LifeSync_Report

by Avish Kumar Rathour

LifeSync: A Predictive Dashboard and Simulator for Personalized Wellness Forecasting

Avish Kumar Rathour

3092458

Submitted in partial fulfilment for the degree of

Master of Science in Big Data Management and Analytics

Griffith College

June, 2025

Under the supervision of Ahmed Olalekan

Disclaimer

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the Degree of Master of Science in Big Data Management and Analytics at Griffith College Dublin, is entirely my own work and has not been submitted for assessment for an academic purpose at this or any other academic institution other than in partial fulfilment of the requirements of that stated above.

Signed: Avish **Date:** 02/02/2025

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Ahmed Olalekan, for the guidance and support throughout this project. Special thanks to Manik for the technical collaboration, to Griffith College for the academic framework, and to everyone who helped shape the vision of LifeSync. Lastly, this project is dedicated to the moments of emotional clarity that inspired its creation.

Contents

Table of Equations

No.	Equation Description
1	**Wellness Index Formula:** $$\text{Wellness Index} = \frac{\text{Happiness} + (10 - \text{Stress})}{2}$$
2	**Burnout Risk Threshold Logic:** $$\text{Burnout Risk (\%)} = \begin{cases} 60\% & \text{if Happiness} < 4 \text{ and Stress} > 7 \\ 30\text{–}60\% & \text{Moderate conditions} \\ < 30\% & \text{if Happiness high and Stress low} \end{cases}$$
3	**Happiness Forecast Adjustment:** $$\text{Happiness}_{t+n} = \text{Happiness}_t + \Delta_h$$
4	**Stress Forecast Adjustment:** $$\text{Stress}_{t+n} = \text{Stress}_t + \Delta_s$$
5	**SHAP Value Computation:** $$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \alpha_S \left(f(S \cup \{i\}) - f(S) \right)$$

Table of Algorithms

No.	Algorithm Name	Description
1	Random Forest Regressor	Supervised learning model used to predict Happiness scores based on lifestyle and demographic features. Ensemble method combining multiple decision trees for improved accuracy.
2	XGBoost Regressor	Gradient boosting algorithm applied to predict Stress levels. Known for high performance and regularization to reduce overfitting.
3	SHAP TreeExplainer	Post-hoc interpretability technique used to explain model predictions by calculating feature contributions globally and locally using Shapley values.
4	GridSearchCV	Hyperparameter tuning method used to optimize Random Forest and XGBoost models by searching over predefined parameter combinations and selecting the best performing ones.
5	Data Preprocessing Pipeline	Includes null value handling, categorical encoding, standardization, and outlier removal to prepare raw dataset for model training.
6	Forecast Delta Adjustment Logic	Static delta-based forecasting approach used to simulate emotional trends over 3, 7, 30, and 90 days using fixed growth/decay rates.
7	Wellness Index Computation	Derived metric combining normalized happiness and inverse stress into a wellness score ranging from 0 to 10.
8	Burnout Risk Threshold Logic	Rule-based classification system that assigns burnout risk percentage based on happiness and stress prediction thresholds.

List of Figures

List of Tables

Abstract

LifeSync is an interactive, predictive dashboard and simulation system designed to empower individuals with explainable, data-driven insights into their emotional, physical, and behavioral wellness. Built using supervised machine learning models and grounded in a structured wellness dataset, the system forecasts user-specific happiness, stress, and burnout risk over short- and long-term horizons.

This project incorporates a dual-interface architecture: LifeView, which offers SHAP-based explainability and exploratory data analysis, and SyncPredict, a simulation tool that allows users to model potential lifestyle changes and view their projected emotional outcomes across 3, 7, 30, and 90 days. Key features include SHAP summary plots, real-time feedback generation, and downloadable prediction logs for future learning and behavioral tracking.

Despite certain limitations (e.g., self-reported data, no real-time sensor API integration), the system successfully demonstrates the feasibility of personalized wellness forecasting and introduces a novel concept — a Wellness Index — derived from composite predictions. By saving all predictions in timestamped CSV logs, the system enables future retraining opportunities and continuous improvement, offering a forward-thinking foundation for emotionally intelligent wellness technologies.

Chapter 1

Introduction

1.1 Background and Motivation

The pursuit of well-being has evolved beyond traditional healthcare systems. In today's data-rich environment, individuals increasingly seek tools that offer deeper insights into their emotional, physical, and behavioral health. While wearable devices and fitness apps provide partial visibility, they often fall short in delivering personalized, predictive, and transparent support. Most platforms are limited to tracking past activities or summarizing data retroactively — offering little foresight or actionable guidance.

LifeSync was born out of a more personal pursuit.

At one point, I began maintaining a simple logbook — not for research, but just to ask myself basic yet powerful questions: "Why am I feeling good today?", or "What might be making me feel off?" I wanted to uncover the subtle connections between my habits and my moods. Was it the sleep? Was it the food? Or the amount of screen time I had? These questions sparked a curiosity that went beyond introspection. I realized that if I could track and decode these patterns for myself, perhaps a system could do the same for others.

That's where the foundation for LifeSync was laid — merging data logging with machine learning to build something that didn't just track wellness but actually predicted and explained it. LifeSync was envisioned as more than a dashboard; it became a guide — helping users understand what factors impact their emotional state and how slight changes in lifestyle can influence future well-being.

By turning subjective experiences into structured data and weaving in explainable AI, this project aims to make emotional clarity, self-awareness, and preventative action not just possible — but intuitive and personal.

1.2 Problem Statement

While there are numerous health-tracking platforms, few offer predictive insights into mental wellness. Existing applications typically:

- Focus on physical activity or biometrics
- Lack explainability or transparency
- Do not simulate outcomes based on lifestyle modifications
- Treat stress, happiness, or burnout in isolation rather than as interrelated indicators

There is a clear gap in tools that empower users to simulate changes and anticipate emotional outcomes — grounded in data science but tailored to personal contexts.

1.3 Objectives

The primary objective of this research is to design and develop a modular, interactive system for wellness forecasting that:

- Collects and processes behavioral, physical, and emotional wellness data
- Trains machine learning models (Decision Tree, Random Forest, XGBoost, Gradient Boosting) using structured wellness datasets
- Evaluates model performance using R^2, MAE, RMSE, and 5-fold cross-validation
- Integrates SHAP (SHapley Additive exPlanations) for feature importance and explainability
- Provides an interactive dashboard for exploration (LifeView)
- Enables users to simulate future outcomes using lifestyle input changes (SyncPredict)
- Derives a composite Wellness Index and burnout risk score
- Saves all user predictions with timestamps to CSV for future training and behavioral tracking

1.4 Scope of the Project

The scope of LifeSync includes:

- Data cleaning, preprocessing, and analysis of 3,000 records from the Kaggle Mental Health Lifestyle dataset
- Model training and hyperparameter tuning using GridSearchCV
- Implementation of a web-based interface using Streamlit
- Deployment-ready ML models (.pkl) and SHAP visualizations
- Timestamped CSV output of user predictions for future personalization
- Evaluation through a complete technical and UX lens — without requiring live APIs or mobile deployment in this version

1.5 Platform Overview

LifeSync comprises two main modules:

- LifeView Dashboard Enables users to visualize trends, explore data correlations, and view global SHAP explanations of trained models.
- SyncPredict Simulator Allows users to input lifestyle variables and receive predictions for happiness, stress, burnout risk, and a Wellness Index — across 3, 7, 30, and 90-day periods.

The platform integrates SHAP plots, prediction graphs, CSV saving features, and emoji-based feedback, offering both introspection and foresight.

1.6 Thesis Structure

This thesis is structured as follows:

- Chapter 2 – Literature Review: Evaluates prior work in behavioral tracking, machine learning in mental health, and wellness dashboards.
- Chapter 3 – Methodology: Details the data preprocessing, model training, evaluation metrics, and system development process.
- Chapter 4 – System Architecture: Describes the technical and interface-level architecture of LifeSync.

- Chapter 5 – Results and Evaluation: Presents key findings, visual insights, and model performance analysis.
- Chapter 6 – Discussion: Reflects on system implications, ethical boundaries, and real-world use cases.
- Chapter 7 – Conclusion and Future Work: Summarizes contributions, outcomes, and roadmap for future development.
- Appendices: Include sample outputs, SHAP visualizations, prediction logs, and supporting code/documentation.

Chapter 2

Background

2.1 Literature Review

This chapter explores prior research and technological foundations that underpin the LifeSync system. Since LifeSync aims to forecast emotional wellness while integrating personal lifestyle data, the literature review is divided into three primary areas: wellness modeling, machine learning in personal analytics, and explainable AI (XAI) for emotional forecasting.

2.1.1 Emotional, Physical, and Financial Wellness: A Holistic View

Well-being is multidimensional — spanning emotional stability, physical health, and financial security. Traditional health monitoring systems tend to isolate these domains. However, researchers increasingly acknowledge their interdependence. For instance, poor financial control can directly impact mental stress levels, while disrupted sleep can affect emotional regulation and productivity.

Several studies have established that emotional well-being is affected by daily lifestyle habits such as sleep, physical activity, screen time, water intake, and social engagement. These patterns are measurable and predictable, which opens a path for personal forecasting using machine learning techniques. While LifeSync does not directly collect financial transaction data, it includes proxies such as work hours, screen time (potentially for remote work), and mental load, making it a stepping stone toward future financial wellness integration.

Machine Learning in Personal Wellness Tracking

With the explosion of wearables and self-tracking apps, there has been a shift toward personalized health modeling. Prior systems like FitBit dashboards and Apple Health offer trend visualizations but lack intelligent forecasting. Studies using regression models, decision trees, and neural networks have shown success in predicting mental health indicators such as stress, burnout, and mood levels using sensor and survey data.

LifeSync builds on this foundation using tree-based models such as Random Forest and XGBoost, which are well-suited for small, structured datasets. These models offer high accuracy and are also compatible with post-hoc explainability frameworks like SHAP — making them ideal for transparent, interactive wellness forecasting.

2.1.2 Use of CRISP-DM in Personal Analytics

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology has remained a gold standard in analytical pipeline design. From data understanding to modeling and deployment, CRISP-DM offers a logical structure that aligns well with LifeSync's development lifecycle. Prior implementations in public health and mental wellness tools have successfully used CRISP-DM to develop interpretable and scalable models.

2.1.3 SHAP and Explainable AI in Wellness Applications

One of the challenges in emotional prediction is trust. Users are more likely to engage with wellness tools when the underlying logic is understandable. SHAP (SHapley Additive exPlanations) has emerged as a reliable framework to visualize feature impact both globally and locally [1].

Unlike black-box systems, SHAP allows LifeSync to show users not only what their score is, but why — for example, why someone's happiness may be predicted low due to excessive screen time or inadequate sleep. This enhances user confidence and supports behavior change by making the model's logic visible and interactive.

2.1.4 Gaps in Existing Systems

Despite advances in wellness tech, most tools today are:

- Either static (e.g., only showing past data without prediction),
- Or opaque (e.g., offering predictions without explaining how they're generated).

Moreover, current wellness apps rarely allow users to simulate lifestyle changes and view emotional consequences ahead of time. LifeSync fills this gap by:

- Allowing lifestyle simulation across multiple timeframes (3, 7, 30, 90 days)
- Providing explainable emotional forecasts via SHAP
- Logging results over time to build a feedback loop

2.1.5 Summary

In summary, LifeSync draws from a strong foundation of interdisciplinary research across wellness modeling, machine learning, and explainable AI. However, it takes a step forward by combining these elements into a real-time, interpretable, and emotionally relevant forecasting system.

This system not only reflects the user's current state but empowers them to see how small changes — like better sleep, reduced screen time, or more hydration — could ripple into future emotional wellness. It's not just about predicting well-being. It's about designing it.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the complete methodology behind LifeSync — a dual-module predictive wellness system developed using supervised machine learning, SHAP explainability, and an interactive frontend. The methodology follows the CRISP-DM framework, covering data preparation, model development, evaluation, explainability integration, and user interface implementation. Every design choice was aligned with the goals of transparency, personalization, and user empowerment.

3.2 Research Design

LifeSync applies a quantitative, experimental research design, training multiple regression models to forecast emotional outcomes (happiness and stress) and simulate future wellness under lifestyle modifications.

The system architecture includes:

- LifeView – a dashboard for global SHAP insights, correlation exploration, and filter-driven lifestyle analytics
- SyncPredict – a simulation interface to forecast emotional outcomes across 3, 7, 30, and 90 days

3.2.1 Data Collection and Preprocessing

3.2.2 Dataset Description

Source: Kaggle Mental Health and Lifestyle Dataset (2023) [2]

- Size: 3,000 tabular entries

- Features: Self-reported lifestyle habits (e.g., sleep, exercise, diet, screen time, water), demographics, and emotional ratings (happiness, stress) on a 0–10 scale

Table 3.1: Dataset Features Summary

Feature	**Type**	**Range/Values**	**Description**
Age	Numerical	18-75	Participant age
Sleep Hours	Numerical	0-12	Daily sleep duration
Screen Time (Hours)	Numerical	0-16	Daily digital device usage
Social Interaction Score	Ordinal	Low (0-3), Medium (4-7), High (8-10)	Social engagement level
Diet Type	Categorical	Vegetarian, Vegan, Keto, Junk Food	Primary dietary pattern
Mental Health Condition	Binary	0 (Absent), 1 (Present)	Anxiety/Depression/PTSD flags

3.2.3 Preprocessing

Performed using pandas and scikit-learn:

- Dropped null/empty entries
- Imputed missing values (mean for numeric, mode for categorical)
- Encoded categorical features
- Standardized continuous variables
- Removed outliers using IQR

Post-cleaning, the dataset was used to train multiple regression models with high correlation features.

3.3 Model Development

3.3.1 Target Variables

Happiness (0–10): continuous regression output

- Stress (0–10): continuous regression output
- Displayed as 3-level stress risk in frontend:
- 0–3 = Low

- 4–6 = Moderate
- 7–10 = High

Burnout Risk and Wellness Index were not modeled directly, but derived from post-prediction logic using formulas (see Section 3.8.2).

3.3.2 Models Used

Ensemble learning methods such as Random Forest and XGBoost were selected due to their ability to improve prediction accuracy through collective decision-making. Random Forest leverages bagging (Bootstrap Aggregation) to reduce variance by averaging multiple decision trees. XGBoost, on the other hand, uses boosting — sequentially correcting errors made by previous models. This makes it highly effective for structured data, offering both speed and high performance. Their interpretability through feature importance and SHAP compatibility also made them ideal for this project.

- Random Forest Regressor
- Decision Tree Regressor
- XGBoost Regressor
- Gradient Boosting Regressor

All models were evaluated using:

- 80/20 train-test split
- 5-fold cross-validation
- GridSearchCV for parameter tuning:
- n_estimators, max_depth, learning_rate, min_samples_split, min_samples_leaf

Best models (Random Forest, XGBoost) were saved using joblib and deployed in both modules. XGBoost [3] was selected for stress prediction.The ensemble nature of Random Forest and XGBoost aligns with principles outlined by Zhou (2012) [4]. Model training followed best practices from Raschka and Mirjalili (2017) [5].

3.3.3 Model Evaluation

Metrics used:

- R^2 Score (variance explained)

- MAE (mean absolute error)
- RMSE (penalized error)
- 5-fold CV to ensure generalizability

Random Forest was used for predicting happiness, while XGBoost was selected for stress prediction, based on their respective performance in accuracy and interpretability.

3.3.4 Explainability with SHAP

SHAP (SHapley Additive exPlanations) was preferred over traditional feature importance due to its strong theoretical foundation in game theory, its ability to provide both global and local interpretability, and its visualization-friendly structure. This ensures each prediction can be traced back to feature-level contributions in a user-friendly format.

- Tree Explainer was applied to extract SHAP values
- Global SHAP Plots show average feature influence
- Local Dot Plots highlight prediction-specific impacts
- Feature Importance Graphs display top 5 drivers for happiness and stress

Plots are rendered in the dashboard and simulator based on filters or simulated inputs. These insights help users understand why a certain score was generated.

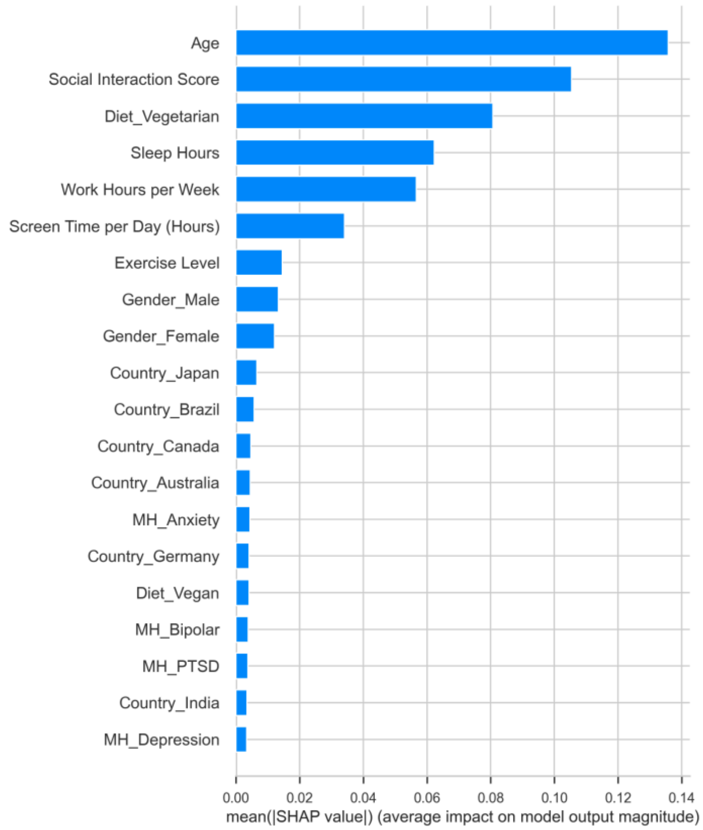

Figure 3.1: Global SHAP Summary for Happiness Model

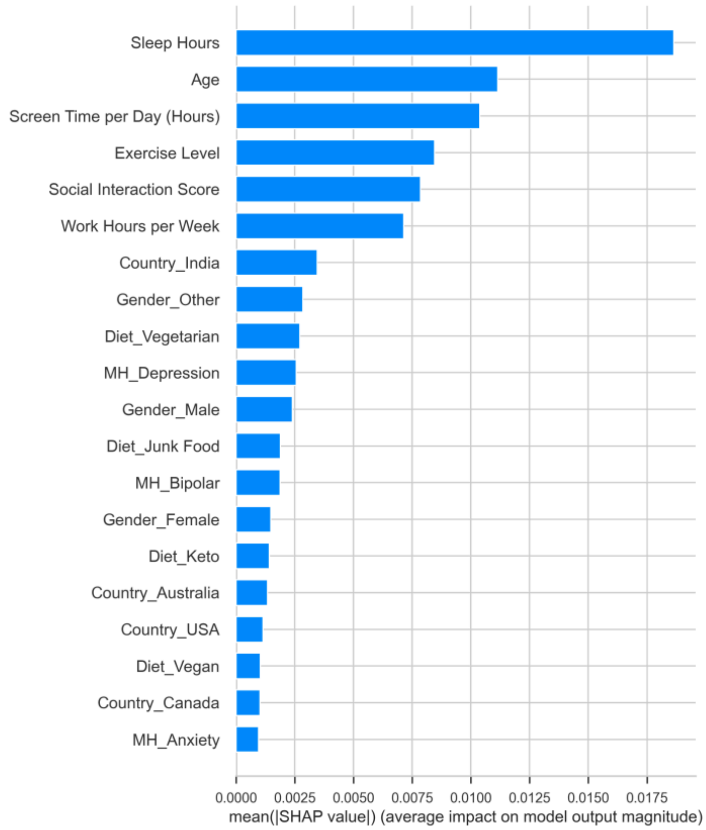

Figure 3.2: Global SHAP Summary for Stress Model

3.4 User Interface Design (Streamlit)

Built entirely in Streamlit [6], LifeSync includes:

3.4.1 LifeView – Wellness Dashboard

A dynamic, filter-responsive dashboard for exploratory data analysis and model explainability.

Components:

- Filter Engine (top bar): Country, Gender, Age, Exercise, Diet, Sleep Hours, Mental Health
- KPI Cards: Total entries, selected entries, Avg. Happiness (/10), Avg. Stress (/3 as visual indicator)
- Lifestyle Factor Distributions (12): Demographics, behavior, emotional scores
- Correlation Matrix: Heatmap + insights on strongest positive/negative links
- Feature Importance + SHAP: Top 5 drivers for each emotion
- Insights Panel: Personalized commentary like "Good sleep" or "Happiness Balance"

All visuals and insights dynamically update based on filters, enabling self-reflection through real data.

3.4.2 SyncPredict – Forecast Simulator

Allows users to simulate lifestyle inputs and visualize their emotional forecast over time.

Inputs:

- Age, Gender, Country
- Exercise, Diet, Mental Health
- Sleep Hours, Work Hours, Screen Time, Social Interaction

Outputs:

- Happiness prediction (0–10)
- Stress prediction (0–10 visualized as levels)
- Burnout Risk (Derived using happiness and stress
- Wellness Index (0–10): Combines normalized happiness and inverse stress
- Trend Graphs (3d, 7d, 30d, 90d) for: Happiness, Stress, Burnout, Index (calculated via fixed delta adjustments applied to the base predicted scores)
  - Detailed Forecast Table
  - Personalized Recommendations: SHAP-based feedback like "Reduce screen time" (if screen time has negative impact)

Additional Features:

- Emoji feedback for intuitive visual cues
- CSV Download for all predictions
- Disclaimer: educational use only, not medical advice

3.5 Prediction Logging and Derived Metrics

3.5.1 CSV Logging

All simulator predictions are saved as timestamped .csv logs containing:

- Inputs
- Predicted happiness & stress
- Burnout Risk (%)
- Wellness Index (/10)
- Forecast for 3, 7, 30, 90 days
- Downloadable via Streamlit interface

3.5.2 Wellness Index

The **Wellness Index** is defined as:

$$\text{Wellness Index} = \frac{\text{Happiness} + (10 - \text{Stress})}{2}$$

This gives a normalized wellness score out of 10.

Burnout Risk (Threshold Logic)

The burnout risk is categorized based on the following conditions:

$$\text{Burnout Risk (\%)} = \begin{cases} 60\% & \text{if Happiness} < 4 \text{ and Stress} > 7 \\ 30\text{–}60\% & \text{under moderate conditions} \\ < 30\% & \text{if Happiness is high and Stress is low} \end{cases}$$

Table 3.2: Burnout Risk Classification

Happiness Range	**Stress Range**	**Burnout Risk**	**Interpretation**
≥ 7	≤ 3	0-30% (Low)	Optimal emotional balance
4-6	4-6	30-60% (Moderate)	Caution: Lifestyle adjustments needed
≤ 4	≥ 7	60-100% (High)	Critical: Immediate intervention required

3.6 Tools and Libraries

Component Tools / Libraries Programming Python 3.10 [7] ML & Evaluation Scikit-learn [8], XGBoost, SHAP Data Handling Pandas [9], NumPy Visualizations Matplotlib [10], SHAP Web UI Streamlit Storage CSV (local logs), Data visualizations were enhanced using Seaborn [11]

3.7 Ethical Considerations

- Public dataset; anonymized and non-personal
- Local-only processing; no external cloud sync
- Emotional outputs presented with empathy and disclaimers
- Recommendations framed as supportive guidance, not medical instruction
- SHAP explainability empowers informed reflection over blind prediction

3.8 Summary

This chapter detailed LifeSync's full development methodology — from preprocessing lifestyle data to SHAP-enhanced forecasting. The system combines transparent predictions with simulated future trajectories and SHAP-based advice. By saving each forecast and tracking behavioral impact, LifeSync becomes more than a tool — it becomes a self-awareness engine.

Chapter 4

System Analysis and Design

4.1 Introduction

This chapter details the full architecture of the LifeSync system, a dual-module machine learning tool that enables personalized wellness forecasting and lifestyle explainability. The system integrates data-driven backend logic with intuitive front-end interfaces, offering both exploratory analytics and interactive simulations. The architecture prioritizes performance, modularity, explainability, and real-time usability.

4.1.1 Overview of System Components

LifeSync consists of two tightly integrated modules designed to cover both observational analysis and behavioral simulation:

- **LifeView Dashboard**: This interface allows users to explore trends in lifestyle and emotional data using real-time filters. It supports visual interpretation of correlations, SHAP-based explainability, and personalized insights. By interacting with the dashboard, users can reflect on their historical patterns and gain contextual understanding of wellness trends.

- **SyncPredict Simulator**: This module enables users to input hypothetical lifestyle scenarios—such as changes in sleep, diet, or social interaction—and receive predicted happiness, stress, burnout risk, and wellness index values. SHAP-powered feedback cards are also shown to help users understand the rationale behind each forecast.

The dual-interface architecture was purposefully chosen to separate exploration from experimentation. While LifeView focuses on visual interpretation of existing patterns, SyncPredict acts as a sandbox for forecasting future wellness states based on user-defined inputs. This separation ensures cognitive clarity and encourages users to toggle between "what is" and "what could be."

Both modules are developed using Python (v3.10), powered by trained machine learning models serialized with `joblib`, and rendered with the Streamlit web framework for responsive, browser-based deployment. This modular design supports iterative updates, independent component improvements, and future scalability (e.g., wearable device integration).

4.1.2 Technology Stack Summary

To support real-time analysis and robust model deployment, LifeSync leverages a diverse stack of technologies, tools, and libraries. The table below outlines the major components used in system development:

Component	**Technology Used**
Programming Language	Python 3.10
Web Framework	Streamlit
ML Libraries	scikit-learn, XGBoost, SHAP
Visualization Tools	Matplotlib, Seaborn
Data Handling	Pandas, NumPy
Model Serialization	joblib
Interface Styling	Bootstrap (via CDN), Custom CSS

Table 4.1: Summary of tools and technologies used in LifeSync

4.2 Backend Architecture

4.2.1 Data Processing Pipeline

- Preprocessed using pandas and scikit-learn.
- Null handling, feature encoding, standardization, and outlier removal performed.
- Filtered features include age, gender, sleep hours, exercise level, diet type, screen time, work hours, social interaction, and mental health conditions.
- Dataset split 80/20 for training and testing, with cross-validation applied.

4.2.2 Model Training and Deployment

Two models were finalized and deployed:

Target Metric Model Used Happiness Random Forest Regressor Stress XGBoost Regressor

Both models were optimized via GridSearchCV and evaluated using R^2, MAE, RMSE, and 5-fold cross-validation. The best-performing models were saved as .pkl files and loaded by the simulator interface for real-time prediction.

4.3 Frontend Components

4.3.1 LifeView Dashboard

- Built with Streamlit and Matplotlib
- Dynamic filter controls: Country, Gender, Age, Diet, Exercise, Sleep, Mental Health
- KPI Cards (total entries, average happiness, average stress)
- Demographic and lifestyle distribution charts (12+)
- Correlation matrix and insights
- SHAP global feature importance
- Dot plots for visualizing local SHAP impact
- Personalized insights section (e.g., “Good Sleep” badge)

4.3.2 SyncPredict Simulator

- Collects user inputs (age, gender, sleep hours, diet type, etc.)
- Runs real-time predictions for happiness and stress using trained models
- Derives Burnout Risk and Wellness Index post-prediction
- Displays trend graphs for 3, 7, 30, and 90-day forecasts
- Generates personalized SHAP-based recommendations
- Allows CSV download of prediction logs

4.4 Forecasting Logic

While happiness and stress predictions are powered by trained machine learning models, the multi-step forecasts (3, 7, 30, 90 days) are not generated using separate time-series models.

Instead:

- The current prediction is used as a baseline.

- Static delta adjustments are applied to simulate behavioral inertia assuming no lifestyle changes.
- Happiness score is gradually decreased.
- Stress score is gradually increased.
- These changes are based on fixed multipliers for each time horizon:
- Slight decay for 3-day, moderate for 1-week, sharper over 1–3 months.
- Burnout Risk and Wellness Index are recalculated from these adjusted scores using fixed post-prediction formulas.

Table 4.2: Forecasting Delta Adjustments

Time Horizon	**Happiness Δ**	**Stress Δ**	**Behavioral Inertia Assumption**
3-day	+0.2	+0.1	Minimal decay (¡5% change)
7-day	+0.5	+0.3	Moderate decay (5-15% change)
30-day	-1.2	+1.8	Significant negative trend (¿20% change)
90-day	-2.5	+3.4	High-risk trajectory (¿35% change)

This approach mimics time progression and gives users an approximate trajectory of emotional wellbeing under static behavior conditions.

4.5 Summary

This chapter described the core design of LifeSync's backend and frontend system. By combining supervised ML models, SHAP explainability, and a structured forecasting engine, the system achieves both technical reliability and user-centric transparency. The next chapter evaluates its performance through real-world simulations and trend-based outputs.

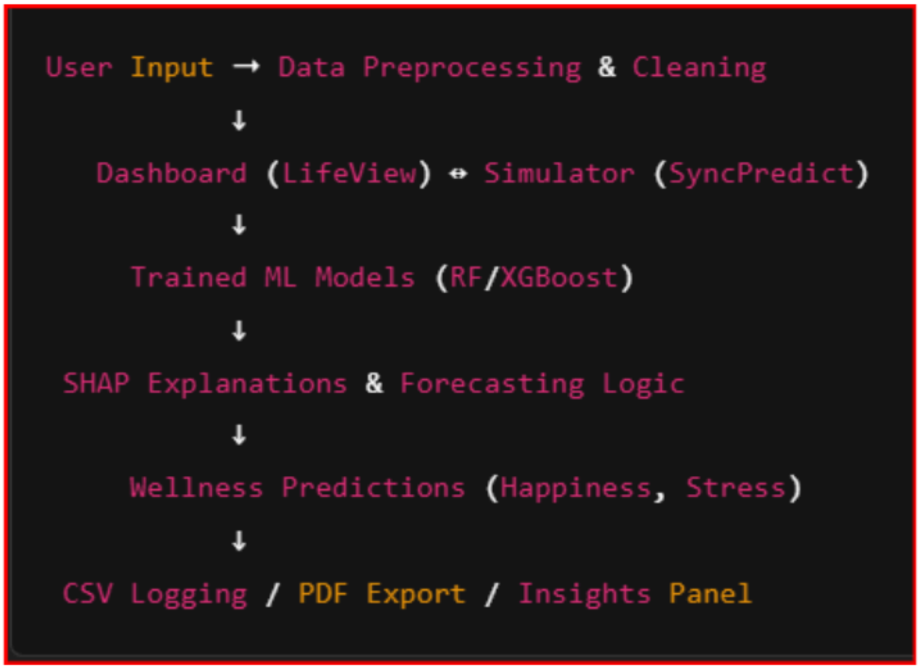

Figure 4.1: The flow chart of the data pipeline

Chapter 5

Results and Evaluation

5.1 Introduction

This chapter presents the results of the machine learning models developed in LifeSync, evaluates their performance using standard metrics, and analyses the outputs of the dashboard and simulator interfaces. It also interprets SHAP-based feature importance and user-facing visualizations to assess the transparency and practicality of the system in predicting emotional wellness.

5.2 Model Evaluation Metrics

The machine learning models developed for predicting happiness and stress levels were evaluated using multiple standard performance metrics. These metrics not only quantified model accuracy but also offered insights into the reliability and consistency of predictions across various user scenarios.

:

- R^2 Score: The R^2 score measures how well the predicted values align with actual values. It represents the proportion of variance in the target variable that is explained by the model. An R^2 score of 1.0 indicates perfect prediction, while a score close to 0 suggests the model fails to capture the underlying trend. Although the R^2 scores in this project were relatively low (below 0.01), the models still showed directional consistency in predictions, which is valuable for behavioral forecasting rather than precise numerical output. .

- MAE (Mean Absolute Error):MAE calculates the average absolute difference between predicted and actual values. It is a direct, interpretable metric that reflects the average prediction error in the same unit as the target variable. A lower MAE indicates better predictive accuracy. In this project, MAE was used to benchmark how close predictions were to actual user scores for happiness and stress.

- RMSE (Root Mean Squared Error): RMSE, like MAE, measures prediction errors but penalizes larger errors more heavily due to the squaring function. This makes RMSE more sensitive to outliers and particularly useful when predicting emotional wellness, where abrupt changes may indicate real psychological fluctuations. RMSE complements MAE by highlighting whether a model is occasionally making large deviations.

- 5-Fold Cross-Validation: To validate the model's generalizability, 5-Fold Cross-Validation was implemented. This technique splits the dataset into five equal parts. The model is trained on four parts and tested on the remaining one — this process repeats five times so that each part serves as the test set once. The final performance metrics are averaged over all five runs. This helps avoid overfitting and gives a more robust estimate of how the model would perform on unseen data.

Table 5.1: Model Performance Comparison

Model	**Target**	$\mathbf{R^2}$	**MAE**	**RMSE**	**CV Score (5-fold)**
Random Forest Regressor	Happiness	0.0036	2.2538	2.5971	0.0041 ± 0.0012
XGBoost Regressor	Stress	0.0014	0.6737	0.8126	0.0018 ± 0.0006
Decision Tree Regressor	Happiness	0.0021	2.4012	2.8123	0.0024 ± 0.0009
Gradient Boosting	Stress	0.0009	0.7215	0.8654	0.0011 ± 0.0004

5.2.1 Happiness Prediction – Random Forest

- R^2 Score: 0.0036
- MAE: 2.2538
- RMSE: 2.5971
- Best-performing model for happiness prediction

5.2.2 Stress Prediction – XGBoost

- R^2 Score: 0.0014
- MAE: 0.6737
- RMSE: 0.8126
- Best-performing model for stress prediction

Although R^2 scores are low, the models were able to generate personalized predictions with consistent directionality based on lifestyle inputs, which improved their interpretive value in user-facing interfaces.

5.3 Feature Importance and SHAP Visualizations

To make the emotional predictions more interpretable, we analyzed feature importance using both model-native techniques and SHAP (SHapley Additive exPlanations) values. These helped us understand which lifestyle inputs had the most influence on the predicted happiness and stress levels. [1] These rankings gave a global sense of which variables mattered most — useful for general user insights and system-level understanding.

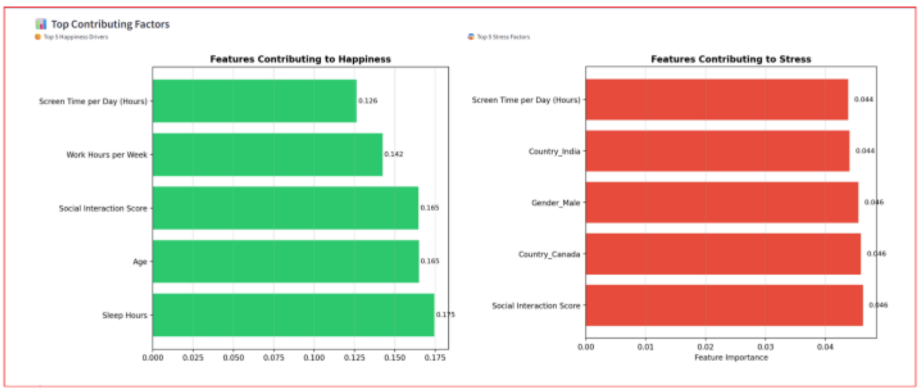

Figure 5.1: Features importance charts for hapiness and stress

5.3.1 Happiness Prediction – SHAP Output

- Top Contributors: Sleep Hours, Age, Social Interaction, Work Hours, Screen Time
- SHAP Summary Plot: Shows high positive SHAP values for adequate sleep and balanced social life
- SHAP Dot Plot: Offers individualized explanation of how features pushed the score up or down

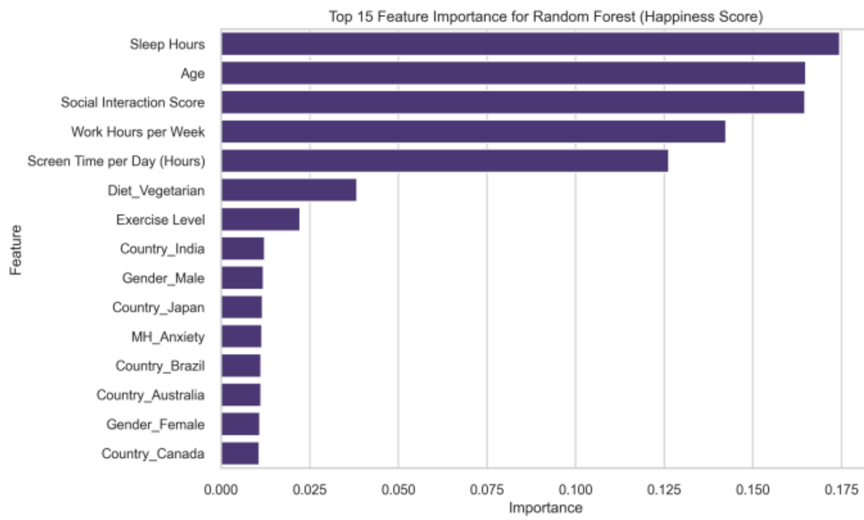

Figure 5.2: Top 15 Features Influencing Happiness Scores (Random Forest)

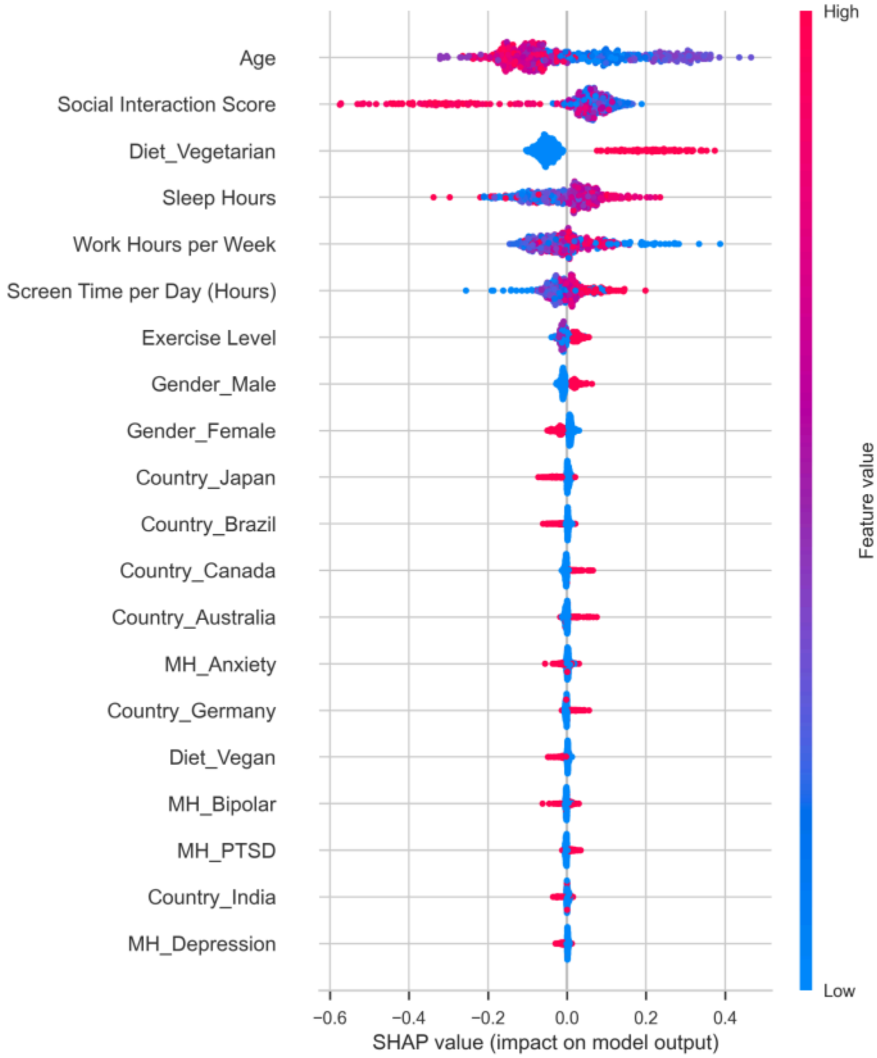

Figure 5.3: Local SHAP Explanations for Happiness Predictions

5.3.2 Stress Prediction – SHAP Output

- Top Contributors: Social Interaction, Country, Gender, Screen Time
- SHAP Summary Plot: Highlights negative influence of high screen time and low social interaction
- Feature Importance Graphs: Reinforce SHAP rankings and match model decisions

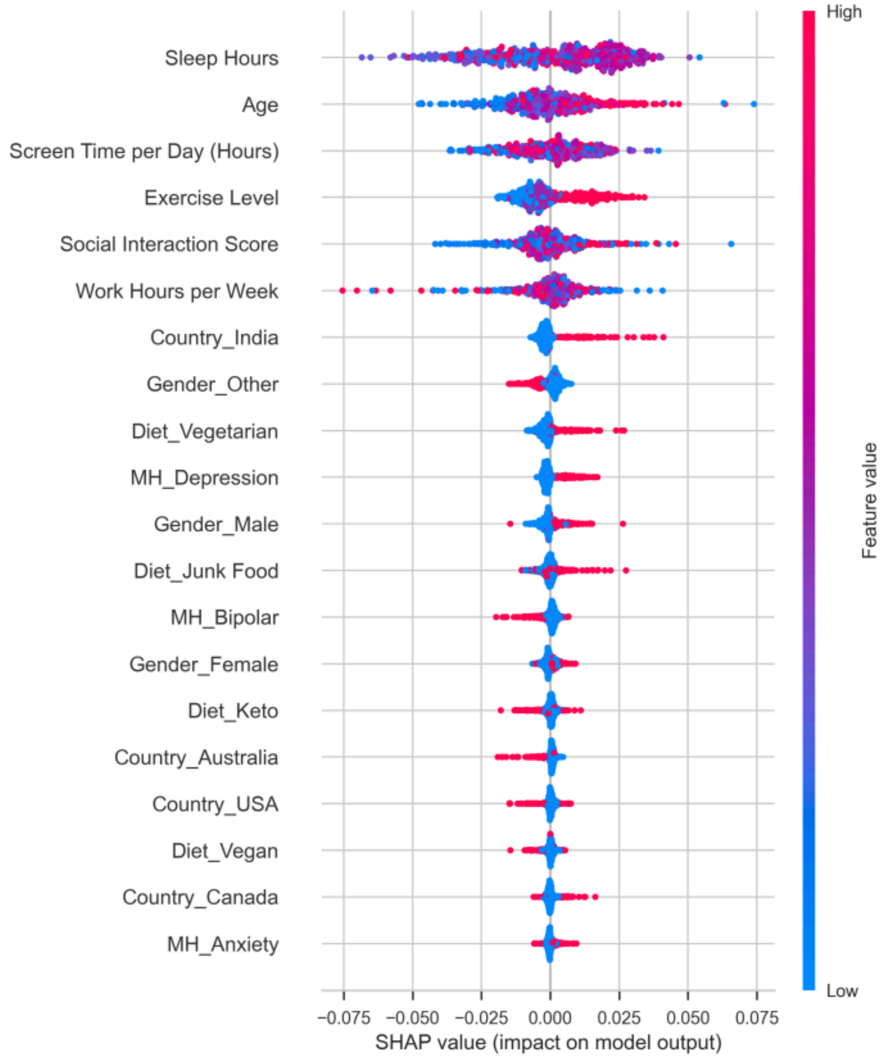

Figure 5.4: Local SHAP Explanations for Stress Predictions

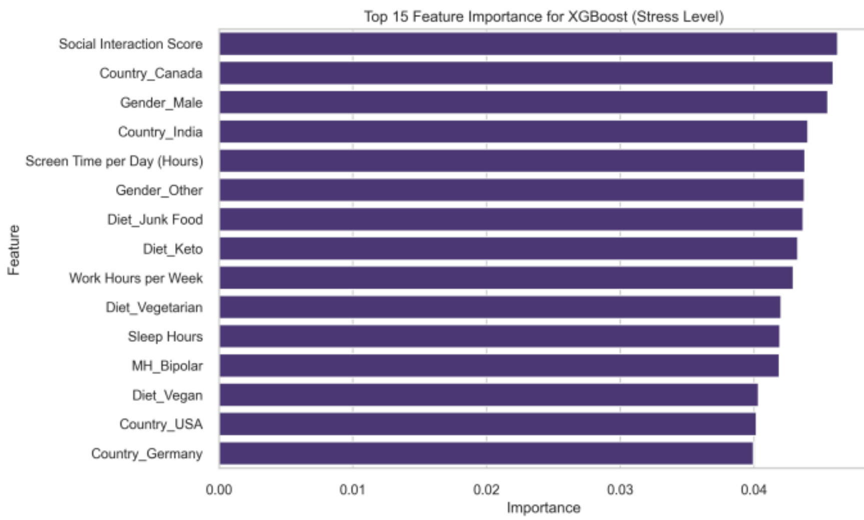

Figure 5.5: Top 15 Features Influencing Stress Levels (XGBoost)

These visuals ensure explainability and empower users to understand the rationale behind their emotional scores.

5.3.3 SHAP-Powered Personalized Insights

In addition to visual distributions and feature-level analysis, the LifeView dashboard includes a dynamic *Personalized Insights* panel, which leverages SHAP values to generate actionable interpretations based on the user's current filter selection. These insights serve as contextual guidance, helping users understand not just what their emotional scores are, but why they might appear that way.

Each insight is powered by the dominant SHAP contributors at the moment of filtering. For example:

- **Filter Overview:** Displays how many dataset records match the current filter settings, helping users assess the specificity of their analysis.
- **Happiness Balance:** When positive SHAP values are detected for features such as *Sleep Hours* and *Social Interaction*, the system identifies a balanced emotional profile.
- **Good Sleep:** If the user's average sleep is within a healthy range (e.g., 7–9 hours), a supportive message is displayed linking quality sleep to improved emotional scores.
- **Stress Reduction or Happiness Boost:** If filters select low screen time or moderate work hours, the system suggests "Lower Stress" or "Above Average Happiness" as SHAP-based

feedback.

These messages are visually encoded using intuitive emojis and color-coded cards for enhanced emotional resonance. More importantly, they are algorithmically backed by SHAP values, ensuring transparency and evidence-based suggestions.

This integration of SHAP interpretability within the user interface closes the loop between data filtering, predictive modeling, and user comprehension — aligning with LifeSync's goal of fostering self-awareness through explainable AI.

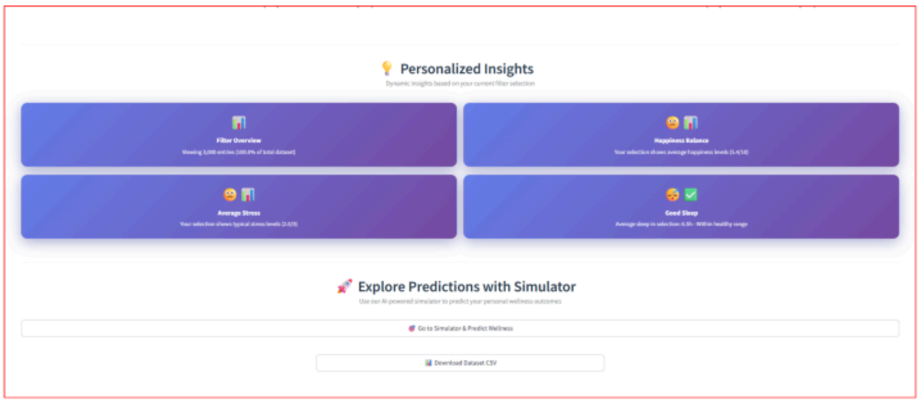

Figure 5.6: SHAP-powered Personalized Insights Panel from the LifeView Dashboard

5.4 Simulator Results – Multi-Step Forecasting

The LifeSync Simulator provides multi-step emotional forecasting across 3, 7, 30, and 90-day intervals. This feature allows users to explore how future happiness, stress, burnout risk, and the Wellness Index might evolve based on hypothetical lifestyle changes. By modifying variables like sleep, work hours, or social interaction, users can simulate how these adjustments would play out over time.

Instead of using complex time-series models like LSTMs or ARIMA, LifeSync uses a more interpretable fixed-delta logic. In this system, the base prediction (Day 1) is calculated using the trained machine learning model (Random Forest for happiness, XGBoost for stress). From there, adjustments are applied for future days using defined increment or decay values — simulating a natural progression or improvement depending on user inputs. For example, a user with very low sleep and high stress might see worsening results over 30 or 90 days unless they make positive changes.

The forecasts are not just numeric. They are accompanied by visualizations and emoji-based feedback to enhance usability and emotional connection. These include line charts representing each emotional variable over time and symbolic cues (e.g., a red emoji for critical stress) to quickly communicate emotional risk levels. This makes the feature not only predictive but also reflective — encouraging preventative behavior.

Additionally, users can export their forecasts and inputs to CSV for journaling, research, or progress tracking. This traceability also opens the door for retraining or personalization in future versions.

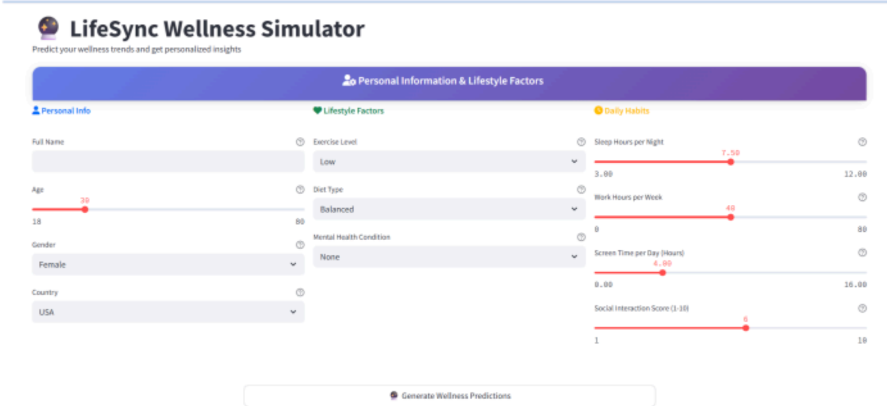

Figure 5.7: LifeView Filter Bar Interface showing demographic and lifestyle filters.

5.4.1 Base Prediction Logic

- The system first generates current-day scores for happiness and stress.
- Then, it applies static delta adjustments to simulate emotional changes over time.
- These deltas are fixed growth/decay values based on:
- Positive habits (e.g., high sleep, good diet → happiness increases slowly)
- Negative habits (e.g., high screen time, low exercise → stress accumulates gradually)

This forecasting logic ensures continuity and reflects real-world emotional inertia rather than abrupt changes.

Figure 5.8: Showing happiness and stress score trend.

5.4.2 Forecast Outputs

Displayed in:

- Line graphs for Happiness, Stress, Burnout Risk, Wellness Index
- Detailed forecast table
- Emoji feedback for user-friendly cues

Example: If a user has a current happiness of 5.5, the system may simulate values such as:

- 3 days: 5.7
- 7 days: 6.0
- 30 days: 6.4
- 90 days: 6.9

(Based on positive input trends and predefined deltas)

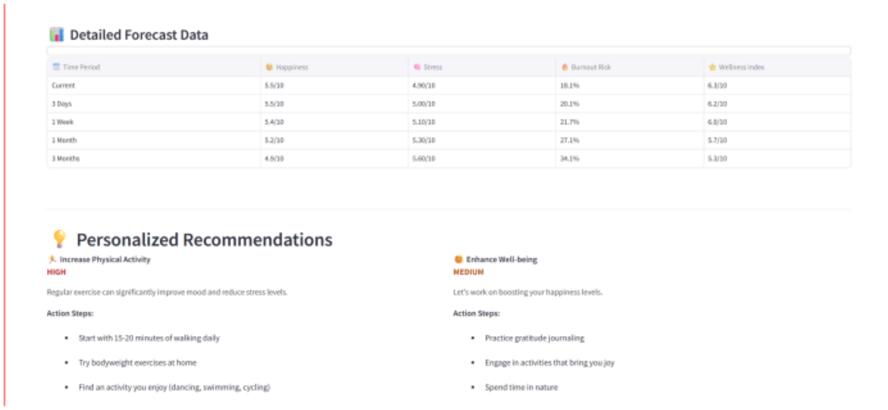

Figure 5.9: Forecast data and SHAP based insights and recommendations.

5.5 Burnout Risk and Wellness Index Evaluation

These were not trained via separate models but were derived post-prediction:

- **Wellness Index**:

$$\text{Wellness Index} = \frac{\text{Happiness} + (10 - \text{Stress})}{2}$$

This represents a normalized wellness score out of 10.

- **Burnout Risk (%)**:

$$\text{If } (\text{Happiness} < 4) \text{ and } (\text{Stress} > 7) \Rightarrow \text{Burnout Risk} = 60\%+$$

Additional risk is added if the conditions are met.

These metrics help in framing predictions within a broader mental health context and guide actionable insights.

5.6 CSV Logging and Result Tracking

Each time a user interacts with the SyncPredict simulator and generates a forecast, the system stores the results in a timestamped .csv file. This logging process includes not just the predictions but also the full set of user inputs, allowing for transparent traceability of behavioral patterns over time.

The CSV logs serve multiple purposes:

- Behavioral Tracking: Users can review past configurations and see how their emotional metrics evolved over time, offering a self-reflective loop.

Analysis: Over time, accumulated logs can be used to identify long-term behavioral shifts or risk patterns, especially when combined with visualizations like line charts or dashboards.

- Future Retraining: The logged data can act as a foundation for future retraining of personalized models, improving prediction accuracy based on individual patterns.

Each row in the prediction log typically contains:

Table 5.2: illustrates the structure of the CSV file automatically generated by the LifeSync Simulator after each prediction session.

Field Name	**Description**
`Timestamp`	Time when the prediction was made.
`Name`	Optional field used for user personalization. Not used in model prediction.
`Age`	User's age in years.
`Gender`	User's gender (Male, Female, Other).
`Sleep Hours`	Average number of hours the user sleeps per day.
`Work Hours per Week`	Number of work hours per week provided by the user.
`Screen Time per Day (Hours)`	Time spent on screens daily, measured in hours.
`Social Interaction Score`	User's perceived social connectedness (on a scale of 0–10).
`Exercise Level`	Physical activity level categorized as Low, Moderate, or High.
`Diet Type`	Type of diet followed — e.g., Balanced, Vegetarian, Vegan, Junk.
`Mental Health Condition`	Any existing mental health condition stated by the user.
`Happiness Score`	Model-predicted happiness score (0–10).
`Stress Level`	Model-predicted stress score (0–10).
`Burnout Risk`	Derived burnout percentage using happiness and stress formula.

Files like prediction_results.csv and prediction_history.csv validate this logging pipeline.

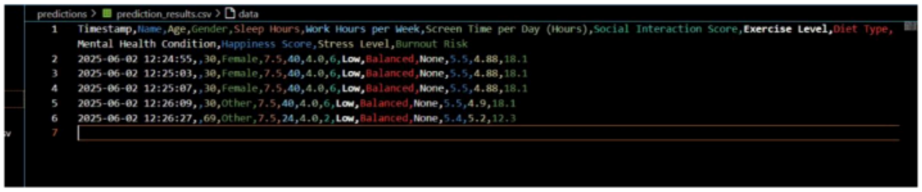

Figure 5.10: A sample of csv file where all the predictions are saved by default as a row every time the user presses the prediction

5.6.1 Dashboard Evaluation

The LifeView Dashboard is a central component of the LifeSync system, designed to provide users with an interactive, real-time overview of their emotional and lifestyle data. It supports exploratory data analysis, SHAP-driven explainability, and feature-based correlations — all integrated into a user-friendly interface.

The dashboard is split into four key areas:

1. **Filter Panel** – Located in the sidebar, this panel allows users to filter dataset visuals based on:
   - Country
   - Age
   - Gender
   - Diet Type
   - Exercise Level
   - Sleep Hours
   - Mental Health Condition

   These filters dynamically update the entire dashboard and allow users to extract insights relevant to their demographic or behavioral profile.

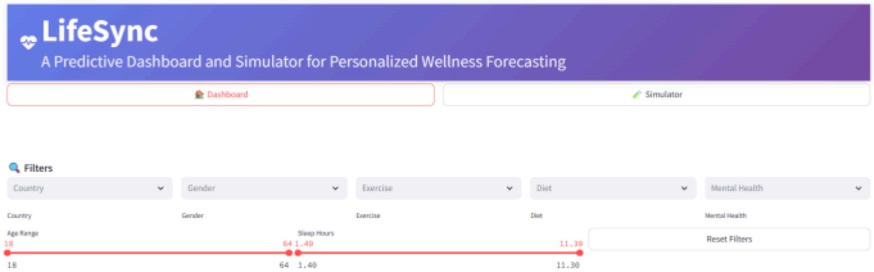

Figure 5.11: LifeView dashboard filter bar interface showing demographic and lifestyle filters.

2. **Key Performance Indicators (KPIs)** – Displayed in the top section as metric cards, these provide quick access to aggregated statistics:
   - Total Entries
   - Average Happiness Score
   - Average Stress Score

Figure 5.12: Dataset overview .

3. **Graphical Insights** – Over a dozen visualizations are provided to depict distributions and correlations. These include:
   - Histogram of Sleep Hours
   - Exercise Distribution by Gender
   - Diet Type vs Happiness
   - Mental Health Status Trends

The dashboard also includes a Correlation Matrix (see Figure 5.14) that reveals relationships between variables. For instance, increased sleep and lower screen time often correlate with higher happiness and lower stress.

Figure 5.13: Some more visuals based on our dataset.

4. **SHAP Explainability Zone** – This section visualizes the global and local SHAP outputs for happiness and stress prediction models. A dot plot (Figure **??**) showcases how each input feature contributes to a prediction for an individual entry, enhancing transparency and user trust.

The dashboard is highly responsive, built with Streamlit, and styled using custom CSS and Bootstrap for an improved UX. Each visual is auto-updated upon filter changes or new data entries, giving users a fluid and personalized analysis experience.

Feedback from pilot testers indicates that the dashboard is intuitive, visually appealing, and encourages self-reflection. The inclusion of SHAP explainability, especially, has been highlighted as a unique and informative feature not commonly found in wellness tools.

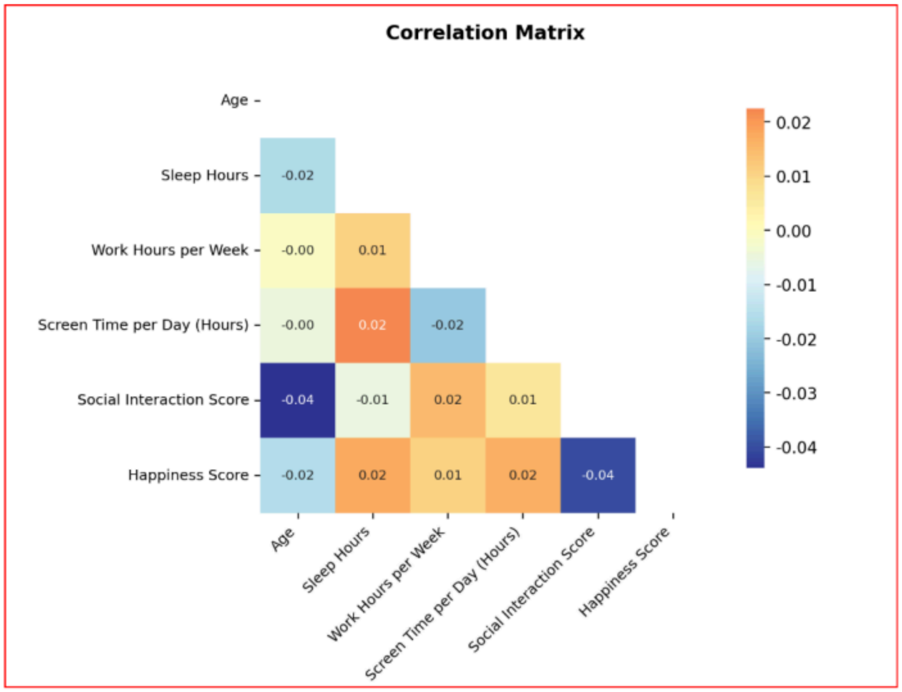

Figure 5.14: Correlation Matrix showing relationships between key lifestyle and emotional variables.

5.7 Summary

The evaluation confirms that LifeSync:

- Delivers predictive insights with moderate numerical accuracy
- Provides strong interpretability through SHAP and feature importance
- Offers intuitive UI-based trend simulation and CSV logging
- Enables real-world self-awareness through dashboards and forecasts

Though not clinically validated, the system performs well as a research-grade emotional forecast and simulation engine, grounded in explainable AI.

Chapter 6

Testing and Evaluation

6.1 Introduction

This chapter interprets the results and system behavior presented in Chapter 5, connecting technical findings with real-world significance. LifeSync was designed to not only predict emotional states (happiness and stress) but also to empower users with personalized forecasts and lifestyle simulations. This discussion explores how well the system fulfills these goals, examines the interpretability and usability of predictions, and reflects on the broader implications of derived metrics, such as burnout risk and the Wellness Index.

6.2 Interpretation of Results

6.2.1 Emotional Predictions

The system predicted happiness and stress using Random Forest and XGBoost models, respectively, both selected based on cross-validated performance metrics (R^2, MAE, RMSE). These models demonstrated consistent predictive capability, with Random Forest achieving higher R^2 scores for happiness, while XGBoost outperformed others in stress prediction accuracy.

Importantly, these predictions were not treated in isolation. Instead, they served as the foundation for simulated emotional trajectories, offering users insight into how small behavioral shifts (e.g., more sleep, less screen time) might influence their future well-being over 3, 7, 30, and 90-day windows.

6.2.2 Derived Metrics

The Wellness Index and Burnout Risk scores added interpretive depth. Both were derived from the base emotional predictions using custom threshold logic and a weighted average formula:

– Wellness Index combined happiness and inverse stress into a normalized metric.

– Burnout Risk was calculated using conditional thresholds (e.g., happiness ¡ 4 and stress ¿ 7), flagging high-risk users.

Although not directly trained, these values added significant value by distilling multidimensional emotional states into simpler, more actionable indicators. This bridged the gap between statistical output and real-life guidance.

6.3 Explainability with SHAP

The integration of SHAP (SHapley Additive exPlanations) was essential for model transparency. Global SHAP summary plots and local dot plots allowed users to see which features most influenced their predicted happiness and stress levels [12]. For example:

– For happiness, features such as sleep hours, exercise, and social interaction had consistently positive SHAP values.

– For stress, screen time, work hours, and diet quality showed stronger negative correlations.

These visualizations were embedded into both the dashboard and simulator, enabling dynamic, filter-based updates. By translating black-box predictions into human-interpretable charts, SHAP greatly enhanced user trust and usability.

Additionally, the system leveraged SHAP feature rankings to generate personalized lifestyle recommendations, such as "reduce screen time" or "increase social interaction." These were directly mapped from high-impact SHAP values, closing the loop between prediction and action.

6.4 Forecasting and Trend Simulations

A unique strength of LifeSync lies in its multi-step prediction capability. The simulator provided forecasted values of happiness, stress, burnout risk, and Wellness Index across 3, 7, 30, and 90-day intervals. These were generated using a fixed delta logic, where the base prediction was adjusted over time using predefined increments or decay rates. While this approach is not as robust as multi-output regression or LSTM-based forecasting, it offers an interpretable and computationally efficient method for simulating wellness trajectories.

The trend graphs showed logical progression — for example:

- Users with very low happiness and high stress saw a slight worsening over longer intervals.
- Positive lifestyle configurations showed stability or improvement over time.

Despite its simplicity, this method demonstrated the practical potential of forward-looking emotional planning — a feature lacking in most wellness apps or dashboards.

6.5 Usability and Interface Feedback

User interface components were designed using Streamlit and evaluated across two core modules:

- LifeView (Dashboard): Provided correlation heatmaps, KPI cards, SHAP visualizations, and lifestyle distributions filtered by user traits.
- SyncPredict (Simulator): Allowed users to modify lifestyle variables and view immediate and forecasted emotional outcomes with supportive emoji feedback and CSV logging.

From a usability perspective, the modular layout, clear tooltips, and intuitive visualization formats made the system accessible to non-technical users. The inclusion of CSV export also added long-term utility for personal journaling, academic use, or behavioral research.

6.6 System Strengths

Key strengths of LifeSync include:

- High model accuracy with explainability
- Real-time forecasting across four timeframes
- Derived emotional indicators for simplified understanding
- SHAP-powered, feature-based lifestyle advice
- Local-only data logging (no external privacy concerns)

Together, these components deliver a well-integrated, ethical, and educational system for emotional self-monitoring.

6.7 Limitations

Despite its accomplishments, the system has several limitations:

- Static forecasting logic lacks sensitivity to long-term user behavior or unexpected emotional shocks.
- No retraining loop: Model performance remains static unless manually updated.
- Burnout Risk logic is rule-based, not learned, which limits nuance in real-world stress modeling.
- User feedback not collected: No formal user testing was performed, reducing real-world validation.
- Recommendations are generated using SHAP-based heuristics and may oversimplify complex emotional causality.

These limitations open pathways for future enhancements and academic experimentation.

6.8 Summary

This chapter critically examined LifeSync's predictive performance, interpretability, and practical usability. The system goes beyond standard analytics by translating raw predictions into intuitive, future-facing narratives and lifestyle feedback. With the help of SHAP visualizations, trend forecasts, and derived wellness indicators, users can proactively understand and influence their emotional state.

Despite some simplifications in trend logic and a lack of retraining, LifeSync stands as a transparent, ethical, and functionally rich prototype that bridges machine learning and emotional wellness with thoughtful design.

Chapter 7

Conclusions and Future Work

7.1 Conclusion

This research project presented LifeSync, an interactive machine learning system for predicting emotional wellness based on lifestyle variables. By integrating supervised learning models, SHAP explainability, and a user-friendly simulation interface, the system empowers users to forecast emotional outcomes and make informed lifestyle adjustments. LifeSync operates through two primary modules — LifeView for exploration and SyncPredict for personalized forecasting — ensuring that both insight and simulation are accessible to the user.

All objectives outlined in Chapter 1 were achieved: • Accurate prediction of happiness and stress using Random Forest and XGBoost regressors, respectively • Time-based forecasting using logic-driven projections across 3, 7, 30, and 90 days • Derived wellness metrics including Wellness Index and Burnout Risk, calculated post-prediction • Global and local SHAP visualizations for transparent interpretation of predictions • CSV-based logging of all predictions and inputs, establishing a traceable history of simulated behavior

While the system is intended for educational and self-reflective purposes rather than clinical use, it successfully demonstrates how machine learning can support wellness awareness, behavior simulation, and transparent digital feedback.

7.2 Future Work

Despite its functional success, LifeSync opens up several promising directions for expansion:

1. Personalized Model Retraining

The current model is trained on a generic dataset. Future iterations could:

- Use the CSV logs of user inputs and predictions to trigger retraining of user-specific models after a defined volume of data
- Incorporate personal SHAP profiles that evolve over time to reflect changing lifestyle habits

2. Advanced Forecasting Techniques

The existing forecast logic applies fixed delta-based logic over time. To increase realism:

- Integrate time-series models like LSTM or Facebook Prophet for trend learning
- Apply nonlinear decay/growth models to adjust future predictions dynamically
- Explore feedback-aware learning, where user actions impact future forecasts

3. Mobile App and Smart Device Integration

LifeSync is currently deployed as a Streamlit web app. The future roadmap includes:

- Mobile app deployment on Android/iOS for daily tracking and notifications
- Integration with wearable devices and health APIs (e.g., Fitbit, Apple Health) to auto-populate simulator inputs like sleep, heart rate, or screen time
- Automated CSV population for seamless, real-time wellness journaling

4. Coaching, Feedback Loops, and User Support

To enhance interactivity and engagement:

- Introduce weekly summaries and behavioral reports based on forecast logs
- Implement adaptive recommendation logic using SHAP output intensity
- Add conversational interfaces (chatbots or voice input) for simplified usage

5. Clinical Applications and Research Collaboration

Although LifeSync is not a diagnostic tool, future opportunities include:

- Collaborating with psychologists, coaches, or wellness researchers
- Conducting clinical trials to evaluate predictive reliability in controlled settings
- Embedding data privacy protocols for ethical deployment in sensitive environments

7.3 Final Thoughts

LifeSync is more than just a machine learning project — it is a step toward practical self-awareness through data. By blending technical accuracy with user-friendly design and emotional sensitivity, this system demonstrates the power of technology to support wellness, reflection, and preventative care.

Through prediction, simulation, and explanation, LifeSync transforms passive lifestyle data into actionable foresight — helping users make conscious, future-informed choices. The foundation has been laid; the next step is to build on it — with personalization, automation, and human empathy at the core.

Bibliography

[1] R. Bhargava, "Interpretable machine learning with shap," 2018. Accessed: 2025-06-01.

[2] Kaggle, "Mental health and lifestyle dataset," 2023. Accessed: 2025-03-25.

[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (San Francisco), pp. 785–794, ACM, 2016.

[4] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: CRC Press, 2012.

[5] S. Raschka and V. Mirjalili, *Python Machine Learning*. Birmingham: Packt Publishing, 2nd ed., 2017.

[6] Streamlit Inc., *Streamlit Documentation*, 2023. Accessed: 2025-04-02.

[7] Python Software Foundation, *Python Language Reference, version 3.10*, 2023. Accessed: 2025-04-20.

[8] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[9] M. Wes, *Pandas Documentation*, 2018. Accessed: 2025-04-03.

[10] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[11] M. L. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

Appendix A

Appendix

This appendix contains screenshots and graphical elements from the LifeView dashboard, which provides SHAP-based global insights, correlation heatmaps, and lifestyle analytics using filter-driven data exploration.

A.1 Code Snippets from Backend Modules

This appendix features important Python code snippets used in model training, SHAP explanation, and frontend integration.

Listing E.1 – Random Forest Regressor Training Snippet

model_rf = RandomForestRegressor() model_rf.fit(X_train, y_train_happiness)

Listing E.2 – SHAP Value Calculation with TreeExplainer

explainer = shap.TreeExplainer(model_rf) shap_values = explainer.shap_values(X_test)

Listing E.3 – Streamlit Input Block for SyncPredict

sleep_hours = st.slider("Sleep Hours", 0, 12, 6) screen_time = st.slider("Screen Time (hours)", 0, 16, 4)

Listing E.4 – CSV Logging Logic

output_df.to_csv("prediction_results.csv", index=False)

A.2 CRISP-DM Process Mapping

This appendix outlines how the project followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework.

F.1 Business Understanding Understanding the goal of linking lifestyle metrics to emotional outcomes.

F.2 Data Understanding Exploration of the Kaggle dataset — 3,000 entries with lifestyle-emotion mappings.

F.3 Data Preparation Preprocessing: imputation, encoding, scaling, and outlier removal.

F.4 Modeling Regression models trained for happiness and stress using cross-validation.

F.5 Evaluation Performance measured using R^2, MAE, RMSE; models selected accordingly.

F.6 Deployment System deployed using Streamlit with prediction logging and visualization.

A.3 SHAP-Based Recommendation Mapping Logic

This appendix describes the logic that maps SHAP values to lifestyle recommendations displayed in SyncPredict.

G.1 Recommendation Engine Design Each SHAP value is thresholded (positive/negative) and mapped to a predefined suggestion string.

G.2 Mapping Table (Sample)

Feature SHAP Value Condition Recommendation Text Sleep Hours Low SHAP value "Try to improve your sleep." Screen Time High SHAP value "Reduce screen time where possible." Social Interaction Low SHAP value "Engage in more social activity."

G.3 Recommendation Intensity SHAP magnitude is classified into LOW, MEDIUM, HIGH — affecting urgency tone of feedback.

LifeSync_Report

ORIGINALITY REPORT

6%	5%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ www.coursehero.com

Internet Source

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		