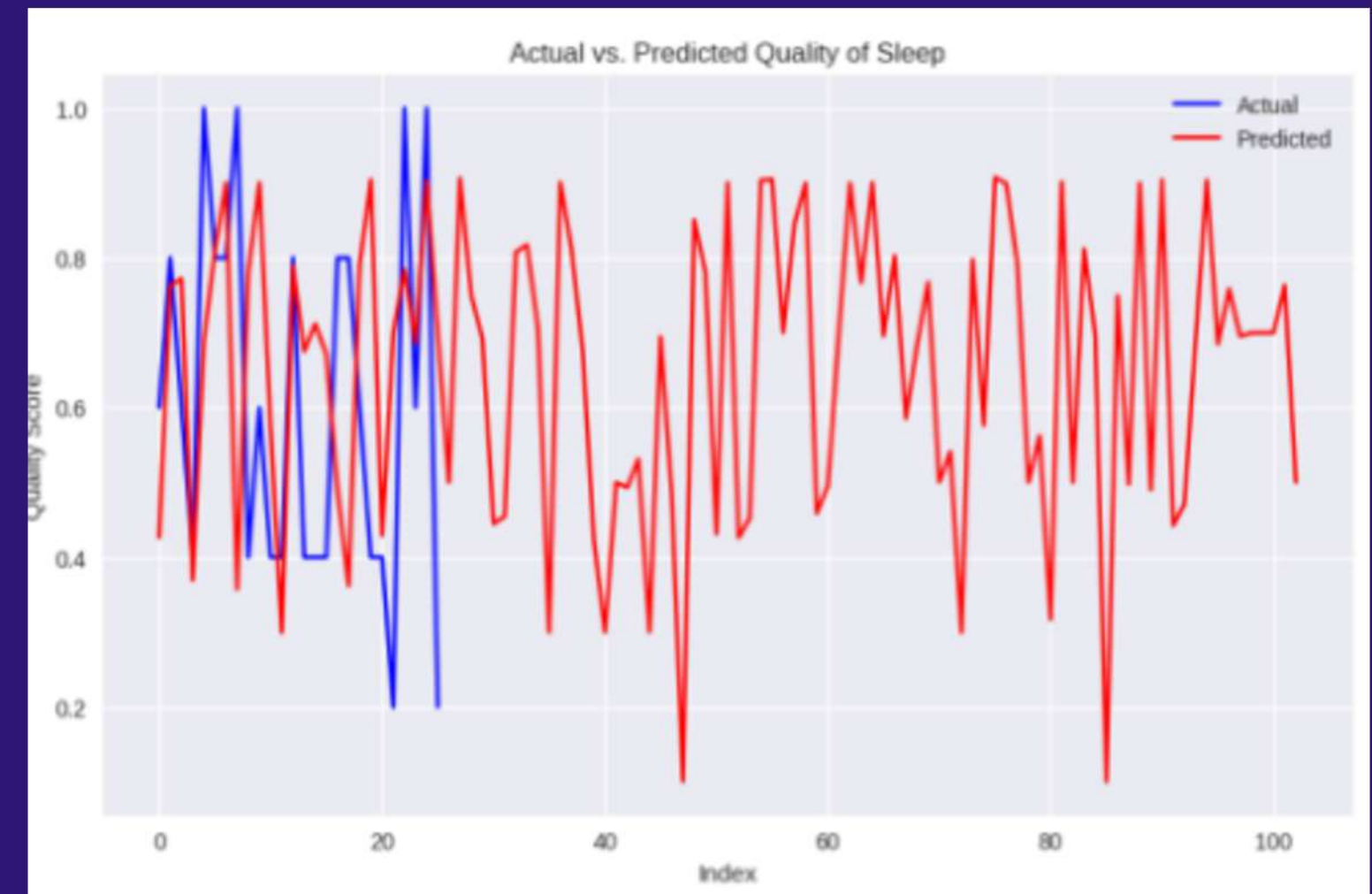
1. Model Building:

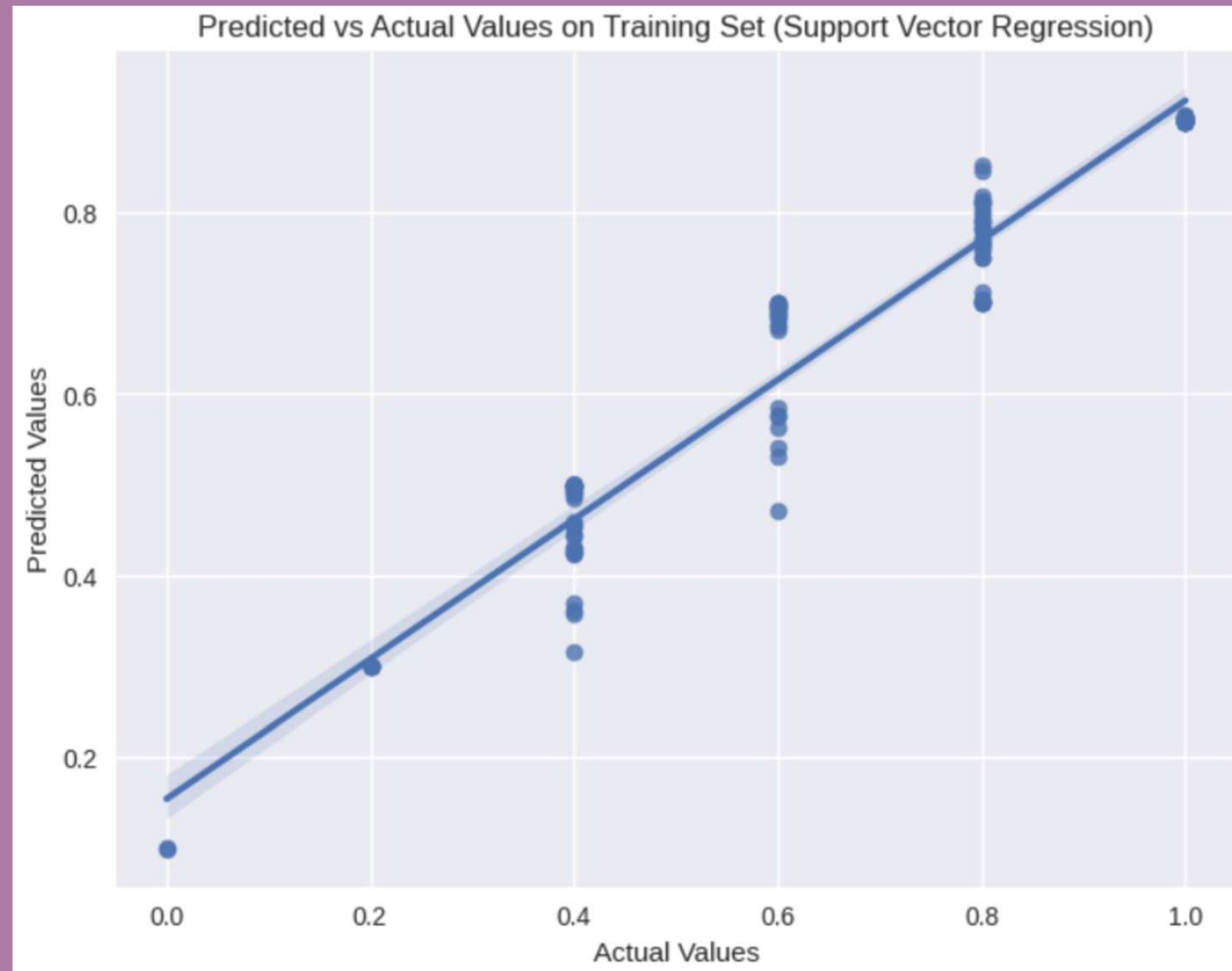
Mean Squared Error (Training): 0.00595

Root Mean Squared Error (Training): 0.07717

R² Score (Training): 0.90409

2. Visualization Prediction:





Best Hyperparameter:

Mean Squared Error: 0.00280

Root Mean Squared Error: 0.05288

R^2 Score: 0.95496

Decision Tree Regression



The model Decision tree with max depth of 3 serve a purpose to balance model complexity and overfitting by fitting model to training data and make predictions using evaluation metrics which assess in prediction accuracy and fit. Max depth adjustment controls model complexity and generalization ability

1. Model Building

Mean Squared Error (Training):

0.0031551296599840284

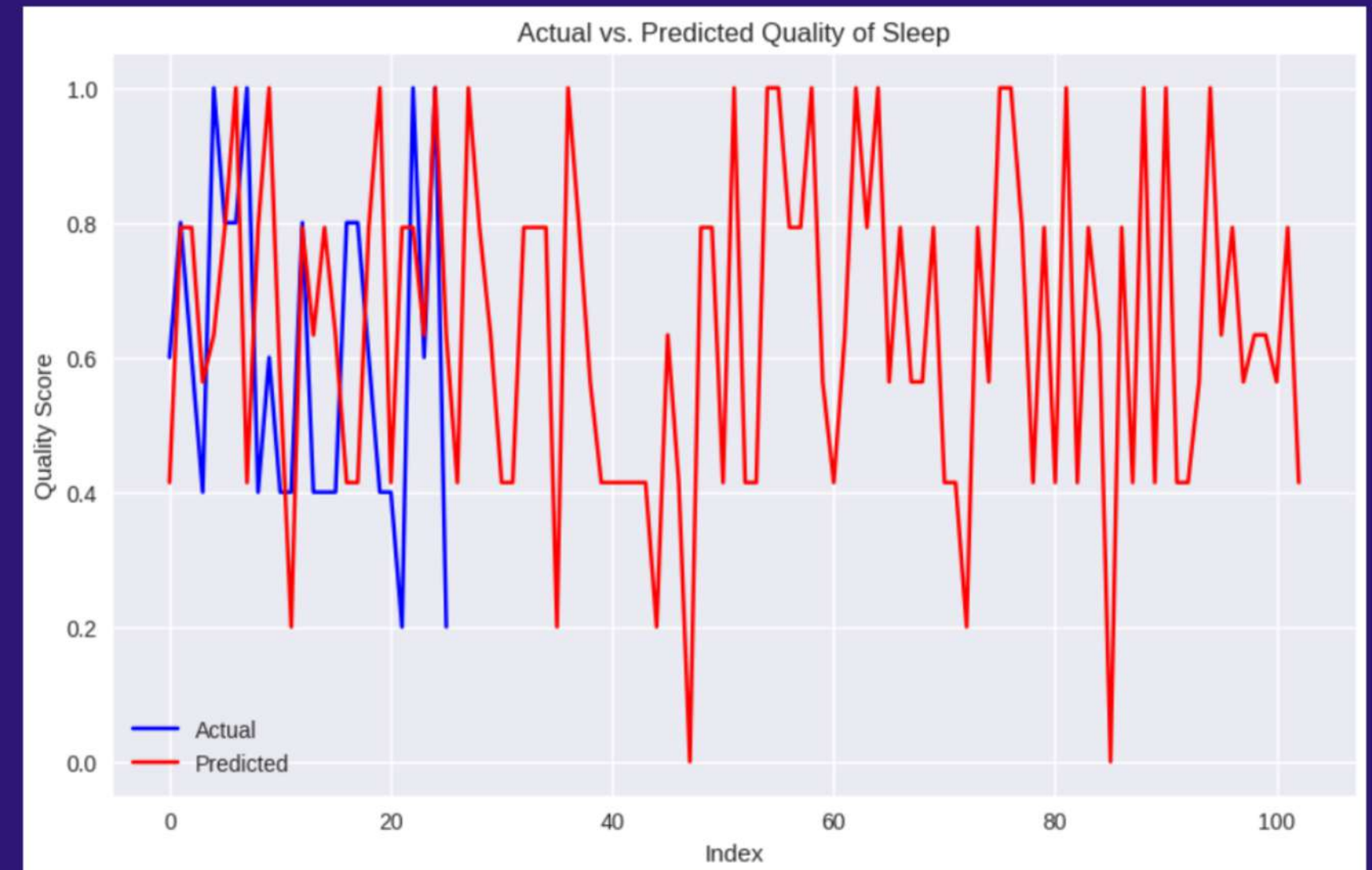
Root Mean Squared Error (Training):

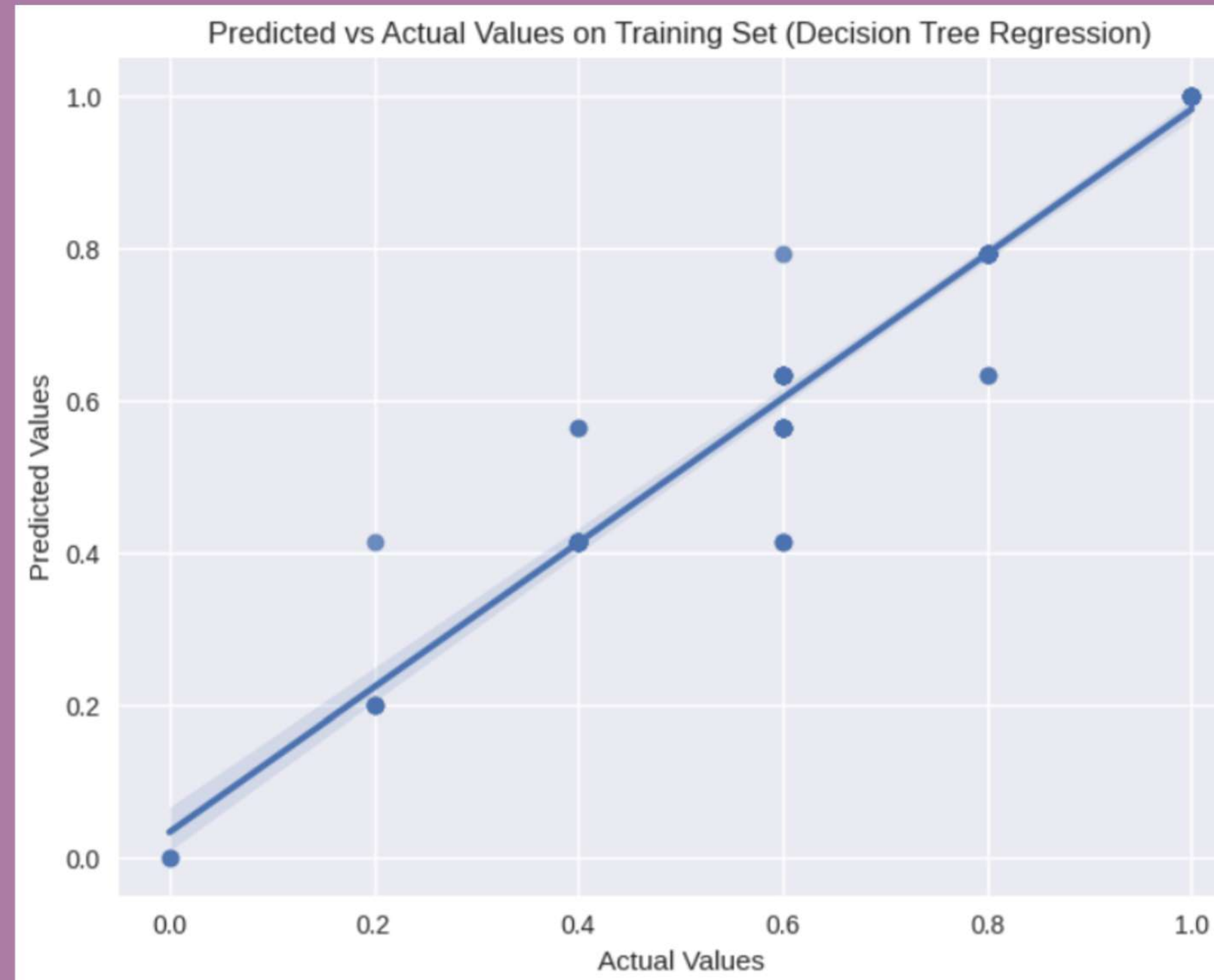
0.056170540855363216

R² Score (Training):

0.9491789588200374

2. Visualisation Prediction





Best Hyperparameter:

Mean Squared Error: 0.00078

Root Mean Squared Error: 0.02787

R^2 Score: 0.98748

Random Forest Regressor



1. Model Building:

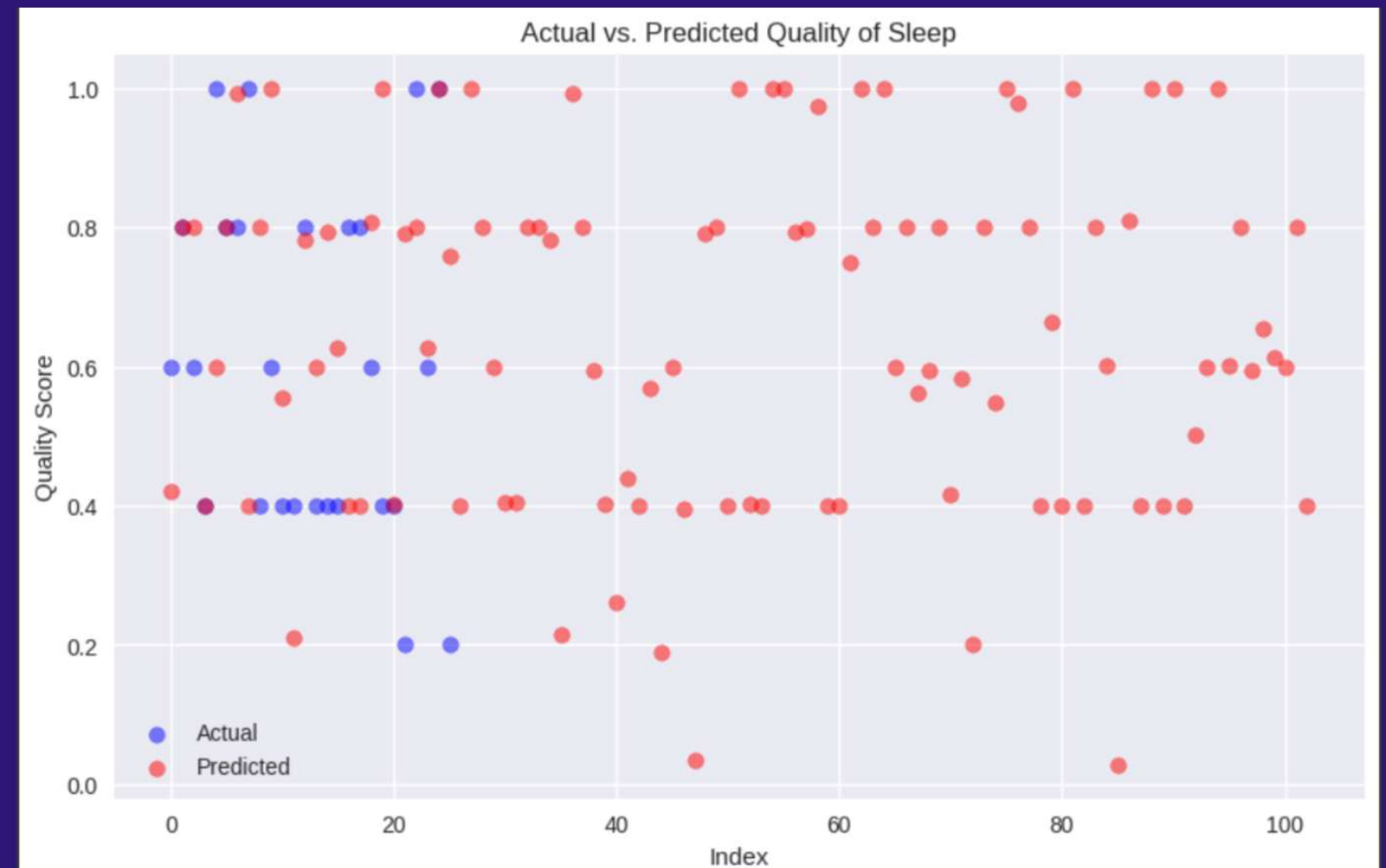
Mean Squared Error (Training): 0.00040

Root Mean Squared Error (Training): 0.01993

R^2 Score (Training): 0.99360



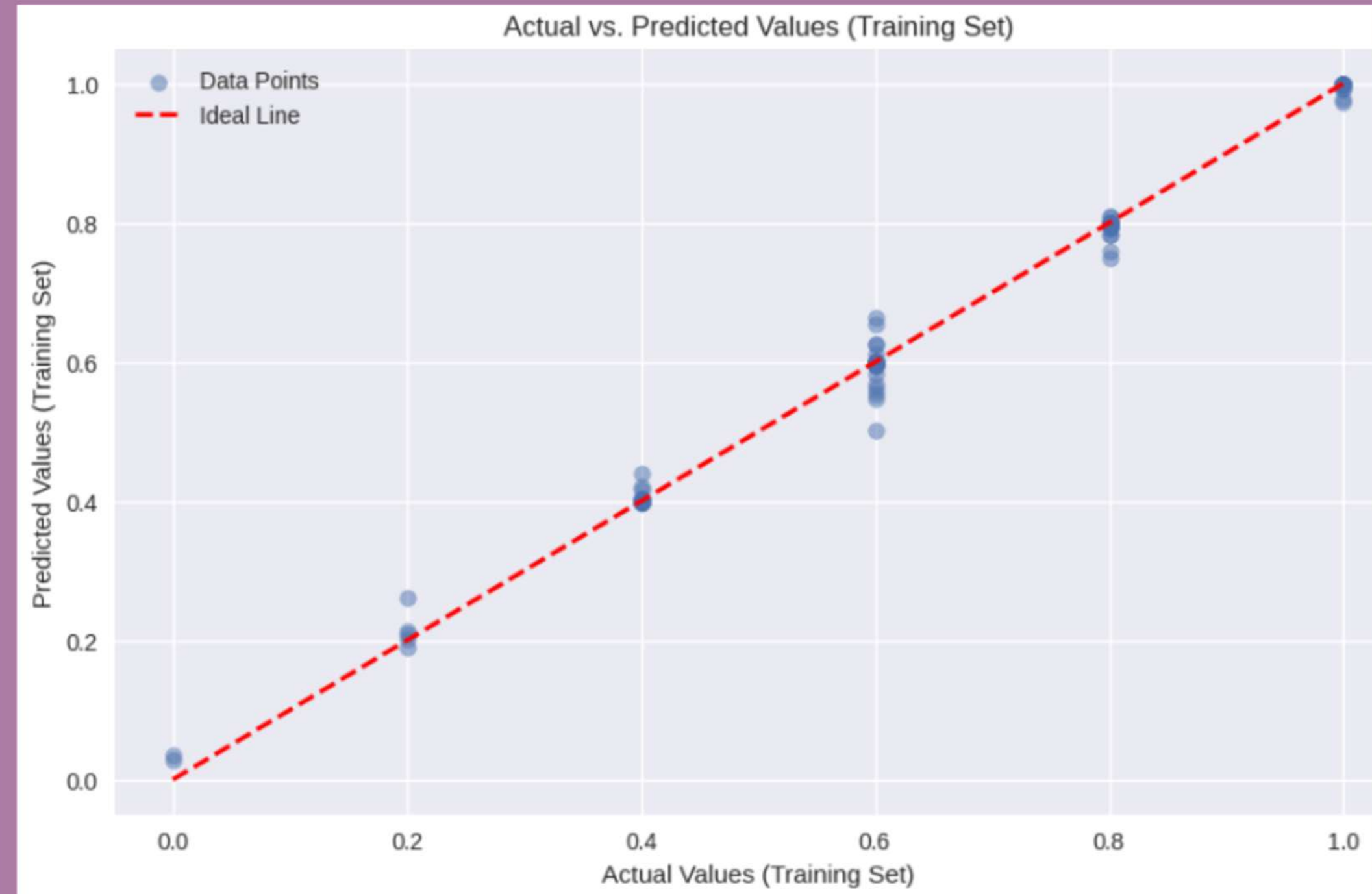
2. Visualization Prediction:



Hyperparameter tuning by using Grid search cross-validation

The best accuracy of predictive model:

1. Mean Squared Error (MSE): 0.00037
2. Root Mean Squared Error (RMSE): 0.01931
3. R^2 Score: 0.99399



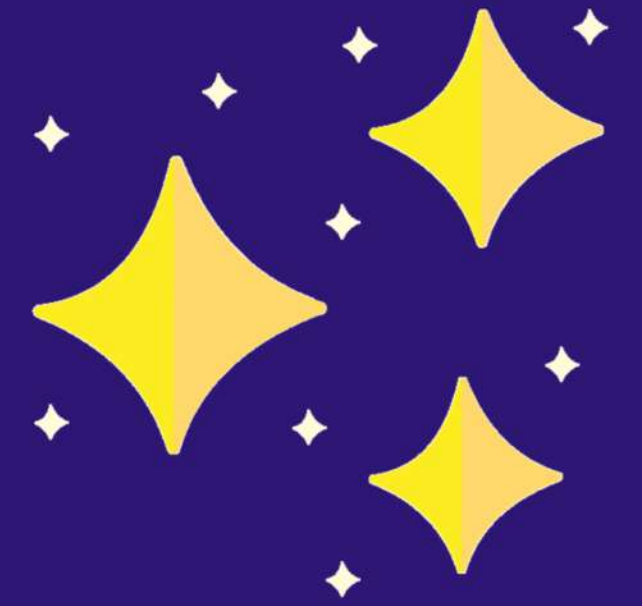
Choose the best model and evaluate on test set

Lists of MSE, RMSE and R2 scores for different models

	MSE	RMSE	R2-Score
Polynomial Ridge Regression	0.002360	0.048576	0.961993
Support Vector Regression	0.002796	0.052880	0.954959
Decision Tree Regression	0.000777	0.027869	0.987489
Random Forest Regression	0.000373	0.019308	0.993995



Model Testing



Tuned Random Forest Regression - Testing Set

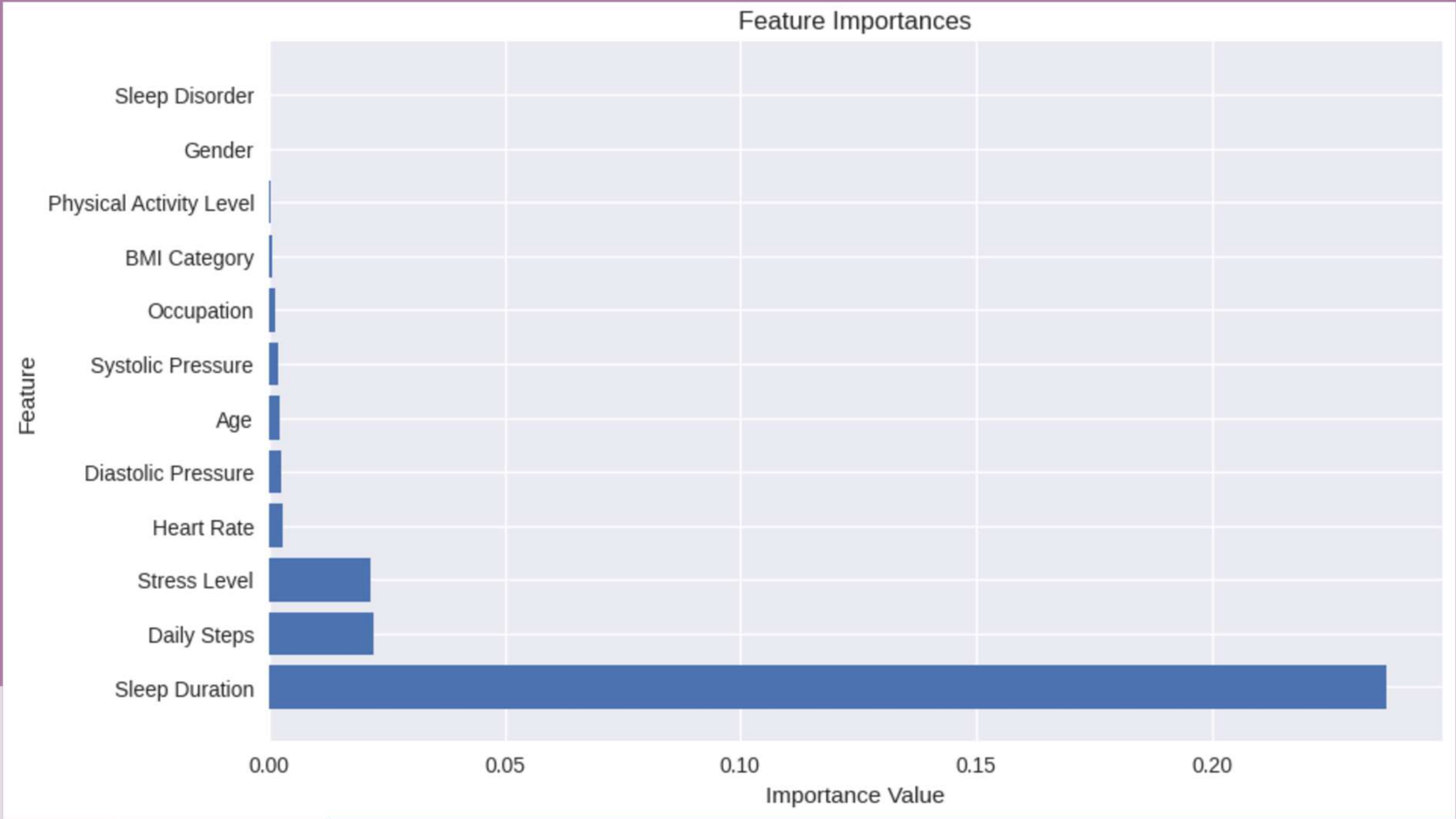
- Mean Squared Error: 0.00470
- Root Mean Squared Error: 0.06859
- R² Score: 0.92152

	MAE	R2-Score
Train	0.000373	0.993995
Test	0.004700	0.921520

A child is shown from the chest up, looking upwards with a sense of wonder. The background is a dark, starry night sky. A large, stylized white cloud graphic is positioned to the right of the child's head. The text 'DATA INTERPRETATION' is centered in the image, enclosed in a white rounded rectangular frame with a dotted line border. The word 'DATA' is on the top line and 'INTERPRETATION' is on the bottom line, both in a bold, white, sans-serif font.

DATA INTERPRETATION

FEATURES IMPORTANCE



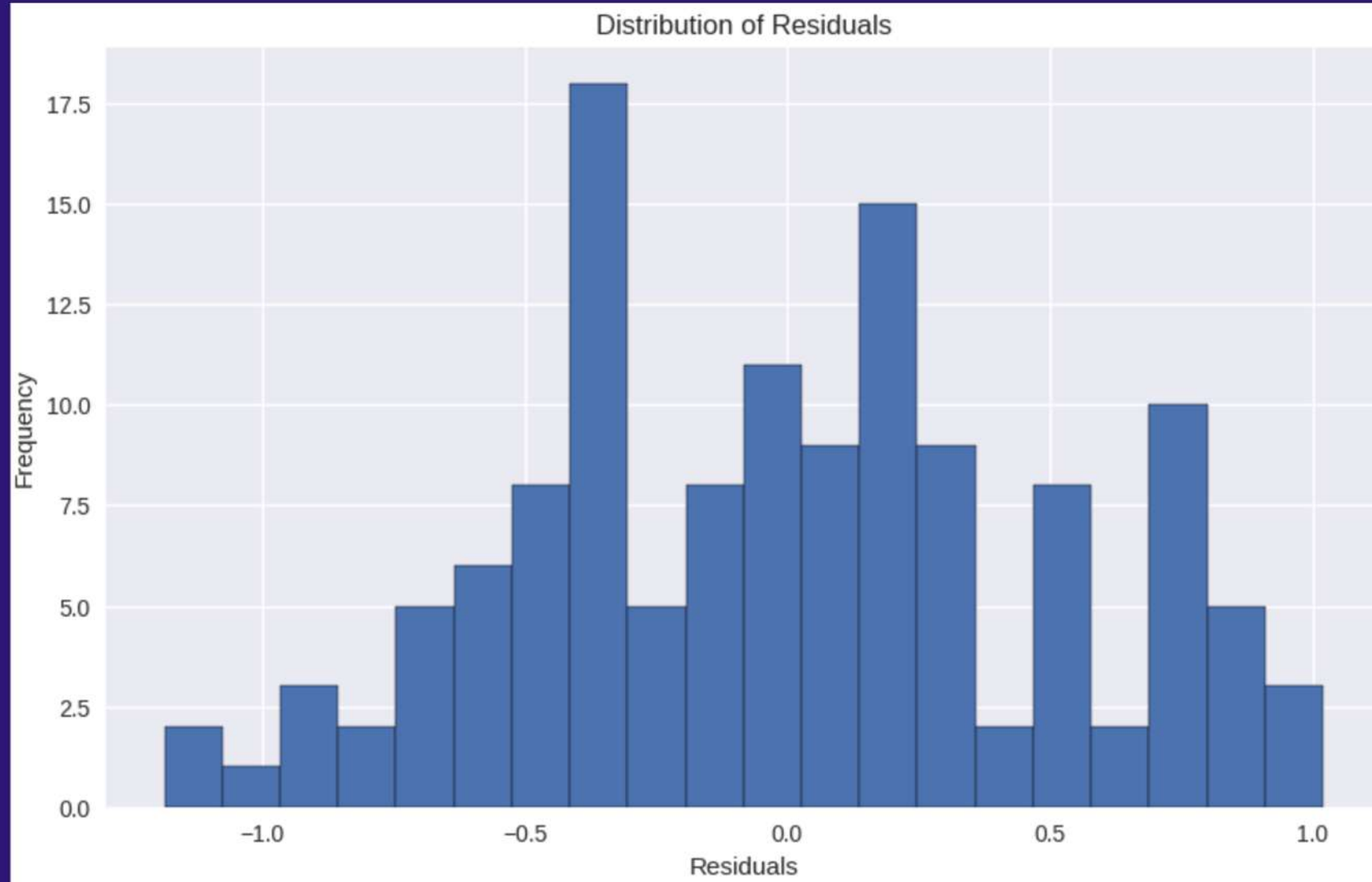
TOP 3 SIGNIFICANT FEATURES

Sleep Duration - 0.237

Daily Step - 0.022

Stress Level - 0.0216

RESIDUAL ANALYSIS

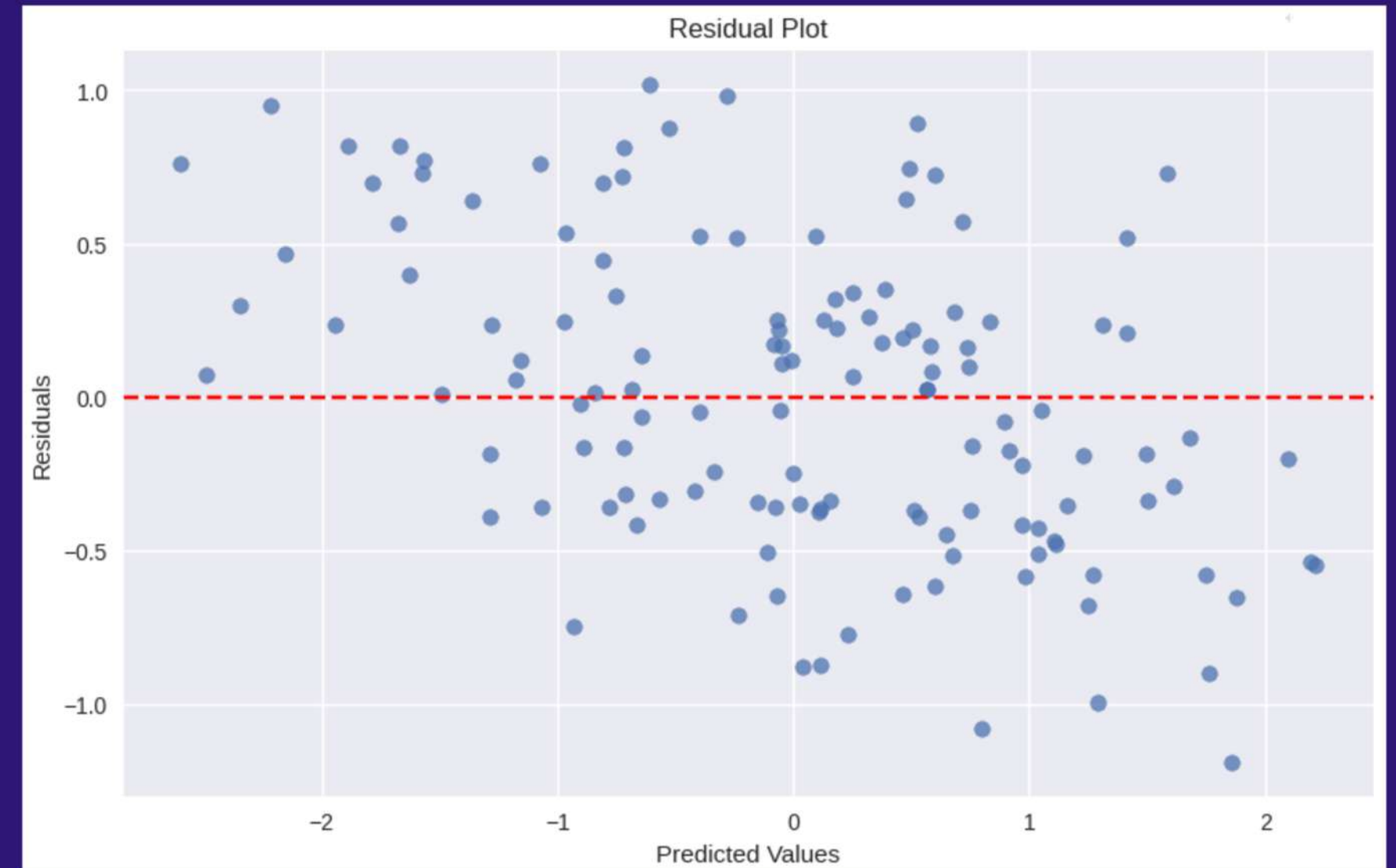


**APPROXIMATELY BELL-SHAPED
(NORMAL DISTRIBUTION)**

- **have few outliers or extreme values**
- **Slightly negative skew**

More negative residuals suggest the model may slightly overestimate observed values on average.

RESIDUAL PLOT



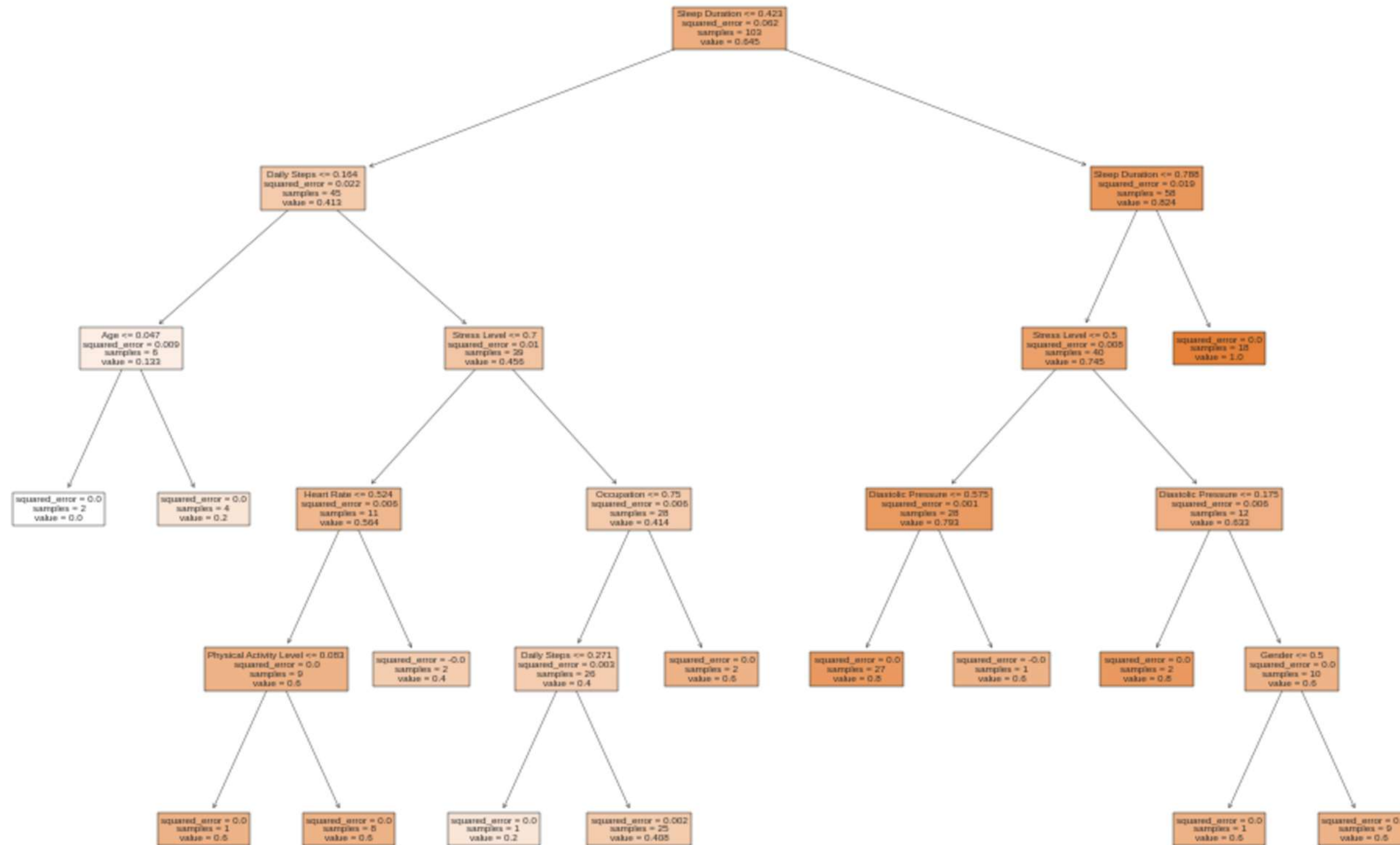
RANDOMLY SCATTERED AROUND ZERO LINE

- **few potential outliers with large residuals**
- **slight asymmetry**

Higher concentration of positive residuals suggests a potential bias in the model, with predictions tending to be higher than actual values.

BEST MODEL VISUALIZATION

Random Forest - Decision Tree Visualization



A child is shown from the chest up, holding a handful of star-shaped confetti. The background is a soft-focus image of the child's face and hands, overlaid with a dark purple tint. A white cloud graphic is positioned to the right of the text. The text 'OUR DATA PRODUCT' is centered within a white rectangular frame that has rounded corners and a dotted border.

OUR DATA PRODUCT

One of the most critical factors affecting sleep

Physical activity levels, often measured by daily steps, are closely linked to sleep quality.

Directly related to sleep quality. High stress can lead to difficulty falling asleep and staying asleep.

An indicator of overall cardiovascular health. It can influence sleep quality.

Affects sleep patterns and requirements.

An indicator of body weight relative to height. It can influence sleep quality, with both high and low BMI associated with different sleep issues.

Sleep Health Analysis

Input and Relationship with Sleep Quality

Sleep Duration: 7.5 hours
Your sleep duration of 6 to 8 hours is generally associated with good sleep quality and overall health. Maintain it!

Stress Level: 2
Low stress levels are generally associated with better sleep quality, as stress can significantly impact how well you sleep.

Stress Level: 2
Low stress levels are generally associated with better sleep quality.

Heart Rate: 70 bpm
Normal resting heart rate is generally associated with good sleep quality.

Age: 30 years
Adults between 18 and 60 years old typically need 7-9 hours of sleep each night to maintain good health and cognitive function.

BMI: 25
Your BMI is 25 or higher, indicating overweight or obesity, which can increase the risk of sleep disturbances, such as sleep apnea, and related health problems.

Overall Sleep Quality:
You have a fair sleep quality! With some adjustments, you can move from fair to good or even excellent sleep quality, significantly enhancing your overall well-being and daytime functioning.

Suggestions for Improvement:
Here are some tips and strategies that might help you improve your sleep quality:
- Increase your daily steps.
- Exercise more and maintain a balanced diet.

http://127.0.0.1:7742 Open in Browser

Sleep Health Analysis

Input and Relationship with Sleep Quality

Sleep Duration: 7.5 hours
Your sleep duration of 6 to 8 hours is generally associated with good sleep quality and overall health. Maintain it!

Stress Level: 2
Low stress levels are generally associated with better sleep quality, as stress can significantly impact how well you sleep.

Stress Level: 2
Low stress levels are generally associated with better sleep quality.

Heart Rate: 70 bpm
Normal resting heart rate is generally associated with good sleep quality.

Age: 30 years
Adults between 18 and 60 years old typically need 7-9 hours of sleep each night to maintain good health and cognitive function.

BMI: 25
Your BMI is 25 or higher, indicating overweight or obesity, which can increase the risk of sleep disturbances, such as sleep apnea, and related health problems.

Overall Sleep Quality:

You have a fair sleep quality! With some adjustments, you can move from fair to good or even excellent sleep quality, significantly enhancing your overall well-being and daytime functioning.

Suggestions for Improvement:

Here are some tips and strategies that might help you improve your sleep quality:

- Increase your daily steps.
- Exercise more and maintain a balanced diet.

Input Fields:

Sleep Duration (hours): Slider from 0 to 12, set at 7.5.

Daily Steps Estimation: Slider from 0 to 20,000, set at 3,000.

Stress Level (from low to high): Slider from 1 to 5, set at 2.

Heart Rate (bpm): Input field with value 70.

Age: Input field with value 30.

BMI: Input field with value 25.

Predict

- Presents a detailed analysis of each input factor and its relationship with sleep quality.
- Users receive personalized feedback on how each factor contributes to their overall sleep health.
- Provides users with an overall assessment of their sleep quality based on the input factors.
- Typically categorized into three levels: poor, fair, or excellent.
- Users receive actionable suggestions tailored to their specific input data to help them improve their sleep quality.

Insights & Conclusion



Surprising Finding from feature important analysis:

- 1.Sleep Duration
- 2.Daily Steps
- 3.Stress Levels



Model Performance:

- Best Model:Random Forest
- High performance indicates robustness and reliability.



Need for Further Research:

- Variables like BMI and sleep disorders may influence sleep quality through intricate relationship.
- Deeper research could reveal how different factors collectively impact sleep quality.

