# Lifestyle Score & Health Type Prediction using Machine Learning

PROJECT REPORT

submitted by

**DEVARSH S R**

to

**IPCS GLOBAL**

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **IPCS Global, Kollam** for providing the opportunity, guidance, and resources required to successfully complete this project titled *“Lifestyle Score & Health Type Prediction using Machine Learning.”* The training environment and structured learning framework offered by the institute played a significant role in enhancing my understanding of data science and machine learning concepts.

I extend my appreciation to the faculty and trainers at IPCS Global for their consistent support and mentorship throughout the course. Their expertise, feedback, and encouragement were instrumental in helping me develop both the theoretical knowledge and practical skills necessary for executing a complete end-to-end machine learning project.

I am also thankful for the access to learning materials, technical sessions, and hands-on workshops provided by the institute, which greatly contributed to strengthening my analytical approach and project handling capabilities. The experience of working through real-world concepts, datasets, and deployment techniques has been invaluable for my academic growth and professional development.

Finally, I acknowledge the collaborative learning environment at IPCS Global, which motivated me to explore, experiment, and implement the various components of this project with confidence. This project stands as a reflection of the comprehensive guidance and support I received during my training period.

# ABSTRACT

This project presents a machine learning–based system designed to evaluate an individual's lifestyle quality using measurable behavioral, physiological, and wellness-related parameters. A comprehensive synthetic dataset consisting of 200,000 records was generated to simulate realistic human lifestyle patterns, including variables related to sleep, physical activity, diet, stress, screen time, and daily habits. Through extensive exploratory data analysis, the most influential features affecting lifestyle outcomes were identified and used to build two predictive models: a regression model to estimate a numerical Lifestyle Score (0–100) and a classification model to categorize individuals into Healthy, Moderate, or Unhealthy lifestyle groups.

Multiple algorithms were tested and compared, including Linear Regression, Random Forest, Support Vector Machines, Logistic Regression, and K-Nearest Neighbors. The final system incorporates the best-performing models, achieving strong accuracy and generalization on unseen data. Both models were deployed through an interactive Streamlit application, enabling users to input their lifestyle information and receive real-time predictions along with meaningful interpretations.

The project demonstrates the practical application of machine learning in the domain of personal health assessment, offering users a data-driven means of understanding their lifestyle quality. The deployment-ready framework can be extended further to incorporate additional health indicators, making it a robust foundation for more advanced digital wellness tools.

# CONTENTS


| Contents | Page No |
|---|---|
| **ACKNOWLEDGEMENT** | **i** |
| **ABSTRACT** | **ii** |
| Chapter 1 . INTRODUCTION | 1 |
| Chapter 2 . OBJECTIVES | 3 |
| Chapter 3 : METHADOLOGY | 5 |
| Chapter 4 : DATASET DESCRIPTION | 9 |
| Chapter 5 : DATA PRE-PROCESSING | 13 |
| Chapter 6 : MODEL DEVELOPMENT | 17 |
| Chapter 7 : RESULTS AND EVALUATION | 21 |
| Chapter 8 : DEPLOYMENT | 25 |
| Chapter 9 : CONCLUSION | 28 |

# CHAPTER 1

# INTRODUCTION

In recent years, the importance of maintaining a healthy lifestyle has become more prominent due to the rising incidence of lifestyle-related diseases such as obesity, diabetes, cardiovascular disorders, and mental health challenges. As individuals adopt diverse habits related to sleep, diet, physical activity, stress, and screen time, understanding the overall quality of one's lifestyle can be complex. Traditionally, lifestyle assessment requires manual tracking, expert evaluation, or health check-ups — processes that can be time-consuming, subjective, and not easily accessible to everyone.

To address this challenge, machine learning techniques offer a powerful approach for analyzing multiple lifestyle-related factors simultaneously and providing consistent, data-driven insights. By examining patterns in an individual's daily habits, ML models can estimate how healthy their overall lifestyle is and categorize them into meaningful health groups such as Healthy, Moderate, or Unhealthy. This enables personalized feedback, early detection of harmful habits, and actionable recommendations that empower people to improve their well-being.

This project, "Lifestyle Score & Health Type Prediction using Machine Learning," aims to build an automated system that predicts:

1) Lifestyle Score (0–100)

A numerical rating that reflects the overall lifestyle quality of a person based on their activity level, diet, stress, sleep patterns, physical habits, and other health indicators.

2) Lifestyle Category

A classification that groups individuals into:

- Healthy
- Moderate
- Unhealthy

To achieve this, a comprehensive synthetic dataset of 200,000 individuals was generated, containing variables such as sleep hours, water intake, steps per day, workout habits, junk food consumption, stress level, screen time, BMI, and more. The dataset allows the machine learning models to learn realistic relationships between lifestyle behaviors and health outcomes.

Two separate models were developed:

- A Regression Model to predict the Lifestyle Score.
- A Classification Model to determine the Lifestyle Category.

Advanced preprocessing, feature selection, correlation analysis, and model comparisons were performed to identify the most influential factors. The best-performing models were then deployed in an interactive Streamlit application, allowing users to input their personal lifestyle parameters and instantly receive personalized predictions.

Ultimately, this project demonstrates how machine learning can be leveraged to create a practical, accessible, and intelligent health-monitoring tool. It bridges data science with wellness, enabling individuals to understand their habits better and make informed decisions to improve their quality of life.

# CHAPTER 2

# OBJECTIVES

## 2.1 Primary Objective

The primary objective of this project is to develop an intelligent machine learning system capable of assessing an individual's lifestyle quality using measurable behavioral and health-related parameters. The system predicts two key outputs: a numerical Lifestyle Score and a categorical Lifestyle Type. Together, these predictions provide a comprehensive evaluation of a person's overall lifestyle health.

## 2.2 Develop a Lifestyle Scoring Model

A major goal is to build a regression-based model that estimates a Lifestyle Score between 0 and 100. This score summarizes the combined effect of factors such as physical activity, diet quality, sleep duration, stress, water intake, and screen time. The scoring model aims to provide users with a quantitative measure of their lifestyle condition.

## 2.3 Develop a Lifestyle Classification Model

In addition to the numerical score, the project aims to categorize individuals into meaningful lifestyle groups. A classification model is developed to assign users to:

- Healthy
- Moderate
- Unhealthy

This allows users to understand their lifestyle status in an interpretable, real-world context.

## 2.4 Build a Realistic Synthetic Dataset

Another objective is to generate a large and realistic dataset that captures a wide range of human lifestyle patterns. The dataset consists of variables such as exercise habits, dietary characteristics, physiological metrics, and behavioral factors. It is carefully designed to reflect realistic distributions, variability, and correlations to ensure reliable model performance.

## 2.5 Perform Exploratory Data Analysis (EDA)

A detailed exploratory analysis is undertaken to study the relationship between different lifestyle features and the target variables. Through correlation matrices, descriptive statistics, and visualizations, the project identifies the strongest predictors and determines which features are most influential for lifestyle assessment.

## 2.6 Evaluate and Compare Multiple Machine Learning Models

Several machine learning algorithms are trained and compared to determine the best-performing models for both regression and classification. The evaluation includes models such as:

- Linear Regression
- Random Forest
- Support Vector Machines
- K-Nearest Neighbors
- Logistic Regression

## 2.7 Deploy a User-Interactive Prediction System

The final objective is to implement a Streamlit-based application that allows users to provide lifestyle-related inputs and receive real-time predictions. The interface is designed to be simple, accessible, and informative, allowing users to understand their lifestyle quality without any technical expertise.

# CHAPTER 3

# METHADOLOGY

## 3.1 Overview

The methodology for this project follows a complete machine learning pipeline, beginning with data generation and preprocessing, followed by exploratory analysis, model development, model evaluation, and ending with the deployment of the final system. Each stage is designed to ensure accuracy, interpretability, and practical usability. The methodology integrates both regression and classification workflows, as the project builds two predictive models for lifestyle assessment.

## 3.2 Data Generation and Understanding

A synthetic dataset of 200,000 records was generated to simulate real-world lifestyle patterns. The variables were designed to reflect realistic behavior, including normal ranges, plausible correlations, and variability across individuals. The dataset incorporated demographic attributes, physical activity information, dietary habits, stress-related factors, and behavioral metrics.

A thorough understanding of these features was necessary before proceeding with model development. Statistical summaries and visualizations were used to verify that the dataset followed expected patterns and exhibited meaningful relationships between inputs and target variables.

## 3.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to identify trends, correlations, and noteworthy patterns. This involved:

- Generating descriptive statistics
- Analyzing feature distributions
- Computing correlation coefficients

- Visualizing relationships through heatmaps

EDA played a crucial role in determining the features that influence lifestyle outcomes, guiding both regression and classification model design.

## 3.4 Model Design and Selection

Two types of prediction models were developed:

### a) Regression Model

Designed to estimate the continuous **Lifestyle Score**, this model used numerical input features selected based on correlation strength and interpretability. Several regression algorithms were evaluated, including:

- Linear Regression
- Random Forest Regressor
- Support Vector Regressor

Linear Regression produced the best performance for the selected feature set and was chosen as the final regression model.

### b) Classification Model

This model predicts the categorical **Lifestyle Type** (Healthy, Moderate, Unhealthy). Multiple classifiers were trained and compared, including:

- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier
- K-Nearest Neighbors

The model with the highest accuracy and generalization performance was selected as the final classifier.

## 3.5 Training and Validation

To ensure fair evaluation, the dataset was divided into training and testing subsets using an 80:20 split. For the classification task, stratified sampling was used to maintain proportional representation across the three lifestyle classes.

During model training:

- Numerical features were standardized
- Categorical variables were encoded
- Hyperparameters were tested where appropriate
- Predictions were validated using the test set

This ensured both robustness and stability in model performance.

## 3.6 Model Saving and Deployment Preparation

After selecting the best models, both the regression and classification models were saved using **joblib** for future use. Additionally, the trained label encoder used for transforming categorical labels was saved to maintain consistency during prediction.

A prediction pipeline was implemented, enabling the models to receive user input in a structured format and return predictions for both lifestyle score and category.

## 3.7 Application Development

A Streamlit-based application was developed as the final user interface. The application:

- Accepts user lifestyle inputs through interactive sliders.
- Prepares the input according to the trained model requirements

- Performs real-time predictions using the saved models
- Displays results in a clear and interpretable manner

This deployment step ensures that the system is not only technically sound but also practically accessible for end users.

## 3.8 Summary

The overall methodology integrates data simulation, analysis, model development, evaluation, and deployment into a cohesive workflow. The approach ensures that the final lifestyle prediction system is accurate, interpretable, and user-friendly, fulfilling both analytical and practical objectives of the project.

# CHAPTER 4

# DATASET DESCRIPTION

## 4.1 Overview

The dataset used in this project is a synthetically generated collection of 200,000 records representing diverse human lifestyle patterns. It includes a wide range of behavioral, physiological, and wellness-related attributes that together provide a comprehensive view of an individual's lifestyle habits. The synthetic nature of the dataset allows for balanced distributions, controlled correlations, and realistic variability across different population groups.

## 4.2 Feature Categories

The dataset contains a total of 19 features, divided into meaningful categories based on the type of information they represent. These categories are described below.

### a) Demographic Features

These variables represent basic personal characteristics:

- Age
- Gender

### b) Physiological Measurements

Attributes describing physical body metrics:

- Height_cm
- Weight_kg
- BMI

### c) Physical Activity Indicators

Daily movement and exercise-related features:

- Steps_per_day
- Workout_days_per_week
- Workout_intensity
- Activity_score

### d) Lifestyle Behaviour Variables

Habits related to sleep, screen usage, food choices, and water consumption:

- Sleep_hours
- Screen_time_hrs
- Water_intake_L
- Junk_food_per_week
- Diet_score

### e) Health Risk Factors

Variables indicating potentially harmful habits or stress conditions:

- Alcohol
- Smoking
- Stress_level

### f) Target Variables

The project predicts two outputs:

- Lifestyle_Score — continuous numerical value

- Lifestyle_Category — Healthy, Moderate, or Unhealthy

## 4.3 Data Characteristics

The dataset was designed to mimic realistic human behavior. This includes:

- Natural variation in lifestyle patterns (e.g., sleep ranges from 3 to 10 hours)
- Logical relationships (e.g., higher stress tends to reduce lifestyle score)
- Distribution across a broad population sample
- Balanced representation across categories
- Consistent scaling of features to maintain model stability

All values were generated using controlled probabilistic distributions combined with domain logic. For example:

- Stress level increases the probability of having poor diet scores.
- Higher physical activity increases predicted lifestyle score.
- Excessive screen time negatively correlates with health outcomes.

These relationships ensure that the dataset behaves similarly to real-world lifestyle data.

## 4.4 Summary Statistics

The dataset exhibits realistic variation across all features. Typical examples include:

- Age: 18–70 years
- Sleep Hours: 3.0–10.8 hours
- Steps per Day: 2,000–15,000 steps
- Stress Level: 1–10
- Diet Score: Approximately 1.5–8.0

- Lifestyle Score: Wide distribution with mean around 21.5

Such variation ensures that the machine learning models can generalize effectively.

## 4.5 Data Quality

The dataset contains:

- No missing values
- Clean labeling
- Standardized formats for all numerical fields
- Categorical variables properly encoded (e.g., Gender and Lifestyle_Category)

This ensures that preprocessing requirements are minimal and modeling can be performed efficiently.

# CHAPTER 5

# DATA PRE-PROCESSING

## 5.1 Overview

Data preprocessing is a crucial step to ensure that the dataset is clean, consistent, and suitable for machine learning. Although the synthetic dataset was generated in a controlled manner, several preprocessing steps were still required to prepare the data for feature analysis and model development. This section outlines the procedures applied, including data cleaning, encoding, feature selection, and scaling.

## 5.2 Handling Missing and Invalid Values

Since the dataset was generated synthetically with predefined constraints, it contained no missing values or invalid entries. This greatly simplified preprocessing, allowing the focus to shift toward preparing the data for analytical and predictive tasks.
A verification step was performed using:

- `df.isnull().sum()` to confirm zero missing values
- Data type checks to ensure consistent formatting

All features were validated before further processing.

## 5.3 Encoding Categorical Variables

The dataset includes two categorical features: **Gender** and **Lifestyle_Category**.
These variables must be converted into numerical form for machine learning models to process them effectively.

- **Gender** was assigned categorical labels representing Male, Female, and Other.

- **Lifestyle_Category** (Healthy, Moderate, Unhealthy) was encoded using **LabelEncoder**. This encoding was necessary for the classification model, ensuring consistency between training and prediction phases.

The fitted label encoder was saved using `joblib.dump()` for use during model deployment.

## 5.4 Feature Selection

To build efficient and interpretable models, a correlation analysis was performed to identify variables that have a meaningful relationship with the **Lifestyle Score** and **Lifestyle Category**.

Key observations included:

- Strong positive correlations with:
    - Activity Score
    - Sleep Hours
    - Workout Days per Week
    - Diet Score
- Strong negative correlations with:
    - Stress Level
    - BMI
    - Screen Time
    - Junk Food Consumption

Based on these insights, the most influential features were selected for both the regression and classification models. This improved model accuracy and reduced unnecessary complexity.

## 5.5 Feature Scaling

Many machine learning algorithms require features to be on similar scales to prevent bias toward variables with larger numeric ranges. Therefore, **StandardScaler** was applied to the selected numerical features before model training.

Scaling ensured:

- Improved convergence for linear models
- Stable performance for SVM and KNN
- Balanced feature behavior during distance-based calculations

The scaler was fitted on the training data and applied consistently to both training and testing sets.

## 5.6 Train-Test Split

To evaluate model performance fairly, the dataset was divided into training and testing subsets using **train_test_split** with an 80:20 ratio.
This ensured:

- Sufficient data to train both regression and classification models
- Reliable estimation of generalization performance
- Stratification for classification to preserve class balance

This step provided a fair benchmark for comparing different machine learning models.

## 5.7 Summary

The preprocessing pipeline included the following key steps:

- Validation of missing values and data quality
- Encoding of categorical variables
- Correlation-based feature selection
- Standardization of numerical features
- Splitting the dataset into training and testing sets

These stages created a clean, efficient foundation for building robust machine learning models.

# CHAPTER 6

# MODEL DEVELOPMENT

## 6.1 Overview

Model development in this project involves building two separate predictive models:
(1) A regression model to estimate the Lifestyle Score, and
(2) A classification model to determine the Lifestyle Category.
Both models were trained using the preprocessed dataset and evaluated using standard machine learning metrics. This section explains the model design, feature selection, algorithm choice, and implementation details.

## 6.2 Feature Selection for Modeling

Based on correlation analysis and statistical interpretation, a subset of features was selected for modeling. These features demonstrated strong relationships with the Lifestyle Score and Lifestyle Category. Factors such as Activity Score, Sleep Hours, Workout Frequency, Diet Score, Stress Level, Junk Food Consumption, Screen Time, and Water Intake were identified as the most influential.

The selected features ensured:

- Reduced model complexity
- Improved interpretability
- Faster training and prediction
- Better generalization performance

These selected variables were used consistently for both the regression and classification models.

## 6.3 Regression Model Development

The regression task aimed to predict the continuous **Lifestyle Score (0–100)**. Several regression algorithms were tested initially, including Linear Regression, Random Forest Regressor, and Support Vector Regressor. Performance was compared using metrics such as $R^2$ score and Mean Absolute Error.

### Chosen Algorithm: Support Vector Regressor

Support Vector Regressor demonstrated the highest performance and stability with the selected feature set. Its advantages include:

- Strong interpretability
- Low variance
- Suitable for linearly correlated data
- Fast computation

The regression model was trained on the scaled training dataset and evaluated on the test set. The resulting model provided consistent and accurate predictions of lifestyle scores.

## 6.4 Classification Model Development

The classification task involved predicting the **Lifestyle Category**, which has three classes: Healthy, Moderate, and Unhealthy. Several classification algorithms were evaluated, including Logistic Regression, Random Forest Classifier, Support Vector Classifier, and K-Nearest Neighbors.

### Model Comparison

Each classifier was assessed using:

- Accuracy score
- Confusion matrix
- Class distribution performance
- Generalization on the test set

### Chosen Algorithm: Support Vector Classifier

Among the tested algorithms, Support Vector Classifier achieved the highest accuracy and balanced performance across all classes and was selected as the final classification model. This model effectively identifies all three lifestyle categories with high reliability.

## 6.5 Training Process

Both regression and classification models were trained using an 80:20 train-test split. The training process included:

- Standardizing numerical features using StandardScaler

- Encoding categorical variables

- Ensuring consistent preprocessing pipelines

- Monitoring performance on unseen test data

The models showed strong generalization capabilities and performed well across evaluation metrics.

## 6.6 Saving the Trained Models

To enable easy deployment and future predictions, the trained models were saved using **joblib**:

- `score.pkl` – Contains the trained regression model
- `category.pkl` – Contains the trained classification model
- `labelencoder.pkl` – Stores the fitted label encoder for Lifestyle Category

Saving these artifacts ensures reproducibility and enables integration with the Streamlit application.

## 6.7 Deployment-Ready Structure

A prediction pipeline was constructed where user inputs are:

1. Converted to a structured DataFrame
2. Passed through the regression model to obtain the Lifestyle Score
3. Passed through the classification model to obtain the Lifestyle Category
4. Decoded using the LabelEncoder for human-readable output

This complete pipeline is used directly in the Streamlit application for real-time prediction.

## 6.8 Summary

The model development process involved selecting meaningful features, comparing multiple algorithms, choosing the best-performing models, training them effectively, and preparing them for deployment. Both models — the regression model for scoring and the classification model for categorization — demonstrated strong performance and reliability in predicting user lifestyle outcomes.

# CHAPTER 7

# RESULTS AND EVALUATION

## 7.1 Overview

The performance of both the regression and classification models was evaluated using appropriate statistical metrics and diagnostic techniques. This section presents the results obtained from testing the models on unseen data, along with an analysis of their effectiveness, strengths, and limitations. The evaluation ensures that the selected models are both reliable and suitable for real-world lifestyle prediction tasks.

## 7.2 Regression Model Evaluation

The regression model was developed to predict the **Lifestyle Score**, a continuous variable ranging from 0 to 100. Model performance was assessed using standard regression metrics.

### 7.2.1 Evaluation Metrics

The following metrics were used:

- **$R^2$ Score**
  Measures the proportion of variance in the target variable explained by the model.

- **Mean Absolute Error (MAE)**
  Represents the average magnitude of prediction errors.

- **Root Mean Squared Error (RMSE)**
  Provides a penalty for larger errors and reflects prediction precision.

### 7.2.2 Regression Results

The Linear Regression model achieved an **$R^2$ score of approximately 0.74 on the test set**, indicating that the model explains roughly 74% of the variance in Lifestyle Score.
The error metrics (MAE and RMSE) showed that the predicted scores are reasonably close to actual values, demonstrating strong performance given the complexity and variability of lifestyle behavior.

### 7.2.3 Interpretation

The regression model performs well in predicting overall lifestyle quality. With strong correlations between selected features and the target variable, the model provides meaningful and accurate score predictions suitable for practical use.

## 7.3 Classification Model Evaluation

The classification model was evaluated on its ability to correctly categorize individuals into three lifestyle categories: Healthy, Moderate, and Unhealthy.

### 7.3.1 Evaluation Metrics

- **Accuracy Score**
  Measures the overall percentage of correct predictions.

- **Confusion Matrix**
  Provides insight into class-level performance.

- **Precision, Recall, and F1-Score**
  Used to analyze the model's ability to identify each class accurately.

### 7.3.2 Classification Results

The best-performing classifier achieved **high accuracy on the test set**, demonstrating strong capability in distinguishing between the three lifestyle categories.
The confusion matrix indicated that:

- The **Healthy** and **Unhealthy** categories were classified with high precision.
- The **Moderate** category, due to overlapping lifestyle patterns, showed slightly lower but still acceptable performance.

### 7.3.3 Interpretation

The classification model demonstrates reliable performance, with strong accuracy and balanced class-level results. It effectively differentiates between lifestyle categories and is suitable for real-time user prediction in the final application.

## 7.4 Comparative Analysis of Algorithms

Multiple algorithms were tested during model development. The selected models outperformed the alternatives due to:

- Higher accuracy and lower error rates
- Better stability across varying inputs
- Lower computational complexity
- Stronger generalization on unseen data

This comparative approach ensured the final system uses the most effective models for both prediction tasks.

## 7.5 Deployment Readiness

Both models were validated to ensure:

- Consistent predictions across diverse input scenarios
- Stability in edge cases (extreme values within valid ranges)
- Compatibility with the Streamlit user interface

The models were then saved using `joblib` and integrated into a unified prediction pipeline, confirming full readiness for deployment.

## 7.6 Summary

The evaluation results confirm that:

- The **Regression Model** provides accurate lifestyle score predictions with strong explanatory power.
- The **Classification Model** effectively categorizes individuals with high accuracy and balanced performance.
- The overall system demonstrates strong predictive capability and is suitable for practical application in lifestyle assessment.

# CHAPTER 8

# DEPLOYMENT

## 8.1 Overview

Deployment represents the final stage of the project, where the trained machine learning models are integrated into a functional system that end users can interact with. The objective of deployment is to bridge the gap between model development and real-world usability by providing an accessible and intuitive interface for lifestyle prediction. This project utilizes a Streamlit-based web application to make the predictive system available in real time, without requiring users to possess any technical background.

## 8.2 Model Export and Storage

Once the best-performing regression and classification models were identified, they were saved using the `joblib` library. The following components were exported:

- **score.pkl** – the trained regression model
- **category.pkl** – the trained classification model
- **labelencoder.pkl** – the label encoder used for mapping lifestyle categories
  These files ensure that the prediction pipeline remains consistent between training and deployment phases. Saving the models separately allows them to be loaded efficiently and used without retraining.

## 8.3 Prediction Pipeline

A dedicated prediction pipeline was designed to prepare user inputs and generate outputs using the saved models. The pipeline performs the following steps:

1. Accepts user lifestyle parameters as input
2. Converts the input into a structured pandas DataFrame
3. Applies the same preprocessing format as used during model training
4. Passes the input through the regression model to generate a Lifestyle Score
5. Uses the classification model to predict the Lifestyle Category
6. Decodes the numerical category into a human-readable label
   This pipeline ensures reliability, consistency, and ease of integration within the application.

## 8.4 Streamlit Application Development

A Streamlit-based web interface was created to serve as the deployment platform for the predictive system. The application includes:

- Slider inputs for all user lifestyle variables
- A backend that loads the saved models and label encoder
- Real-time generation of lifestyle score and category
- A clean and organized layout for ease of use

The application allows users to experiment with different lifestyle parameters and instantly visualize how changes affect their predicted lifestyle quality.

## 8.5 User Interaction and Output Presentation

The deployed interface presents the predicted results clearly and concisely. Outputs include:

- A numerical Lifestyle Score between 0 and 100
- A Lifestyle Category (Healthy, Moderate, or Unhealthy)
- A brief interpretation of the results to assist user understanding

This allows users to gain insights into their lifestyle quality and identify potential areas for improvement. The focus is on delivering a simple, practical, and informative experience.

## 8.6 Execution and Accessibility

The application can be launched locally using the command:

```
streamlit run app.py
```

Once executed, the application opens in the user's web browser. While the current deployment operates locally, the system is structured such that it can be easily deployed to cloud platforms such as:

- Streamlit Cloud
- Heroku
- AWS EC2
- Azure Web Apps

This ensures future scalability and accessibility to a wider audience.

## 8.7 Summary

The deployment phase successfully transforms the machine learning models into a user-oriented prediction tool. By using a Streamlit application and a well-structured prediction pipeline, the system provides real-time, interactive lifestyle assessments. This deployment framework makes the model practical, accessible, and ready for real-world demonstration or integration into larger health monitoring systems.

# CHAPTER 9

# CONCLUSION

This project successfully demonstrates the development of a comprehensive machine learning–based system for evaluating lifestyle quality using measurable behavioral and health-related parameters. By combining synthetic data generation, thorough exploratory analysis, feature selection, and model development, the system provides two key outputs: a numerical Lifestyle Score and a categorical Lifestyle Classification. These outputs enable individuals to gain an informed understanding of their lifestyle habits and overall well-being.

The project implemented two predictive models—Linear Regression for scoring and a high-performing classifier for lifestyle categorization. Both models exhibited strong performance on test data, demonstrating their ability to generalize effectively across diverse input scenarios. The regression model achieved a substantial level of explanatory power, while the classification model produced high accuracy in distinguishing among Healthy, Moderate, and Unhealthy categories.

A user-friendly Streamlit application was developed to deploy these models, enabling real-time predictions through an accessible web interface. This deployment step transforms the machine learning workflow into a practical tool that can be used by individuals without technical expertise, reinforcing the project's applicability to everyday health awareness.

Overall, the project highlights the potential of machine learning in the domain of personal health assessment. By integrating data-driven insights with an intuitive interface, the system offers a meaningful way for users to monitor their lifestyle patterns and identify areas for improvement. The methodology used in this project can be extended to incorporate additional health indicators over time, making it a strong foundation for more advanced digital wellness applications.