Machine Learning Project Report

Project Title: Social Media & Well-Being Prediction using Machine Learning

Problem Statement:

To predict a user's Happiness Index using behavioral and lifestyle factors such as screen time, stress levels, sleep quality, and exercise frequency.

Dataset Description:

• Total Rows: 14,495

• Total Columns: 10

• Features include Age, Gender, Screen Time, Sleep Quality, Stress Level, Days Without Social Media,

Exercise Frequency, Social Media Platform, and more.

• Target Variable: Happiness Index (1–10).

EDA Summary:

• Null values, duplicates inspected and handled.

• Distributions plotted for numerical variables.

• Outliers detected using IQR & boxplots.

• Correlation heatmap analyzed.

Models Used:

• Linear Regression

• Random Forest Regressor

• XGBoost Regressor

• LightGBM Regressor

• Preprocessing included OneHotEncoding, scaling, and train-test split (80/20).

Performance Summary:

Baseline model error: ~46% (improved with XGBoost & LightGBM).

Future Improvements:

• Hyperparameter tuning (GridSearch/Optuna).

• Feature engineering.

• ANN experimentation.

• Model deployment using Streamlit/Flask.