1. Introduction

1.1 Purpose

The purpose of this Software Requirements Specification (SRS) document is to define the functional and non-functional requirements of the software system titled “Data-Driven Estimation of Obesity Risk from Lifestyle Habits.” This system is designed to utilize machine learning algorithms to analyze lifestyle and physical data in order to predict an individual’s obesity category.

1.2 Scope

The software will allow users to input data related to their physical condition and daily habits. This includes attributes such as age, weight, height, physical activity, eating habits, and more. The software will process the data using classification and regression techniques to determine obesity levels, and clustering algorithms will be used to find patterns in user behavior. The system is intended for use by health professionals, researchers, and individuals interested in monitoring health metrics.

1.3 Definitions, Acronyms, and Abbreviations

- **BMI**: Body Mass Index
- **ML**: Machine Learning
- **SRS**: Software Requirements Specification
- **CSV**: Comma-Separated Values

1.4 References

- IEEE Recommended Practice for Software Requirements Specifications (IEEE 830-1998)
- https://www.kaggle.com/datasets
- https://scikit-learn.org/stable/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7682147/
- https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight
- https://towardsdatascience.com
- https://scholar.google.com/
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning.
- Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
- Obesity Classification Dataset – Kaggle
- Journal of Obesity and Metabolic Syndrome

1.5 Overview

This document includes a detailed overview of the proposed system, including system features, external interfaces, constraints, and other essential details. The primary aim is to serve as a foundation for system development and verification.

---

2. Overall Description

2.1 Product Perspective The product is a standalone developed using Python libraries such as Scikit-learn, Pandas, NumPy, and Matplotlib. It will support both manual data entry and batch input via CSV files.

2.2 Product Functions

- Capture and validate user data via a user-friendly interface.
- Normalize and preprocess the data for machine learning input.
- Train classification and regression models on the dataset.
- Predict obesity levels based on trained models.

- Visualize results in an interpretable format.
- Group similar users using unsupervised clustering.

2.3 Operating Environment

- Google Colab
- Required Libraries: Flask/Streamlit, Scikit-learn, Pandas, NumPy, Matplotlib
- OS: MacOS

2.4 Design and Implementation Constraints

- The prediction accuracy is directly related to the quality and variety of input data.

2.5 Assumptions and Dependencies

- The data used for training is accurate and representative.
- The application will have access to required libraries and dependencies.

---

3. External Interface Requirements

3.1 Software Interfaces

- Python libraries (e.g., Flask/Streamlit, Scikit-learn)

---

4. System Features

4.1 Data Preprocessing Module

- Handling missing values
- Label encoding and normalization
- Train-test data splitting

4.2 Machine Learning Module

- Classification: Logistic Regression, SVM, Random Forest, Gradient Boosting
- Regression: Linear Regression, Ridge, Lasso, Random Forest Regressor
- Clustering: KMeans for grouping lifestyle behaviors

4.3 Prediction Module

- Predict obesity level
- Calculate accuracy, precision, recall, and RMSE (for regression)

4.4 Visualization Module

- Confusion Matrix, Scatter Plots, Cluster Maps, etc.

4.5 Result Display Module

- Output obesity category

---

5. Appendices

Appendix A: Sample Input Features

- Gender

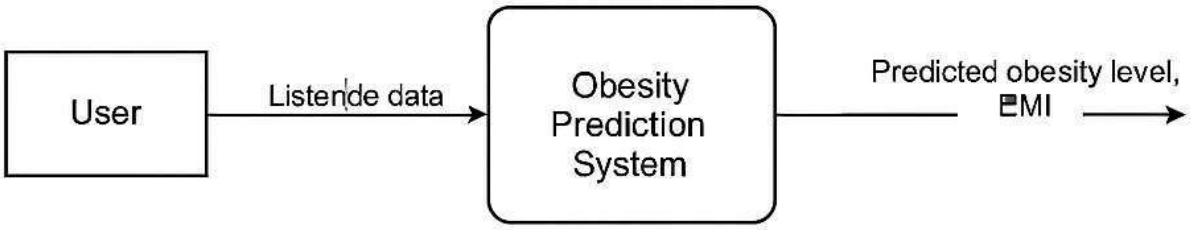
- Age
- Height
- Weight
- Frequency of eating high-calorie food
- Physical activity level
- Water consumption
- Smoking habits
- Daily screen time

Appendix B: Output Categories

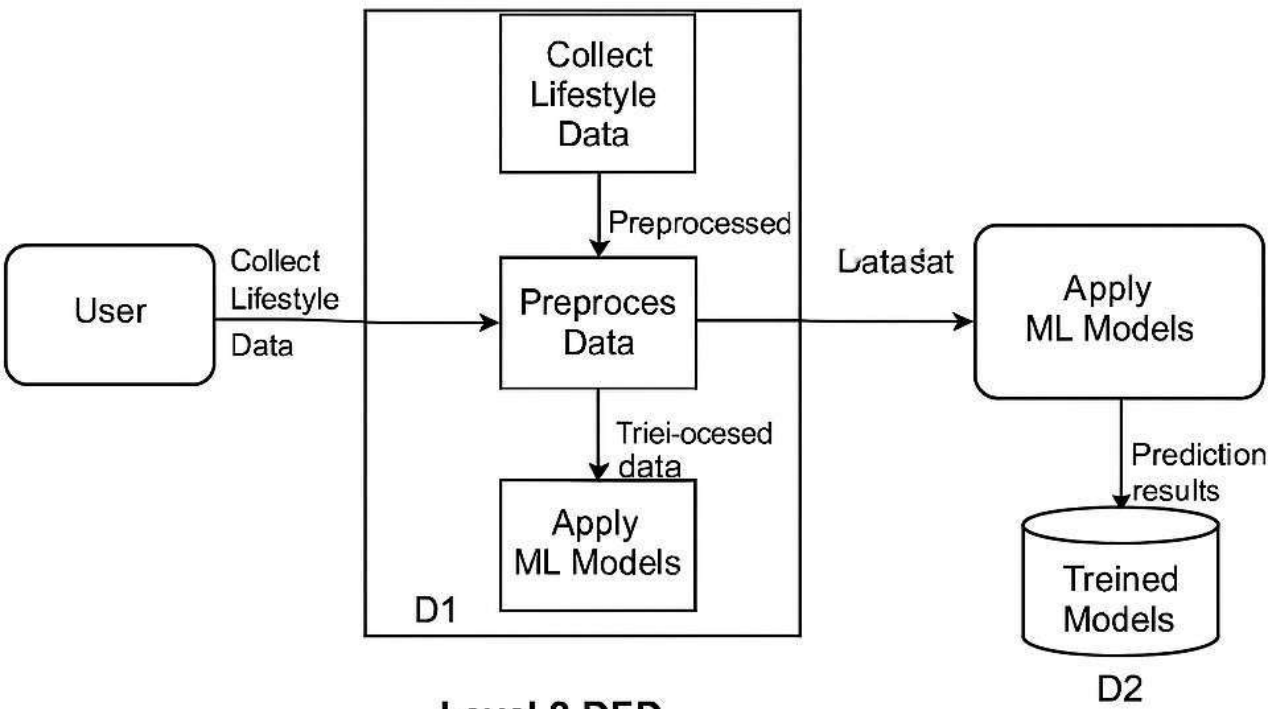
- Underweight
- Normal Weight
- Overweight Level I
- Overweight Level II
- Obesity Type I
- Obesity Type II
- Obesity Type III

6. Data Flow Diagram (DFD)

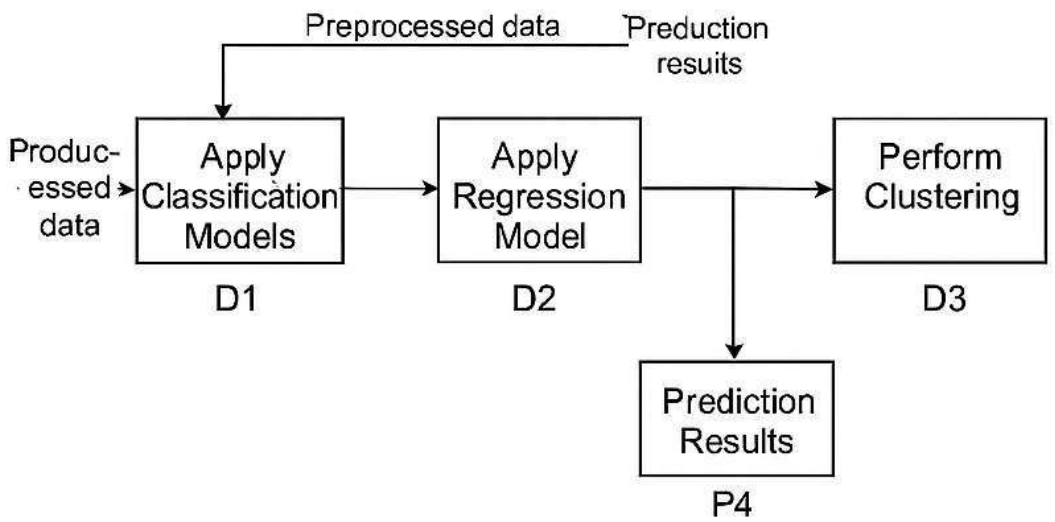
Context Level DFD (Level 0)



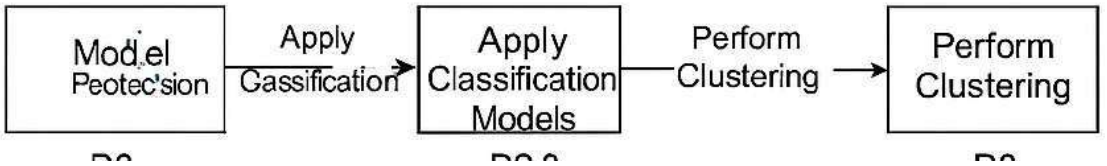
Level 1 DFD



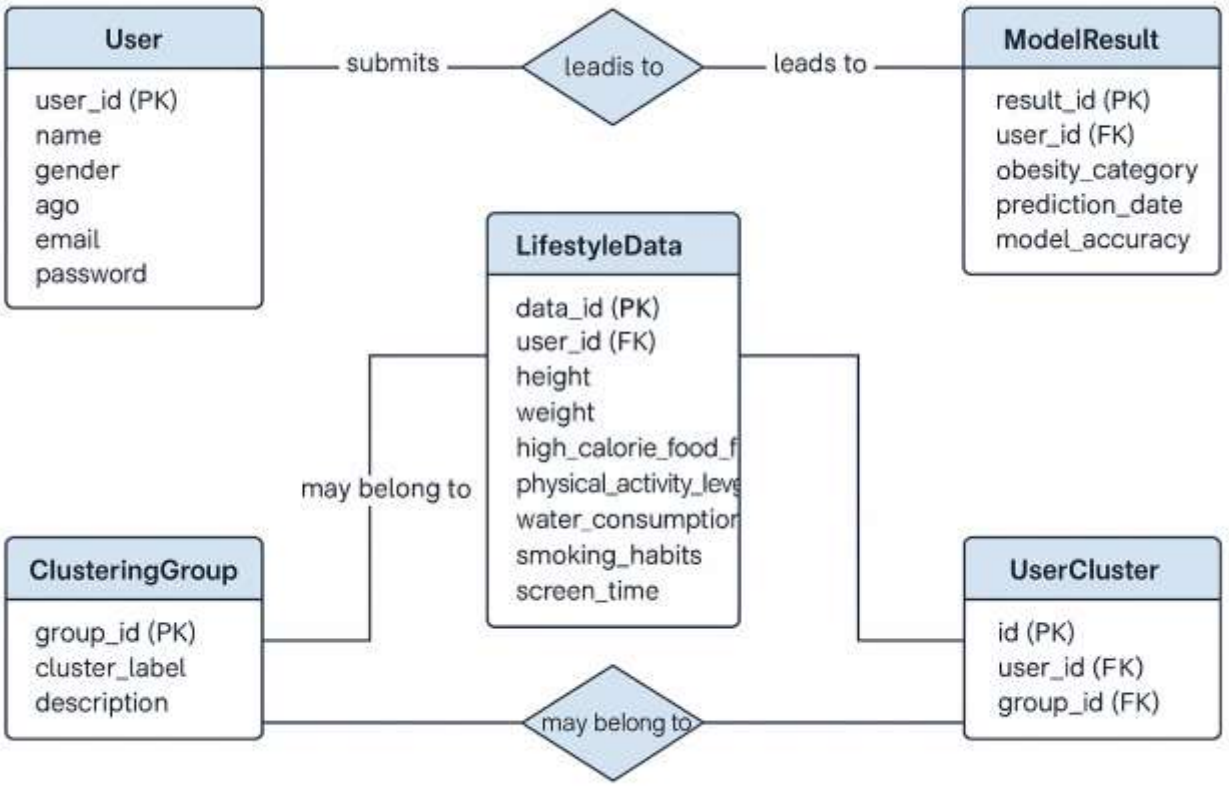
Level 2 DFD



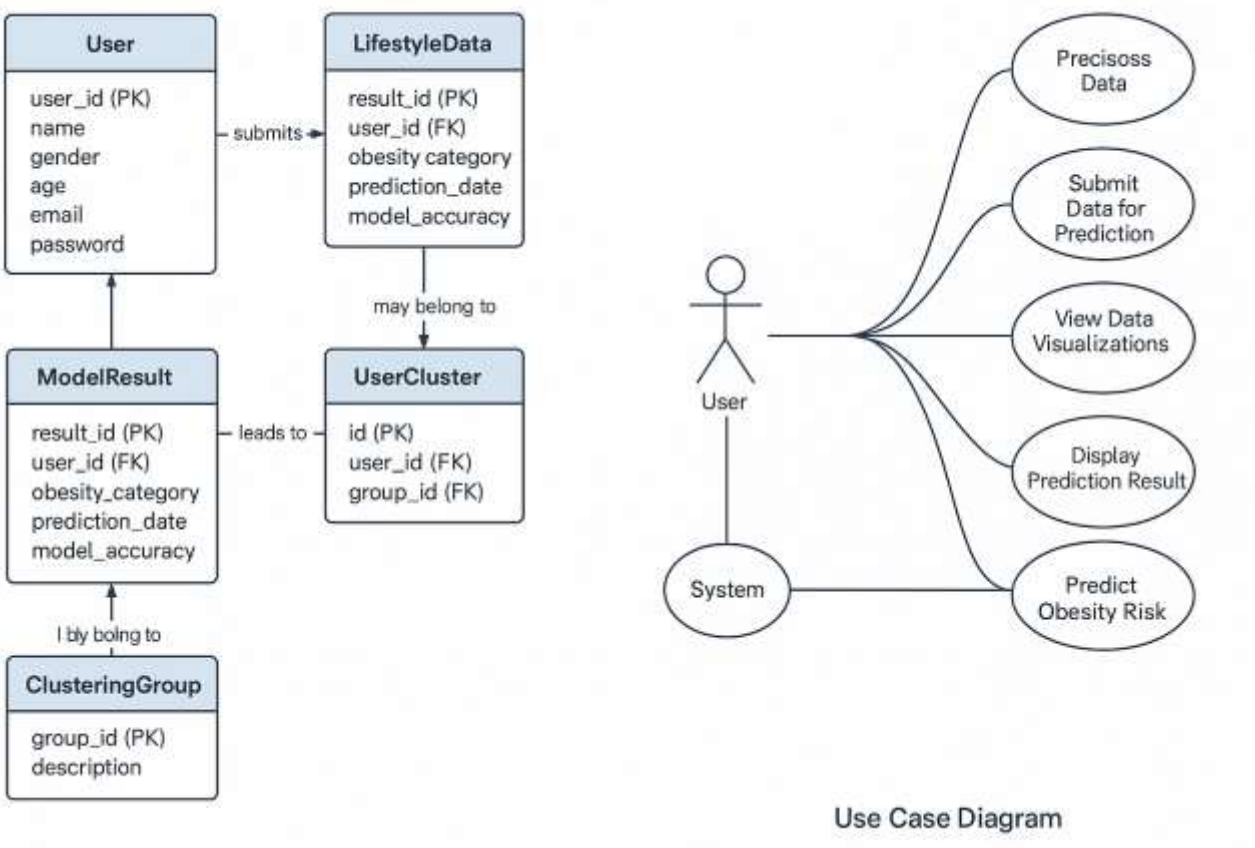
Level 2 DDF



7. Entity Relationship Diagram



8. Use Case



Use Case Diagram