PROJECT DESCRIPTION

Hanee Rai, Shipra Guin, Akash Koottungal, Neeraj Chauhan

July 2024

1 Introduction

This document outlines the development and implementation of the Obesity Level Prediction Dashboard. The project aims to integrate advanced data analytics and AI to provide users with insightful and actionable data on obesity risk factors, contributing to better health management and informed decision-making. By leveraging cutting-edge machine learning techniques and user-friendly visualizations, the dashboard offers an innovative approach to understanding and managing obesity.

2 Project Goals

The primary aim of this project was to create an Obesity Level Prediction Dashboard. This interactive platform uses advanced data analytics and machine learning to provide insights into factors influencing obesity, offering personalized risk assessments and recommendations for obesity management. The goal was to make these complex data interactions accessible to both laypersons and professionals, enabling informed decision-making through an intuitive user interface. Additionally, the project seeks to raise awareness about obesity risk factors and promote healthier lifestyles through data-driven insights.

3 Intended Audience

The dashboard is primarily designed for the general public, including individuals keen on managing their personal health and understanding risk factors associated with obesity. It provides easy-to-understand risk assessments and data-driven insights that can also be valuable for health enthusiasts and educators who wish to promote healthy living practices. The dashboard is designed to make complex data easy to understand, helping users make informed health decisions with confidence.

4 Dataset Information

The project utilizes a comprehensive dataset consisting of 2,111 records, collected from populations in Mexico, Peru, and Colombia. This dataset includes 17 distinct attributes such as age, gender, family medical history, daily diet, and physical activity levels, among others. Notably, 77% of the data was generated using the SMOTE filter in the Weka tool to balance the classes, while the remaining 23% was collected directly from users through a web platform. This diverse dataset provides a robust foundation for predicting obesity levels and understanding the contributing factors.

5 Design Decisions

Our design decisions were driven by the need to make the interface accessible and informative for the end user:

- **Visualization Tools:** We incorporated a variety of interactive charts and graphs to allow users to visually explore data correlations and distributions.
- **Predictive Modeling:** We used machine learning algorithms to predict individual obesity risk levels based on input features, providing users with personalized risk assessments.
- **Explanation Features:** The interface includes detailed explanations of the model's predictions to help users understand the reasoning behind their individual assessments. Additionally, SHAP values are used to explain the impact of each feature on the predictions.

6 Interface Description

The Obesity Level Prediction Dashboard interface is designed to be user-friendly and interactive, ensuring that users can easily navigate and access the information they need. Built with Streamlit, the interface includes a sidebar for navigation, allowing users to switch between different sections such as Home, Dataset Overview and Exploration, Model Performance, Explainability, Obesity Estimation Tool, and User Feedback.

7 Challenges Faced

- **Data Quality and Preparation:** The dataset used in this project includes a significant proportion—77%—of synthetic data, generated using the Weka tool and the SMOTE algorithm to enhance the volume and variety of scenarios for analysis. This approach allowed us to address the common issue of class imbalance in medical datasets, which can skew

the learning process. To ensure that the synthetic data accurately represented real-world conditions without introducing bias, extensive preprocessing was undertaken. This included normalizing features and encoding categorical variables. Rigorous data analysis protocols were followed to enhance the overall reliability and applicability of the dataset for obesity risk prediction.

- **Model Accuracy and Interpretability:** Developing a model that maintained high accuracy while remaining interpretable to non-expert users was challenging. Balancing these two aspects required careful selection and tuning of algorithms, as well as the integration of tools to explain model predictions clearly.

- **Conflict Between Features:** For example, the FCVC feature, where the dataset description on UCI mentions it as an integer, but further analysis of the research paper and dataset exploration revealed that the dataset contains continuous values (floating point values). The research paper suggests binning them into categories. We encountered similar issues with a few other variables as well, necessitating careful preprocessing and feature engineering.

8 Interface

To provide a visual overview of our interface, we include the following screenshots:

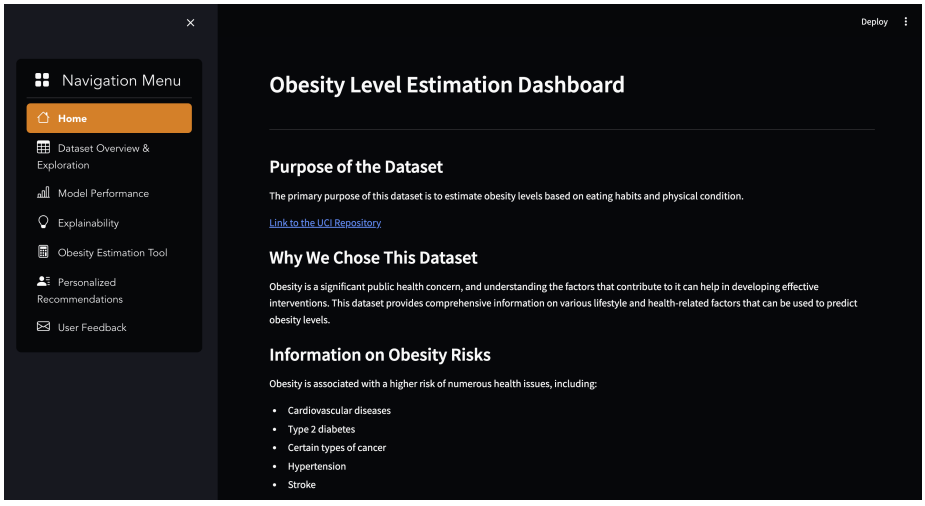

Figure 1: Risk Assessment Tool.

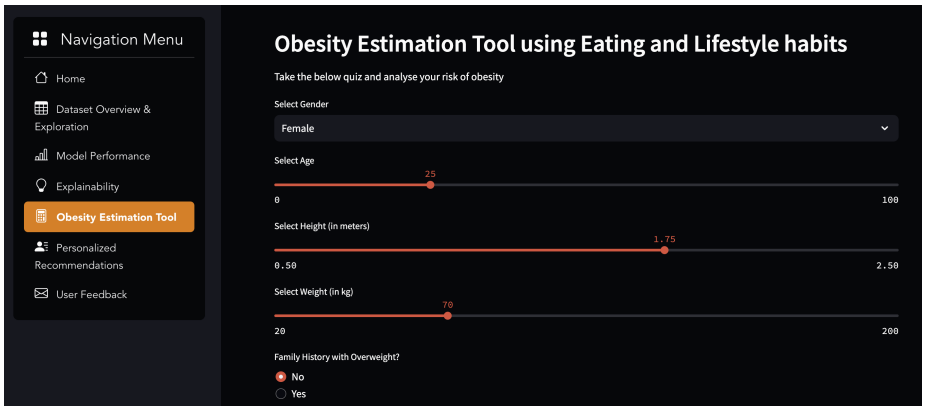

Figure 2: Detailed Input Interface.

9 Reflections

Through this project, we gained invaluable insights into the practical applications of AI in healthcare. The iterative design and feedback process highlighted the importance of user-centric design and adaptive learning systems in developing practical health informatics tools. This project also underscored the critical nature of data quality and the need for robust preprocessing to inform health-related predictions accurately. Moreover, the experience emphasized the potential of machine learning to provide actionable health insights and the importance of making these insights accessible to a broad audience.

10 Future Directions

Future enhancements will focus on incorporating real-time data integration, improving predictive accuracies, and expanding the tool's features to include more personalized health recommendations based on user feedback. Potential developments include integrating wearable device data for real-time health monitoring, enhancing the model with additional health indicators, and providing more detailed lifestyle recommendations to users.

11 Conclusion

The Obesity Level Prediction Dashboard simplifies the power of integrating AI with user-friendly interfaces to address complex health issues. By demystifying data analytics and making predictive insights accessible, this AI tool aids in proactive health management, ultimately contributing to better health outcomes and a more informed public. The project demonstrates the potential of AI

in healthcare and the importance of making data-driven insights available to everyone.