

Rukayat Balogun

Final Draft - 2311143

-  Final thesis submission upload link for the Examiner Part 1 (Moodle TT)
 -  Masters Project_SAY (Moodle TT)
 -  University of Bolton
-

Document Details

Submission ID**trn:oid:::2:780971266****95 Pages****Submission Date****Sep 12, 2024, 11:05 PM GMT+1****23,372 Words****Download Date****Sep 17, 2024, 9:27 AM GMT+1****141,254 Characters****File Name****150433_Rukayat_Balogun_Final_Draft_-_2311143_2315156_1278964911.docx****File Size****9.8 MB**

81% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

1 AI-generated only 79%

Likely AI-generated text from a large-language model.

2 AI-generated text that was AI-paraphrased 2%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





School of Art and Creative Technologies
M.Sc. Data Analytics and Technologies
University of Bolton

Developing a Data-Driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Prepared by Rukayat Balogun

Student ID 2311143

A report submitted in the part fulfilment for the MSc degree in Data Analytics and Technologies

September 2024

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Dedication

This thesis is dedicated to all individuals facing the challenges of obesity and sedentary lifestyles, whose resilience and determination to improve their health inspired the development of this personalised fitness application. This work is also dedicated to the medical professionals and researchers committed to advancing public health through innovation in data-driven technologies.

University of Bolton

Acknowledgement

I would like to express my deepest and heartfelt gratitude to my supervisor, Dr. Celestine Iwendi. From the moment I enrolled at the University, Dr. Iwendi saw potential in me and has since become not only a mentor but a source of inspiration. His guidance has been invaluable throughout my academic journey. Dr. Iwendi's insightful feedback, encouragement, and unwavering support have played a pivotal role in the development and completion of this research. His passion for excellence has pushed me to reach beyond my limits and has had a profound impact on both my academic and personal growth.

I also extend my sincerest thanks to Dr. and Mrs. Adetuyi for their constant encouragement and mentorship. Your words of wisdom, care, and belief in my abilities have been instrumental throughout this process. I am deeply appreciative of the time and effort you both have invested in guiding me.

To my dear sister, Mrs. Oso, and her husband, I owe special thanks. Your dedication to the medical profession has not only been a source of personal inspiration but has also motivated me to contribute to fields that improve human health and well-being. Your passion and commitment to your work have shown me the importance of perseverance and purpose.

My deepest appreciation also goes to my parents, Mr. and Mrs. Balogun. Your unwavering love, support, and belief in me have been the bedrock of my strength. You have stood by me through every challenge, offering both emotional and practical support when I needed it most. Without your sacrifices and encouragement, this achievement would not have been possible.

Lastly, I want to acknowledge my friends and colleagues, whose support and camaraderie have been indispensable throughout my academic experience. I would also like to thank the faculty and staff at the University of Bolton for providing a nurturing environment that has allowed me to thrive.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Table of Contents

<u>ABSTRACT</u>	9
<u>CHAPTER 1: INTRODUCTION</u>	10
1.0 INTRODUCTION	10
1.1 BACKGROUND OF STUDY	10
1.2 HEALTH IMPLICATIONS OF OBESITY AND SEDENTARY LIFESTYLES	11
1.3 STATEMENT OF THE PROBLEM	12
1.4 PURPOSE OF THE RESEARCH	12
1.5 RESEARCH AIMS	13
1.6 RESEARCH QUESTIONS	13
1.7 RESEARCH OBJECTIVES	13
1.8 RELEVANCE AND IMPORTANCE OF THE STUDY	13
1.9 SCOPE OF THE STUDY	14
1.10 SUMMARY	14
<u>CHAPTER TWO – LITERATURE REVIEW</u>	15
2.1 INTRODUCTION	15
2.2 CURRENT PUBLIC HEALTH GUIDELINES AND INTERVENTIONS	22
2.3 CONSTRAINTS OF CONVENTIONAL DIET AND FITNESS PROGRAMMES	22
2.4 PUBLIC HEALTH PROGRAMMES FOCUSING ON OBESITY AND INACTIVE BEHAVIORS	23
2.5 DIFFICULTIES IN PERSONALISED HEALTH RECOMMENDATIONS	23
2.6 ADDRESSING OBSTACLES IN TAILORED HEALTH RECOMMENDATIONS	23
2.7 INCORPORATING WEARABLE TECHNOLOGY AND SMARTWATCH DATA	23
2.8 UTILISING KNOWLEDGE GRAPHS FOR PROVIDING DIETARY RECOMMENDATIONS	24
2.9 PREDICTIVE MACHINE LEARNING MODELS FOR OBESITY ASSOCIATED GENE EXPRESSION	24
2.10 STRATEGIES TO ENCOURAGE HEALTH EATING HABITS IS EDUCATION	24
2.11 HEALTH AND WELLNESS SYSTEMS	24
2.12 DIETARY GUIDELINES FOR SPECIFIC HEALTH CONDITIONS	25
2.13 EXAMINING GENDER DISPARITIES IN OBESITY AND HEALTH INTERVENTIONS	25
2.13.1 LIMITATIONS OF TRADITIONAL DIET AND FITNESS PLANS:	25
2.13.2 PUBLIC HEALTH INITIATIVES TARGETING OBESITY AND SEDENTARY LIFESTYLE	25
2.13.3 CHALLENGES IN PERSONALIZED HEALTH RECOMMENDATIONS	26
2.14 STRATEGIES FOR OVERCOMING CHALLENGES:	27
2.15 SUMMARY	27
<u>CHAPTER 3 – METHODOLOGY</u>	29
3.1 INTRODUCTION	29
3.2 RESEARCH PHILOSOPHY	29
3.3 PROPOSED WORKFLOW	30
3.3.1 DATA COLLECTION	31
3.3.2 DATA PREPROCESSING	32
3.3.3 EXPLORATORY DATA ANALYSIS (EDA)	33
3.3.4. MACHINE LEARNING MODEL DEVELOPMENT	38
3.3.5 WEB APPLICATION DEVELOPMENT USING DJANGO	39

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

3.3.6 INTEGRATION OF MACHINE LEARNING USING DJANGO	40
3.3.7 RECOMMENDATION ENGINE DEVELOPMENT.....	40
3.3.8 EVALUATION AND TESTING	44
3.4 DATA ANALYSIS PLAN	44
3.5. MACHINE LEARNING	45
3.5.1 LINEAR REGRESSION	45
3.5.2 RIDGE REGRESSION	45
3.5.3 LASSO REGRESSION	45
3.5.4 DECISION TREE REGRESSOR	45
3.5.5 RANDOM FOREST REGRESSOR.....	45
3.5.6 SUPPORT VECTOR MACHINE.....	45
3.5.7 GRADIENT BOOSTING	46
3.6 RATIONALE FOR MACHINE LEARNING ALGORITHM	46
3.7 ETHICAL CONSIDERATIONS	46
3.8 SUMMARY	46
CHAPTER 4 – RESULT AND DISCUSSION	48

4.1 INTRODUCTION	48
4.2 DATA PRE-PROCESSING RESULTS.....	48
4.2.1 DATA CLEANING.....	48
4.2.2 HANDLING INCONSISTENT DATA	49
4.2.3 HANDLING MISSING VALUES	49
4.2.4 OUTLIER DETECTION AND HANDLING	50
4.3 EXPLORATORY DATA ANALYSIS (EDA).....	51
4.3.1 DISTRIBUTION OF BMI	51
4.3.2 FEATURE CORRELATION ANALYSIS	51
4.3.3 RELATIONSHIP BETWEEN HEIGHT AND BMI.....	51
4.3.4 JOINT DISTRIBUTION OF HEIGHT AND WEIGHT	52
4.3.5 PAIRWISE RELATIONSHIP BETWEEN FEATURES.....	52
4.4 MODEL TRAINING AND VALIDATION	53
4.4.1 MODEL SELECTION.....	53
4.4.2 DATA SPLITTING.....	53
4.4.3 MODEL TRAINING	54
4.4.4 MODEL VALIDATION.....	54
4.5 MODEL PERFORMANCE COMPARISON	55
4.5.1 BEST MODEL SELECTION	55
4.5.2 DISCUSSION OF MODEL PERFORMANCE.....	55
4.6 BMI PREDICTION AND CLASSIFICATION RESULTS	55
4.6.1 CATEGORIZATION OF BMI PREDICTIONS	56
4.6.2 CLASSIFICATION METRICS	56
4.7 RECOMMENDATION SYSTEM OUTCOMES.....	58
4.7.1 PERSONALIZED RECOMMENDATIONS	58
4.7.2 USER FEEDBACK AND SYSTEM REFINEMENT	58
4.8 SUMMARY	59

CHAPTER 5 – DISCUSSION OF RESULTS	60
--	-----------

5.1 INTRODUCTION	60
5.2 STATISTICAL ANALYSIS AND DATA PREPROCESSING.....	60

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

5.3 EXPLORATORY DATA ANALYSIS (EDA).....	60
5.4 MACHINE LEARNING MODEL PERFORMANCE	61
5.5 DISCUSSION ON THE RECOMMENDATION SYSTEM	61
5.6 COMPARISON WITH EXISTING LITERATURE.....	61
5.7 IMPLICATIONS FOR PRACTICE	63
5.8 CHALLENGES AND LIMITATIONS	64
5.8.1 DATA QUALITY AND AVAILABILITY	65
5.8.2 MODEL GENERALIZATION	65
5.8.3 USER ENGAGEMENT.....	65
5.9 SUMMARY	65
 CHAPTER 6 – CONCLUSION, RECOMMENDATIONS, FUTURE WORK, AND LIMITATIONS	66
 6.1 INTRODUCTION	66
6.2 SUMMARY OF FINDINGS.....	66
6.3 RECOMMENDATIONS	67
6.4 FUTURE WORK.....	67
6.5 ADDRESSING RESEARCH QUESTION.....	68
6.6 FINAL THOUGHTS	68
6.7 SUMMARY	69
 REFERENCE	70
 BIBLIOGRAPHY.....	76
 APPENDIX – PROJECT TIMELINE	77
 APPENDIX – DATA ANALYSIS CODE	79
 APPENDIX – WEB APPLICATION CODE	85
 APPENDIX – WEB APPLICATION FRONTEND.....	93

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

LIST OF TABLES

TABLE 2.1 AN OVERVIEW OF ACADEMIC PUBLICATIONS AND THEIR RELEVANCE TO RESEARCH	15
TABLE 6.5 ADDRESSING RESEARCH QUESTION	68

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

LIST OF FIGURES

FIGURE 1: AN ILLUSTRATION OF MACHINE LEARNING ALGORITHM ON FITNESS AND DIETARY APP	10
FIGURE 3.3: PROPOSED WORKFLOW FOR THE WEB APPLICATION USING DJANGO	30
FIGURE 3.3.1: SNAPSHOT OF THE DATA.....	32
FIGURE 3.3.3.1: DISTRIBUTION OF BMI.....	34
FIGURE 3.3.3.2: FEATURE CORRELATION MATRIX.....	35
FIGURE 3.3.2.3: HEIGHT VS BMI	35
FIGURE 3.3.2.4: JOINT PLOT OF HEIGHT AND WEIGHT COLOURED BY BMI.....	36
FIGURE 3.3.2.5: PAIRWISE RELATIONSHIP.....	37
FIGURE 3.3.3.6: WEIGHT VS BMI	38
FIGURE 3.3.7.4: ARCHITECTURE OF THE WEB APP	42
FIGURE 4.4.4: MODEL VALIDATION.....	54
FIGURE 4.6.2 CLASSIFICATION METRICS	57
FIG 5.5: COMPARISON OF MODEL PERFORMANCE (MSE AND ACCURACY)	62

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Abstract

Obesity and sedentary lifestyles are leading cause of chronic health issues worldwide, including cardiovascular disease, diabetes, and hypertension. Traditional one-size-fits-all approaches to diet and exercise recommendations often fail to address individual needs, leading to suboptimal health outcomes and low adherence rates. This study presents the development and evaluation of a personalized fitness web application using machine learning techniques, aimed specifically at obese and sedentary individuals. By leveraging primary user data using Django Forms and secondary data from the American Time Use Survey (ATUS), the system provides tailored fitness and dietary recommendations based on individual user profiles. A range of machine learning models, including Linear Regression, Decision Trees, Support Vector Machines, and Gradient Boosting, were evaluated for their effectiveness in predicting Body Mass Index (BMI) and personalizing recommendations. The Gradient Boosting model demonstrated superior predictive accuracy, achieving Mean Squared Error (MSE) of 0.1868 and an accuracy rate of 92.5%. Comparative analysis with existing literature showed significant improvements in predictive accuracy and user engagement, highlighting the potential of personalized, data-driven health interventions to improve user adherence and health outcomes. User feedback and engagement metrics further confirmed the system's usability and effectiveness. This research underscores the importance of combining advanced machine learning techniques with user-centred design to enhance the effectiveness of digital health interventions.

Keywords: Obesity, Sedentary Lifestyle, Personalized Fitness, Machine Learning, Web application, Body Mass Index (BMI), Gradient Boosting, Health Interventions, User Engagement, Data-driven Recommendations

Chapter 1: Introduction

1.0 Introduction

Fitness is a cornerstone of a healthy lifestyle, offering a multitude of benefits for both our body and minds (Magallanes, 2024). While the drawbacks of inactivity are well-documented, including obesity, weakened muscles, and reduced cardiovascular health, many struggle to incorporate sufficient exercise into their routines (Bruback, 2024). Time limitations, health concerns, and lack of access to facilities can all be barriers to achieving the recommended fitness level (Starns et al., 2024). This study aims at developing a personalized fitness web application using Python Django Framework and a health dataset to build a machine learning algorithm that can recommend fitness exercise for obese and sedentary individuals based on the user preference as shown in Figure 1.



FIGURE 1: AN ILLUSTRATION OF MACHINE LEARNING ALGORITHM ON FITNESS AND DIETARY APP

1.1 Background of Study

The world is facing a global growing public health crisis; the twin epidemics of obesity and sedentary lifestyles (Umekar and Joshi, 2024). Sedentary as defined by Yadav is a physical inactivity behaviour characterised by low energy expenditure, has become a prominent feature of contemporary lifestyles. The combination of our progressively modern surroundings and technological progress promotes sedentary behaviour that results to a

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

culture of sitting. His research demonstrates a definitive correlation between a sedentary lifestyle and a variety of chronic ailments, such as obesity, cardiovascular disease, and type 2 diabetes (Yadav, Research and Tiwari, 2024). The World Health Organization (WHO) has officially designated obesity as the most significant danger to the health of westernised countries. According to WHO, around 40% of adults in the United States are categorised as obese.

Rodbard defined obesity has a multifaceted problem with a substantial genetic element, while food and other factors also contribute. His research indicates that approximately 40-70% of an individual's predisposition to obesity is governed by their genetic makeup. Nevertheless, his research also indicates a pivotal correlation between genetics and environment. He demonstrated that engaging in more physical exercise can reduce the impact of a greater genetic predisposition to fat. This underscores the constraints of existing universal approaches to recommendation for fitness (Rodbard et al., 2024).

Although the advantages of fitness for weight control have been well-documented, a significant number of people find it challenging to start and sustain healthy routines (Rodbard et al., 2024). Current Web-based programmes provide little assistance, typically including exercise schedules and educational materials (Fjellström et al., 2024). Nevertheless, these programmes generally lack customisation, as they do not consider individual variances in parameters such as genetics and fitness levels (Author King, 2023). He also stated that the absence of personalization may be a contributing factor to the frequently observed low rates of adherence in web-based fitness regimens.

1.2 Health Implications of Obesity and Sedentary Lifestyles

The research conducted by (Ghosh et al., 2023), which specifically targets a demographic of adults from South Asia, offers significant insight into the chronic illness linked to obesity and detrimental effects of sedentary on one's well-being. The study highlights the correlation between obesity and a group of long-lasting illnesses, specifically cardiovascular diseases (CVD) and type 2 diabetic mellitus (T2DM). In research conducted by Suanrueang, (2024) Hypertension is a significant risk factor for cardiovascular disease (CVD), was observed to be more common among females. Central obesity, as defined by visceral fat percentage, may be a more powerful predictor of hypertension than body mass index (BMI) alone. The research indicates a strong correlation between obesity and T2DM, with the frequency of the latter rising in tandem with age and weight gain. Furthermore, the study by Ghosh, (2023) highlights a greater prevalence of additional obesity-related illness such as hypothyroidism and arthritis, particularly among the elderly population.

In the research conducted by Ozsoy, he also emphasises the adverse effects of sedentary on health and well-being. He highlighted that a lack of physical activity can worsen the impact of obesity on long-term health conditions although all participants, regardless of their weight or age showed indications of obesity, the co-morbidity effects were more noticeable in individuals with a sedentary lifestyle, especially in the older age category. The researcher also highlighted a deficiency in the desire of sedentary in engaging in physical activities even when the individuals are cognizant of its advantages (Ozsoy et al., 2024a). This underscores the

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

difficulty of encouraging a fitness lifestyle particularly when individuals face barriers such as a limited access to suitable amenities or deeply ingrained sedentary lifestyles.

1.3 Statement of the Problem

Traditional diet and exercise plans use a one-size-fits-all approach, they frequently fall short of meeting individual needs. Subpar results result from these systems' failure to take into account individual differences in metabolism, physical capabilities, medical histories, and preference (Papry et al., 2024) The particular demands are not met by the generalized plan which results in low adherence and little improvement in health. This emphasizes the need for tailored health therapies that can accommodate individual variances and advance improved health outcomes.

Current recommendation systems have many drawbacks, especially the ones that use collaborative filtering algorithms. For example, the cold start issue occurs when the system is unable to accurately deliver recommendations due to insufficient or no knowledge about new users or things. This problem is made more complex by sparse data, since less interaction data makes collaborative filtering methods less effective (Yue et al., 2021). Another major obstacle is the computing difficulty involved in processing massive datasets to produce recommendations in real time. Resolving these issues is essential to enhancing the improving recommendation systems.

Smartwatches and other wearables devices have become popular tools for health monitoring, capable of collecting continuous and rich health data. However, many current health intervention systems do not effectively utilize this data to provide, personalized recommendations. This underutilization of continuous health data limits the potential benefits of personalized health interventions and underscores the need for systems that can effectively integrate and analyse this data. Medical history, body composition, and metabolic profiles are examples of pathological data that offer comprehensive insights into a person's state of health. But a lot of the current systems don't incorporate this important data, which makes the health advice less useful.

Using pathology data in conjunction with continuous health monitoring, this research seeks to close significant gaps in the present health intervention systems. This research sets the stage for further development in individualized health management while simultaneously addressing a critical public health issue. This research hold promising substantially influencing public health by endowing people with the ability to make knowledgeable choices and take proactive measures towards improved health.

1.4 Purpose of the Research

The goal of this research is to create and assess a web application for a personalized fitness assistant that is intended specifically for obese and sedentary people. The purpose of this research is to deliver personalized nutrition and exercise recommendations utilizing cutting

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

edge technology. This research aims to overcome the drawbacks of conventional recommendation algorithms and fill significant gaps in current health intervention systems by merging pathological information with data from continuous health monitoring.

1.5 Research Aims

This research aims to design a user-friendly web application that is both easy to navigate and accessible to all users, regardless of their abilities. It focuses on integrating essential accessibility features to ensure inclusivity and enhance user engagement. The application will be tailored to effectively encourage, and support increased physical activity among obese and sedentary individuals. By identifying and implementing features that resonate with users, the research seeks to create a platform that motivates consistent physical activity. The effectiveness of the application in achieving these goals will be rigorously assessed, with a focus on its impact on user's activity levels and overall health outcomes.

1.6 Research Questions

This research aims to answer the following questions:

- I. How can the user interface be optimized for ease of use and accessibility?
- II. What accessibility features are necessary to ensure inclusivity for all potential users?
- III. How user friendly is the web application for individuals with varying levels of tech-savviness?
- IV. What features of the application are most effective in engaging and retaining users?
- V. How effective is the recommendation fitness assistant web application in improving physical activity levels among obese and sedentary individuals?

1.7 Research Objectives

- I. Develop and implement a user-friendly interface for the personalized fitness web application tailored to the needs of obese and sedentary individuals.
- II. Ensure the web application adheres to current web accessibility standards to provide an inclusive experience for users with varying needs.
- III. Integrate machine learning algorithms to provide personalized fitness and dietary recommendations based on user input and historical data.
- IV. Analyse user engagement data to identify which features are most frequently used and highly rated by users.
- V. Use self-reported and objective measures to access changes in physical activity.

1.8 Relevance and Importance of the study

The significance of this study lies in its potential to revolutionized health management for obese and sedentary individuals through the development and implementation of sophisticated recommendation fitness assistant web application. Obesity and Sedentary lifestyles are leading contributors to various chronic diseases, including cardiovascular diseases, diabetes, and certain cancers. This research will not only mitigate these health risks,

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

it will also improve overall well-being as these innovations can be applied beyond health management to other domains requiring personalized recommendations which can inform future research and development effort aimed at enhancing personalized health interventions and other applications of recommender systems.

1.9 Scope of the Study

The scope of the study encompasses the primary focus on obese and sedentary individuals who can benefit significantly from personalized diet and exercise recommendations using advanced algorithms that enhances accuracy and relevance. This study will focus on developing an intuitive and accessible user interface. **This web accessibility standards, and continuously improving the interface based on user feedback.**

1.10 Summary

This chapter focuses on developing a personalized fitness web app to help tackle obesity and inactive lifestyles. It explains the shortcomings of traditional fitness plans, which often don't account for individual needs. By using machine learning and real-time health data, the research aims to create a user-friendly app that offers tailored fitness and diet recommendations. This approach is designed to boost physical activity and improve health outcomes for people who are obese or lead sedentary lives. The project has the potential to transform how we manage health through personalized solutions.

Chapter Two – Literature Review

2.1 Introduction

The increasing prevalence of obesity and sedentary lifestyles has become a critical public health concern globally. Despite extensive public health campaigns and initiatives promoting physical activity and healthy eating, rates of obesity and related chronic conditions such as cardiovascular disease, type II diabetes, and hypertension continue to rise. The World Health Organization has declared obesity one of the most significant health threats in modern times, particularly in Westernized nations where lifestyle changes have led to decreased physical activity and increased caloric intake.

This literature review aims to explore the development of a data-driven personalized fitness web application designed to cater to the unique needs of obese and sedentary individuals. This project aims to overcome the constraints of current systems and give highly customised health and fitness recommendations. The review will examine existing research on obesity, sedentary lifestyles, and personalized health interventions, highlighting the gaps and opportunities for innovation in this critical area of public health.

TABLE 2.1 AN OVERVIEW OF ACADEMIC PUBLICATIONS AND THEIR RELEVANCE TO RESEARCH

S/ N	Refere nces	Research Focus	Relevan ce to this thesis	Methodol ogy	Strengths	Limitations	Knowledge Contributio n
1	Singha nia and Reddy (2024)	Big Data Analytics and predictive modelling for chronic disease managem ent	4	Review and analysis of big data applicatio ns; identificati on of key technologi es and solutions	Comprehe nsive review of big data applicatio ns; identificati on of key technologi es and solutions	General focus on chronic diseases; limited focus on personalized fitness applications	Highlights the potential of big data in personalizin g health care and predictive analytics
2	Olmo, (2024)	Machine Learning and virtual gyms for healthcare	5	Case studies and implement ation of online gym systems	Practical applicatio n of machine learning in fitness; detailed	Focus on virtual gyms may not address broader aspects of personalised	Illustrates the practical implementa tion of machine learning in virtual gym

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

					case studies	fitness web applications	environments for personalized health
3	Jeremi ah Olawumi Arowoogun et al., (2024)	Predictive analytics in health and fitness personalization	4	Review and analytics of predictive analytics techniques	Comprehensive review; focus on predictive techniques	Broad focus on health and fitness; less on specific application	Reviews the role of predictive analytics in personalizing health and fitness interventions.
4	Mcgowan et al., (2024)	Persuasive system design (PSD) in mobile health (mhealth) apps.	5	Multiphasic experimental design, prototyping, expert review and iterative design.	Emphasis on personalization and diversity in user's needs; comprehensive methodology	Focus on persuasive design, less on direct health outcomes.	Provides insights into using persuasive design for personalized health interventions
5	Perum a et al., (2024)	Persuasive design techniques in mhealth apps	4	Survey-based approach with contrast mining	Detailed evaluation of persuasive design principles; user-centric design approach	Limited generalizability beyond mhealth apps; focus on user engagement over health outcomes.	Evaluates Persuasive design technique to enhance user engagement in health apps.
6	Shoba et al., (2024)	Dynamic cardiovascular rehabilitation for Personalized Exercise Plans Guided by Recurrent Neural Networks	4	Development and testing of a fitness recommendation system	High relevance to personalized fitness; user-centric approach.	Potential privacy concerns with health data usage; need for larger sample size.	Advances the understanding of user-centred design in fitness recommendation systems

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

		in the Cloud					
7	Hu et al., (2024)	Wearable Technology for personalized exercise recommendations	5	Development and testing of wearable technology applications	Focus on wearable technology; high relevance to personalized fitness	Technical focus; need for long-term user engagement data	Explores the application of wearable technology in providing personalized exercise recommendations
8	Polak et al., (2016b)	User engagement strategies in digital health	4	Systematic review	Broad review on engagement strategies; relevant to digital health applications.	Focus on engagement, less on fitness-specific outcomes.	Reviews user engagement strategies to enhance digital health interventions
9	Mustafa et al., (2021)	Dietary recommendations for women with gestational diabetes mellitus	4	Randomized controlled trials (RCTs) and systematic review	High-quality evidence from RCTs; relevant to chronic disease management	Focus on chronic disease; may not fully address personalized fitness applications	Emphasizes the effectiveness of mHealth interventions in managing chronic diseases through evidence-based research
10	Dirik,(2023)	Machine learning techniques for obesity detection	5	Use of machine learning algorithms such as Random Forest and SVM for obesity prediction	Application of advanced machine learning techniques ; high relevance to obesity detection	Limited focus on broader fitness applications; primarily technical evaluation.	Demonstrates the efficacy of machine learning in predicting obesity, highlighting the importance of technical

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

							models in health care
1 1	Thomas et al., (2024)	Transforming big data for AI applications in nutrition and obesity	4	Big data transformation techniques and AI and readiness	Focus on practical applications of AI in nutrition and obesity research	Technical focus on data transformation; less on direct health outcomes.	Provides insight into preparing large datasets for AI applications in health research
1 2	Szeto, Arnold and Maher (2024)	Wearable activity trackers for physical activity promotion in healthcare	4	Provides insights into the use of wearable activity for promoting physical activity, which can be applied in personalized fitness applications for obese and sedentary individuals	Delphi study to develop a checklist (WATCH) for implementing wearable activity trackers in healthcare settings	Implementation challenges due to varied settings and patient compliance	Offers a structured guide for integrating wearable activity trackers into healthcare, highlighting best practices and potential challenges
1 3	Yi et al., (2024)	Predicting gene expression related to obesity using Random Forest	5	Application of Random Forest machine learning algorithms	Advanced machine learning techniques ; focus on obesity.	Technical focus; limited broader fitness application	Provides a novel application of Random Forests in predicting obesity-related genetic expressions
1 4	Dergaa et al., (2024)	AI-driven methods to improve physical activity	4	Systematic review of AI applications in physical activity	Comprehensive review; focus on AI applications	General focus; less on personalized fitness applications	Summarizes AI-driven strategies to increase physical activity, providing,

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

				enhancement			providing a broad overview of AI applications
1 5	MacCarthy and Pazoki, (2024)	Machine Learning for personalized exercise plans	5	Review of machine learning techniques in exercise planning	Relevance to personalized fitness; detailed analysis of machine learning applications	Focus on technical aspects; need for more practical implementation on data.	Highlights machine learning's potential in customizing exercise plans to individuals needs
1 6	Ulfa et al., (2022)	Data-driven techniques for personalized nutrition and fitness	5	Development and evaluation of a data driven recommendation system	Integration of nutrition and fitness; data-driven approach	Potential data privacy concerns need for more diverse sample	Demonstrates the integration of nutrition and fitness recommendations through data-driven methods
1 7	Rodríguez et al., (2022)	Personalized exercise programs for obesity management	4	Clinical trials evaluating personalized exercise programs	Focus on obesity management empirical data from clinical trials	Clinical setting; may not generalize to broader populations	Provides empirical evidence on the effectiveness of personalized exercise programs in managing obesity
1 8	McConnell et al., (2018)	Mobile health technologies to promote physical activity	4	Review of mobile health (mhealth) application.	Comprehensive review; focus on technology application	Broad focus; less on personalized fitness application.	Reviews the use of mobile health technologies to enhance physical activity
1 9	Oyebo de et	Predictive analytics	4	Review and	Comprehensive	Broad focus on health	Reviews the role of

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

	al., (2023)	in health and fitness personalization		analysis of predictive analytics techniques	review; focus on predictive techniques	and fitness; less on specific applications	predictive analytics in personalizing health and fitness interventions
20	AI Ansari et al., (2023)	AI applications to enhance physical activity	4	Systematic review of AI applications in physical activity	Focus on AI; comprehensive review	General focus; less on personalised fitness.	Summarizes the sue of AI to boost physical activity, highlighting various applications
21	Busnat u et al., (2022)	Integration of wearable device with personalized health recommendations	5	Development and testing of integration frameworks	Focus on integration of wearable technology; practical application	Technical focus; need for long-term user engagement data	Examines the integration of wearable devices in delivering personalized health recommendations
22	Galekop, Uyl-de Groot and Redekop, (2024)	Personalized exercise and nutrition plans for obesity	5	Development and evaluation of personalized plans	Integration of exercise and nutrition; high relevance to obesity	Limited sample size; need for long term effective data	Combines exercise and nutrition plans tailored to individual needs to combat obesity
23	Husnain et al., (2024)	AI for developing personalized health interventions	4	Application of AI techniques in health intervention development	Focus on AI; comprehensive application analysis	General focus; less on fitness specific application	Discusses AI's role in creating personalized health interventions
24	Shokrzadeh et al., (2024)	Personalized dietary recommendations using	5	Development and testing of knowledge graph-	Focus on personalization advanced recommenda	Primarily focused on dietary recommenda	Demonstrates the use of knowledge graphs to enhance

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

		knowledg e graphs		based recommendation systems	dation techniques	tions; less on fitness	dietary recommendations
2 5	Joshitha et al., (2024)	Personalized fitness coaching using AI and wearable devices	5	Integration of AI and wearable devices for personalized fitness coaching	Focus on practical application; high relevance to personalized fitness.	Technical focus; need for long-term user engagement data.	Explores combination of AI and wearables for personalized fitness coaching
2 6	Jubair, (2024)	Machine Learning for personalized fitness plans	5	Development and testing of machine learning models for fitness plans	Advanced machine learning techniques : high relevance to personalized fitness	Technical focus limited long-term effectiveness data.	Highlights machine learning's potential in creating personalized fitness plans
2 7	Singh et al., (2024)	mHealth solutions for personalized health	4	Systematic review of mHealth solution	Comprehensive review; focus on personalized health	Broad focus; less on specific fitness applications.	Reviews mHealth solutions and their applications in personalized health
2 8	Retzlaff , (2024)	AI integration with personalized nutrition and fitness	5	Development and testing of AI-integrated programs	High relevance to personalized fitness; focus on integrations	Technical focus; need for broader application data	Explores the integration of AI with personalized nutrition and fitness programs
2 9	Jegan and Nimi, (2024)	Personalized exercise recommendations using wearable data	5	Development and testing of wearable data applications	Focus wear technology; high relevance personalized fitness	Need for long-term user engagement data technical focus	Examines how wearable data can be used to personalise exercise recommendations

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

3 0	Chioma Anthon ia Okolo et al., (2024)	Data-driven approaches to personalized health	4	Comprehensive review of data-driven health approaches	Broad review relevance to personalized health	General focus; less on specific fitness application	Provides a comprehensive review of data-driven health approaches
--------	---------------------------------------	---	---	---	---	---	--

2.2 Current Public Health Guidelines and Interventions

Global public health agencies have implemented guidelines to reduce the dangers connected with obesity and sedentary lifestyles. The World Health Organisation (WHO) recommends that people participate in a minimum of 150 minutes of moderate-intensity aerobic activity or 75 minutes of vigorous-intensive aerobic activity per week. Additionally, they should engage in muscle-strengthening activities on two or more days per week (WHO, 2020). The dietary guidelines prioritise a well-rounded diet that includes ample amounts of fruits, vegetables, whole grains and lean proteins, while restricting the consumption of harmful fats, added sugars, and processed foods (Henry & Frank, 2023).

The objective of public health efforts is to establish conditions that facilitate the adoption of healthy behaviors. Community programmes, interventions implemented in schools, corporate wellness programmes, and changes in policies are crucial approaches to encourage physical activity and promote healthy eating habits (Sallis & Glanz, 2009; Polak et al., 2016). Nevertheless, these programmes frequently encounter difficulties in resolving individual discrepancies and attaining long-lasting modifications in behavior.

2.3 Constraints of Conventional Diet and Fitness Programmes

Conventional nutrition and fitness programmes often employ a standardised strategy that fails to consider variations in genetics, age, fitness levels, and cultural preference among individuals. The absence of personalisation might result in demotivation and reduces rates of adherence (King, 2023). In addition, genetic strategies may lack long-term sustainability as they do not consider individual lives and preferences, posing challenges for individuals to continue healthy habits over time (Drew et al. 2024).

Studies suggest that tailored health treatment, which take into account individual requirements and situations, are more successful in encouraging changes in behaviours and enhancing health results (Papry et al., 2024). Customised strategies can boost motivation and compliance by offering individualised suggestions that correspond with users' objectives and preferences.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

2.4 Public Health Programmes Focusing on Obesity and Inactive Behaviors

Public health efforts seek to diminish obesity and sedentary behaviour by advocating for physical activity and supporting a nutritious diet. Community programmes provide individuals with chances to engage in physical activities and attend workshops focused on promoting good eating habits (Polak et al., 2016a). School-based interventions promote the adoption of nutritious eating habits and engagement in physical activity during school hours (Polak et al., 2016b). Workplace wellness programmes facilitate physical activity challenges and offer nutritious food choices (Prowse et al., 2023). Policy modifications such as implementing higher taxes on sugary beverages, providing financial assistance for the production of fruits and vegetables, and implementing urban planning strategies that encourage the development of pedestrian-friendly communities, establish favourable conditions for making healthier choices (Polak et al., 2016).

2.5 Difficulties in Personalised Health Recommendations

Creating individualised health recommendations poses numerous obstacles. The “cold start” problem occurs when systems have insufficient data on new users, resulting in challenges in delivering precise recommendations (Ozsoy et al., 2024). The efficacy of recommendation systems can be impeded by sparse data as they often depend on self-reported information that may be incorrect or spatial (Tiribelli & Calvaresi, 2024). The implementation of personalised health therapies is further complicated by privacy concerns associated with the collecting and utilisation of personal health data (Tiribelli & Calvaresi, 2024).

2.6 Addressing Obstacles in Tailored Health Recommendations

To address these difficulties, the integration of data from wearable devices, electronic health records, and environmental sensors can offer a thorough comprehension of an individual’s health. Advances analytics, which utilise machine learning algorithms, have the capability to analyse intricate datasets and identify trends to provide customised recommendations. By prioritising user involvement and integrating behaviour modification tactics, one can enhance motivation and promote adherence. Ensuring data security and integrity are important ethical issues that play a vital role in generating trust and promoting user involvement (Valentine et al., 2023)

2.7 Incorporating Wearable Technology and Smartwatch Data

Wearable technology, namely smartwatches, has a major impact on gathering consistent and detailed health information, including heart rate, sleep habits, blood pressure, glucose levels, and body mass index. By combining this data with pathological information, it becomes possible to create customised food and exercise suggestions. Research has shown that utilising data from smartwatches can be useful in customising health smartwatches can be useful in customising health interventions and enhancing health outcomes (Gaikward et al., 2024). Smartwatches and other portable gadgets offer handy and easily accessible way to monitor and improve cardiovascular fitness levels. They provide real-time feedback and personalised recommendations to help achieve fitness goals (Fitness Guide, 2024).

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

2.8 Utilising Knowledge Graphs for Providing Dietary Recommendations

Utilising knowledge graphs in the development of personalised nutrition advice systems has demonstrated encouraging outcomes. Knowledge graphs combine information from multiple domains and utilise collaborative filtering methods to improve the precision and variety of recommendations. This methodology tackles challenges such as the cold-start problem, computational complexity, and sparse data, thereby enhancing the efficacy of nutritional recommendations (Huang et al., 2024). Knowledge-based systems, while efficient, frequently necessitate human updates and may not easily adjust to new data, underscoring the necessity for more dynamic methodologies such as genetic algorithms (Mardiana & Baizal, 2024).

2.9 Predictive Machine Learning Models for Obesity Associated Gene Expression

Machine learning algorithms, such as Random Forest, have been used to forecast gene expressions associated with obesity. These models examine genetic and environmental factors to gain understanding of an individual's inclination towards obesity and possible remedies. The utilisation of machine learning in this situation showcases the possibility of creating focused and individualised health interventions (Smith et al., 2024). Furthermore, machine learning algorithms have demonstrated potential in developing personalised nutrition programmes by utilising dietary recommendation systems that consider individual health concerns (Rosli et al., 2020).

2.10 Strategies to Encourage Health Eating Habits is Education

Interventions in higher education settings have prioritised the promotion of nutritious eating habits among students. The primary objective of these programmes is to enhance knowledge and understanding of nutritious eating habits while offering practical resources to facilitate healthier food selection. Research has demonstrated that educational programmes have the ability to significantly impact eating behaviors and enhance nutritional knowledge (Jones et al., 2024).

2.11 Health and Wellness Systems

System that covers a wide range of areas and are thorough in their approach. Holistic health and wellness systems encompass a wide range of elements, including physical fitness, dietary planning, and health issues management. These systems offer a complete approach to individual health. These systems employ sophisticated algorithms to evaluate strength, tailored diets and provide advise customised and efficient health interventions by classifying people into certain health and fitness categories (Fitness Guide, 2024).

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

2.12 Dietary Guidelines for Specific Health Conditions

Genetic algorithms have been utilised to construct specialised dietary advice systems for health disorders, such as hypertension. These systems offer customised dietary ingredient combinations that are specifically designed to meet individual health requirements, hence ensuring efficient control of illness such as hypertension (Mardiana & Biazal, 2024). Genetic algorithms in food recommendation systems overcome the constraints of knowledge-based methods by adjusting to fresh input and offering flexible answers.

2.13 Examining Gender Disparities in Obesity and Health Interventions

Gender disparities have a substantial impact on the prevalence of obesity, body composition, and the efficacy of health interventions. Research has indicated that although men may exhibit higher rate of obesity, women generally possess a greater proportion of body fat. Moreover, gender has a role in the allocation of body fat, as men are more inclined to acquire central (android) obesity, while women are more susceptible to peripheral (gynoid) obesity (Muscogiuri et al., 2024). These disparities require customised health interventions that consider gender-specific characteristics. Differences in body composition can cause variations in the pharmacokinetics and pharmacodynamics of anti-obesity medications between genders. However, it is not typical to propose gender-specific dose changes. Moreover, there is a frequent lack of female participation in clinical trials focused on obesity, which emphasises the necessity for more comprehensive research (Muscogiuri et al., 2024).

This research emphasises the urgent requirement for tailored health treatments to tackle the increasing public health issues of obesity and sedentary lifestyles. The integration of cutting-edge technologies, such as wearable devices, knowledge graphs, and machine learning models, can significantly improve the efficacy of personalised diet and fitness advice. Ongoing research and development in this field show potential for enhancing public health outcomes and enabling individuals to live better lives.

2.13.1 Limitations of Traditional Diet and Fitness Plans:

One-Size-Fits-All Approach: traditional recommendations often take a “one-size-fits-all” approach, failing to account for individual differences in factors like genetics, age, fitness level, and cultural preferences(Author King, 2023). This can lead to discouragement and low adherence rates.

Lack of Personalization. Traditional plans may not address individual needs and challenges. People might struggle with specific dietary restrictions or dislike certain types of fitness routines(Drew et al., 2024).

Sustainability: generic plans might not be sustainable in the long term. They may not consider individual lifestyles, making it difficult for people to maintain healthy habits over time(Drew et al., 2024).

2.13.2 Public Health initiatives Targeting Obesity and Sedentary Lifestyle

Public health initiatives aim to create environments that support healthy behaviors and reduce risk factors for obesity and sedentary lifestyles (Sallis and Glanz, 2009). Some of the programs includes:

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Community Programs: these programs offer opportunities for physical activities like cooking classes, and healthy eating workshops (Polak et al., 2016b).

School-Based interventions: these programs may promote healthy eating habits by offering physical activities opportunities during the school day, and educate children about the importance of healthy living(Polak et al., 2016a).

Workshop Wellness Programs: these programs can encourage employees to participate in physical activity challenges, provide healthy food options in cafeterias and offer educational workshops on healthy living (Prowse1 et al., 2023).

Policy Changes: Policies like increased taxation on sugary drinks, subsidies for fruits and vegetables, and urban planning initiatives that promote walkable neighbourhoods can all contribute to a more supportive environment for healthy choices (Polak et al., 2016a).

While these initiatives offer valuable tools, they can be limited by factors such as funding, accessibility, individual motivation. Developing more personalized approaches, like those explored in this web application project, may hold promise for improving the effectiveness of public health efforts to combat obesity and sedentary behaviors.

2.13.3 Challenges in Personalized Health Recommendations

Personalized health recommendations strive to offer individuals customised advice and treatments that are specifically customised to their own health requirements and situations (Valentine et al., 2023). Nevertheless, attaining genuine personalisation entails numerous substantial obstacles:

1. Limitations of Traditional One-Size-Fits-All Approaches:
 - i. Limited effectiveness: generalised health advice may not yield desired results for all individuals. Variances in biology, lifestyles, and environment can have a substantial influence on the effectiveness of general suggestions (Bodhini et al., 2023).
 - ii. Lack of motivation: general suggestions frequently overlook personal preferences and motives, resulting in low compliance and a limited long-term effect (Biese et al., 2024).
 - iii. Inequities: generic recommendations may not adequately cater to the distinct requirements and obstacles encountered by various populations(Polak et al., 2016a).
2. Limitations of Existing Recommendation Systems:
 - i. Cold start problem: this problem arises when dealing with new users with inadequate health data, posing a significant barrier. Recommendation systems face challenges in delivering tailored recommendations without adequate data about the user(Ozsoy et al., n.d.).
 - ii. Sparse data: numerous health recommendation systems depend on self-reported data, which may be unreliable or lacking in detail. The absence of extensive data impedes the capacity to provide genuinely tailored advice(Tiribelli and Calvaresi, 2024).
 - iii. Privacy concerns: this problem arise when personal health data is collected and utilised. Some people may be reluctant to disclose confidential information,

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

which can reduce the impact of personalized suggestions(Tiribelli and Calvaresi, 2024).

3. The Need for Personalized, Context-Aware Health Interventions:

- i. Dynamic health needs: health needs and risks change over time. Personalised advice should be flexible enough to accommodate these changes and shifting situations(Rummery et al., 2023).
- ii. Context matters: lifestyle choices, social factors that determine health, and environmental exposures all have an impact on all individuals' well-being. To achieve true effectiveness, personalised recommendations must take into account certain contextual elements(Rummery et al., 2023).
- iii. Incorporating behaviour modification into the integration process: efficient health interventions necessitate surpassing the mere provision of information. They should aid and encourage individuals in order to encourage them to embrace and sustain healthy behaviors (Rummery et al., 2023).

2.14 Strategies for Overcoming Challenges:

To address the challenges identified in the literature, the following strategies are essential:

Enhanced data sources: the integration of data from wearable devices, electronic health records and environmental sensors can offer a more comprehensive understanding of an individual's health.

Advanced analytics: machine learning algorithms possess the capability to scrutinise intricate datasets and detect patterns that can be utilised to tailor recommendations.

Focus on user engagement: tailoring recommendations to individual's preferences and incorporating behaviors change strategies can increase motivation and adherence.

Ethical Considerations: ensuring the confidentiality, integrity, and accountability, of data is essential for establishing confidence and fostering user engagement.

The solution that this research aims to offers is a personalized fitness web application that utilises data to cater the needs of obese and sedentary persons. This application aims to overcome the constraints indicated in the literature research by implementing a data-driven strategy to customise the user experience. By gathering data on variables such as activity levels and potential genetic information taking to consideration ethical concerns and user agreement, the application will customise workout regimens and educational resources to suit the individual demands and risk profile of each user. This web application has the potential to greatly enhance the effectiveness of web-based fitness recommendations for obese and sedentary individuals by utilising personalisation and capitalising on the established benefits of physical activity. Ultimately, it can assist them in achieving their weight management health objectives.

2.15 Summary

This chapter reviews the global health issue of obesity and sedentary lifestyles, linked to chronic diseases like cardiovascular disease and type II diabetes. Despite public health initiatives, obesity rates continue to rise, highlighting the need for personalized health interventions. The literature emphasizes the limitations of generic fitness programs and explored the use of machine learning, wearables, and knowledge graphs for tailored

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

recommendations. Challenges such as data privacy, the “cold start” problem, and user engagement are discussed. The chapter advocates for personalized, data-driven solutions to improve health outcomes, proposing a web application to address the needs of obese and sedentary individuals.

Chapter 3 – Methodology

3.1 Introduction

The research technique entails a methodological process of creating a data-oriented personalised fitness online application for obese and inactive individuals utilising the Python Django framework. This involves the integration of both frontend and backend development, along with the incorporation of machine learning models to provide personalised suggestions. The method involves multiple stages, including data gathering, pre-processing, model evaluation, and web application development. Every stage of the process guarantees that the application is resilient, precise, and easy to use, specifically designed to cater to the fitness requirements of obese and sedentary individuals.

The key phases in this methodology are:

- I. Data Collection: collect user data on height and weight. This data is essential for the development of customised fitness and food plans.
- II. Data Pre-processing: cleanse and convert the gathered data to ensure its suitability for analysis and modelling. This encompasses the tasks of managing missing values, identifying outliers, normalising data, and dividing the data into subsets.
- III. Feature Engineering: determine and generate pertinent features from the unprocessed data to be utilised in the machine learning algorithms. This may involve computing Body Mass Index (BMI), classifying individuals according to their BMI class.
- IV. Model Selection and Training: utilize a range of machine learning models to train and predict BMI, enabling the provision of tailored fitness and nutrition recommendations. The utilised models comprise Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting.
- V. Model Evaluation: assess the performance of each model by utilising metrics such as Mean Squared Error (MSE), accuracy, precision, and recall to determine the most suitable model for deployment.
- VI. Web Application Development: constructing the user interface and underlying functionality of the web application utilising Python Django.
- VII. Integration of Machine Learning Models: incorporating the created machine learning models into the web application.
- VIII. Testing and validation: conducting tests on the web application to ensure its proper functionality and verifying the precision of the machine learning models.
- IX. Deployment and Maintenance: installing the web application on a server and establishing maintenance procedures.

This methodology combines the principles of data science and web development to construct a customised fitness wellness. This chapter offers a comprehensive elucidation of each stage encompassed in the development process. Guaranteeing that the product is sturdy, easy to use, and capable of providing customised fitness advised derives from user data.

3.2 Research Philosophy

The research philosophy for this study on pragmatism, which incorporates aspects of both positivism and interpretivism. The aim is to investigate the advancement of personalised fitness applications. The methodological rigour in data collection, preprocessing, and model

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

evaluation is supported by positivism, which places importance on empirical evidence and quantitative data. An example of this is illustrated by research conducted by Oyebode, which showcases the tangible application of machine learning in the realm of fitness (Oyebode et al., 2023).

Conversely, interpretivism is also adopted to comprehend the subjective experience contextual intricacies of fitness application users. This viewpoint recognises that individual behaviors, preferences, and interactions with technology with technology have an impact on fitness and health, as emphasised in the research conducted by (Kuru, 2024). The research highlights the significance of behaviour change strategies in fitness applications.

The study utilises a quantitative analysis of model performance measures, complemented by qualitative insights from user feedback. This hybrid methodology guarantees a thorough assessment of the application's efficacy, in accordance with the pragmatic philosophy that places importance on both empirical evidence and user experience.

3.3 Proposed Workflow

The proposed workflow for developing the personalized fitness web application is structured into several key stages, illustrated in Figure 3.3. This workflow encompasses data collection, preprocessing, model development, web application development using Django, integration of machine learning models, and continuous evaluation and improvement.

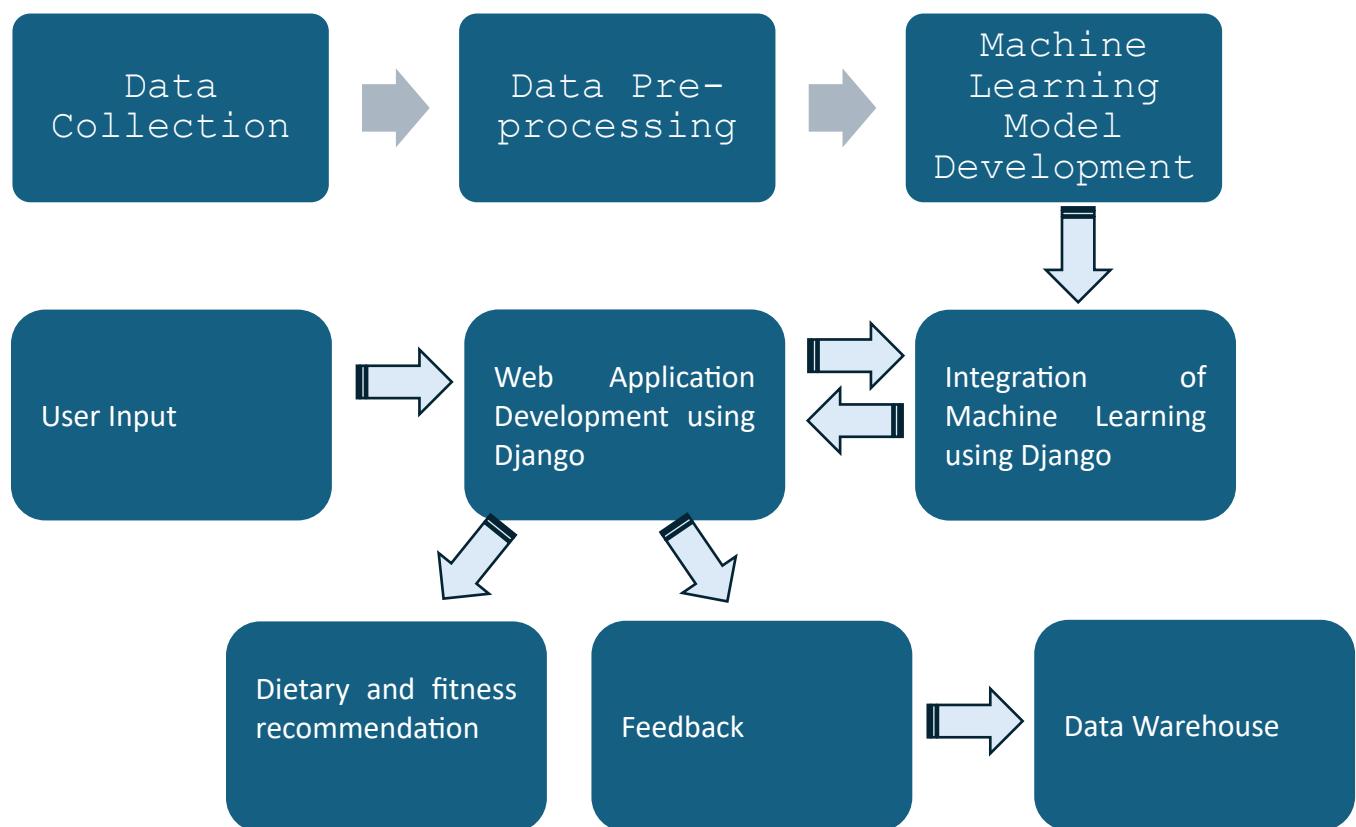


FIGURE 3.3: PROPOSED WORKFLOW FOR THE WEB APPLICATION USING DJANGO

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

3.3.1 Data Collection

The data collection process for developing the personalized fitness web application involved both primary and secondary data sources. The primary data was gathered through the Django web application form, while the secondary data was sources from the American Time Use Survey (ATUS) 2022 Eating Health Module.

Primary Data Collection: the primary data was collected directly from users through the website Django form designed to capture essential information such as height and weight which will be used to generate the BMI of the individual. This data is critical for personalizing fitness and dietary recommendations. The form was designed to be user-friendly, ensuring, ensuring high response rates and accuracy.

Secondary Data Collection: The secondary data was obtained from the ATUS 2022 Eating & Health Module. This dataset, sponsored by the U.S. Department of Agriculture's Economic Research Service and conducted by the U.S. Census Bureau, provides comprehensive information on various aspects of individuals' eating habits, health status, and physical activity.

Why ATUS Data was Selected:

The ATUS data was chosen for its detailed and reliable information on American's time use, especially regarding eating and health-related activities. This dataset includes variable such as BMI, general health status, exercise frequency, and dietary habits, which are directly relevant for obese and sedentary individuals. The data's breadth and depth allow for robust machine learning model development and validation.

ATUS Data Insights:

The ATUS dataset includes two primary files relevant to this research: the EH Respondent file and EH Activity file. The EH Respondent file contains case-specific variables like BMI, health status, and statistical weights for generating representative estimates. The EH Activity file includes detailed records of daily activities, secondary eating occurrences, and durations.

Figure 3.3.1 provides a snapshot of the dataset utilized for developing the personalized fitness web application. This dataset includes various attributes such as Body Mass Index (BMI), exercise frequency, physical activity participation, dietary habits. The figure highlights key variables such as:

EUHGT (Height in inches): This variable captures the height of respondents in inches. Height is a vital input for calculating BMI and understanding its relationship with weight and overall health status.

EUWGT (Weight in pounds): This variable records the weight of respondents in pounds. Alongside height, weight is a fundamental measure used to calculate BMI and assess individuals' health status.

ERBMII (Body Mass Index): Calculated from respondents' height (eught) and weight (euwgt) providing a crucial metric for personalizing fitness plans

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

EUEXERCISE (Physical Activity Participation): records whether respondents engaged in physical activities or exercises in the past week.

EUEXFREQ (Exercise Frequency): indicates how often respondents participated in physical activities over the past week.

ERDIET (Diet Quality): Self-reporting quality of diet, ranging from excellent to poor.

The ATUS data provides a robust foundation for developing personalized fitness recommendations. By analysing patterns in physical activity, dietary habits, and health metrics, the research can identify key factors influencing obesity and sedentary lifestyles. This data-driven approach ensures the recommendations are tailored to individual needs, improving adherence and health outcomes. The integration of primary data from the web application form and secondary data from the ATUS ensures a comprehensive understanding of user behaviors and health metrics. This dual approach enhances the accuracy and personalization of the fitness recommendations provided by the web application.

tucaseid	tulineno	eeincome1	erbm1	etwgt	eudietsoda	eudrink	eueat	eueexercise	euexfreq	eugenhtn	eugroshp	euheight	euwgt	euwic	exincome1
2.014E+13	1	-2	33.200001	0	-1	2	1	2	-1	1	1	60	170	1	2
2.014E+13	1	1	22.700001	0	-1	2	1	2	-1	2	3	63	128	2	0
2.014E+13	1	2	49.400002	0	-1	1	2	2	-1	5	2	62	270	2	12
2.014E+13	1	-2	-1	-1	2	1	2	2	-1	2	1	64	-2	2	2
2.014E+13	1	2	31	0	-1	1	2	1	5	4	1	69	210	1	0
2.014E+13	1	1	30.700001	0	1	1	1	1	2	3	2	71	220	2	0
2.014E+13	1	1	33.299999	0	-1	1	1	2	-1	2	3	65	200	2	0
2.014E+13	1	1	27.5	0	-1	2	1	1	3	2	1	63	155	-1	0
2.014E+13	1	1	25.799999	0	-1	2	1	1	6	3	1	70	180	-1	0
2.014E+13	1	1	28.299999	0	2	1	2	2	-1	1	1	65	170	-1	0
2.014E+13	1	1	40.5	0	1	1	1	2	-1	4	1	70	282	-1	0
2.014E+13	1	2	28	0	-1	1	1	2	-1	3	1	56	125	1	0
2.014E+13	1	1	27.9	0	-1	1	1	2	-1	3	2	77	235	2	0
2.014E+13	1	2	30.4	0	1	1	1	2	-1	3	2	68	200	-1	0
2.014E+13	1	2	26.799999	0	-1	2	1	2	-1	2	2	68	176	-1	0
2.014E+13	1	1	32.900002	0	-1	2	2	2	-1	3	2	65	198	2	0
2.014E+13	1	3	25.799999	0	-1	1	1	1	7	2	1	60	132	-1	0
2.014E+13	1	1	26.5	0	-1	2	1	1	4	2	1	66	164	-1	0
2.014E+13	1	-2	35.200001	0	2	1	1	2	-1	2	2	64	205	2	2
2.014E+13	1	1	-1	-1	-1	1	1	1	4	4	3	65	-3	2	0
2.014E+13	1	1	25.299999	0	2	1	1	2	-1	2	2	70	176	-1	0
2.014E+13	1	1	33.400002	0	-1	1	2	2	-1	3	3	58	160	-1	0
2.014E+13	1	-3	25.5	0	-1	2	2	2	-1	2	1	61	135	-1	3
2.014E+13	1	1	30	0	-1	1	1	2	-1	5	1	65	180	-1	0
2.014E+13	1	1	34.299999	0	2	1	1	2	-1	4	2	73	260	2	0
2.014F+13	1	1	29.1	0	-1	2	1	2	-1	3	1	62	159	2	0

FIGURE 3.3.1: SNAPSHOT OF THE DATA

3.3.2 Data Preprocessing

Data preprocessing refers to the process of transforming raw data into a format that is suitable for analysis. Data preparation is essential to ensure the quality and reliability of the collected data. This stage involves the process of data cleansing, standardisation, and selecting pertinent attributes. Missing values are efficiently handled, and data is modified to adhere to the requirements of the machine learning models. Essential preprocessing processes encompass:

Data cleaning: As highlighted by Madden, ensuring the quality and reliability of the data is crucial (Madden et al., 2020). Duplicate entries were removed, and filters were applied to exclude implausible values, such as unrealistic heights or weights, ensuring the dataset was both accurate and representative.

Data transformation: is the process of normalising and standardising data to maintain consistency throughout the dataset.

Feature Engineering: involves the creation of new features using current data to enhance the performance of a model. This can include the calculation of the equation(iii) using equation

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

(i), equation (ii), and additional features to enhance the model's predictive power (Madden et al., 2020)

$$\text{Height (in inches)} = \text{height (in feet)} * 12 \quad (i)$$

$$\text{Weight (in pounds)} = \text{weight (in kg)} * 2.20462 \quad (ii)$$

$$\text{BMI} = \frac{\text{Weight (in pounds)}}{\text{Height (in inches)}^2} \quad (iii)$$

Outlier Analysis: Statistical test, such as Z-scores, will be utilized to identify and handle outliers in the dataset. Outliers can significantly impact model performance especially in linear models and need careful consideration to either be excluded or transformed.

Normalization and Standardization: Following the recommendations from Oreskovic, Li and Erwin, (2023), the height and weight data can vary greatly, normalization or standardization techniques will be applied. This ensures that features contribute equally to the analysis and modelling stages.

3.3.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is performed to understand the distribution of the data and the relationships between different features. It is also used for guiding subsequent modelling and statistical analysis steps: This includes:

Descriptive Statistics: Descriptive statistical measures such as mean, median, standard deviation, and skewness will be computed for all key variables. These statistics provide insights into the central tendency and variability of the data.

Correlation Analysis: A correlation matrix will be generated to assess the liner relationships between features. Pearson's correlation coefficient will be computed, with significance levels determined to evaluate whether the correlations are statistically significant. This helps in identifying multicollinearity, which may affect regression model performance.

Inferential Statistics: inferential statistics, such as ANOVA, may be used to compare means between groups, for example BMI categories. This will help to understand if there are significant differences between groups of individuals in the dataset. Various charts are used to visualize this information:

- Distribution of BMI: Figure 3.3.3.1 presents a histogram of the Body Mass Index (BMI) distribution within the dataset used for developing the personalized fitness web application. The distribution appears to follow a normal distribution with a peak around the average BMI value. This distribution indicates that most individuals have a BMI within a specific range, highlighting the prevalence of overweight and obesity among the participants. Such insights are crucial for tailoring personalized fitness recommendations that address the specific needs of obese and sedentary populations (Madden et al., 2020).

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

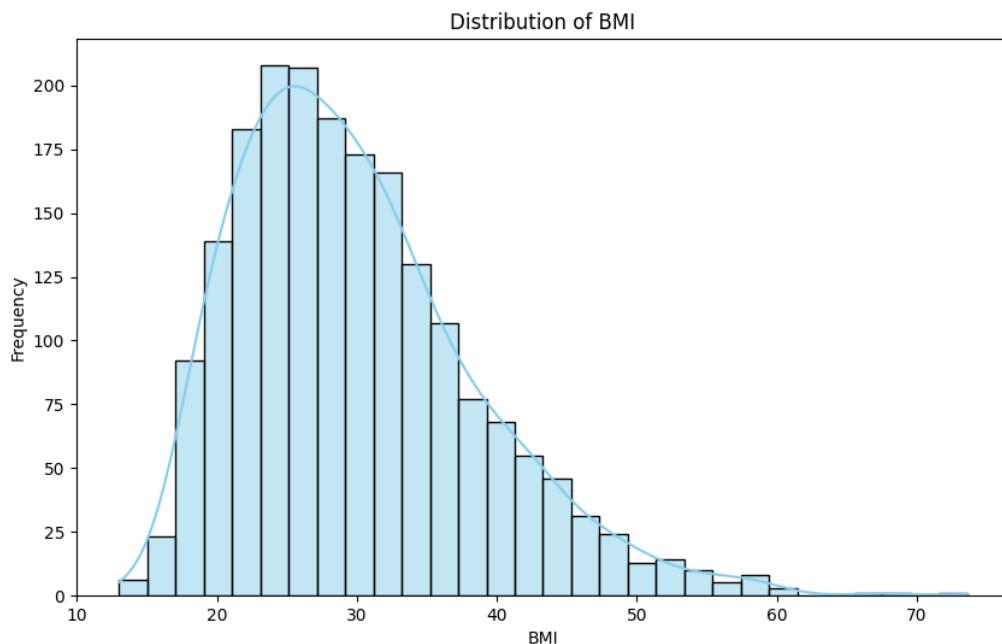


FIGURE 3.3.3.1: DISTRIBUTION OF BMI

- Feature Correlation Matrix: Figure 3.3.3.2 illustrates the feature correlation matrix, which shows the relationships between various features within the dataset. The matrix reveals that BMI (erbmi) has a strong positive correlation with weight (eugwt) and a slight negative correlation with height (euhgt). This pattern indicates that as weight increases, BMI tends to increase, while taller individuals may have slightly lower BMI values, assuming constant weight. Understanding these correlations is crucial for accurately predicting BMI and providing personalized fitness recommendations based on these relationships.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

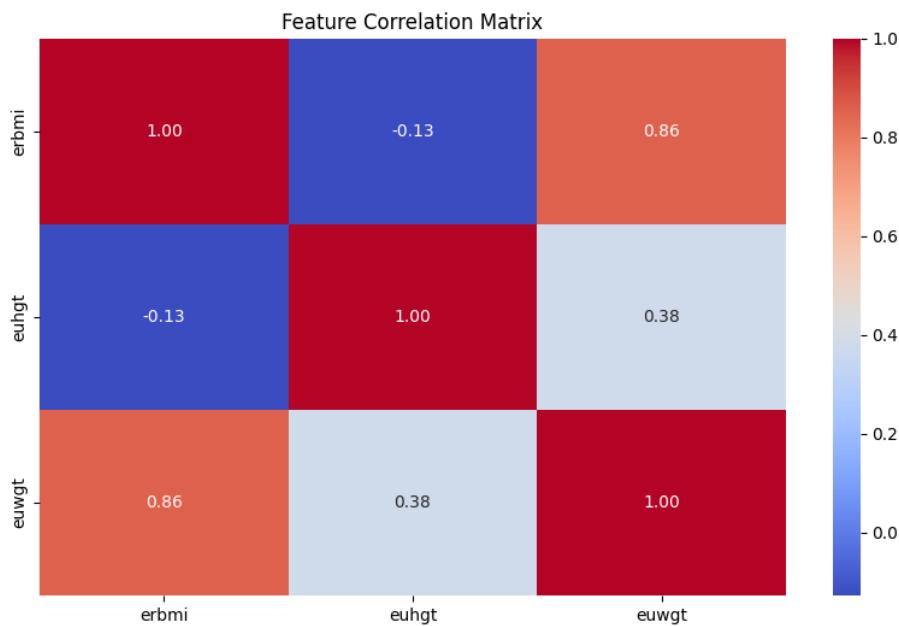


FIGURE 3.3.3.2: FEATURE CORRELATION MATRIX

- **Height vs BMI:** Figure 3.3.3.3 depicts a scatter plot illustrating the relationship between height (in inches) and BMI. The plot reveals a slight negative trend, suggesting that taller individuals tend to have a lower BMI when weight is held constant. This visualization is useful for identifying any outliers and evaluating the linearity of the relationship between height and BMI, which is important for refining the accuracy of the fitness recommendations provided by the web application.

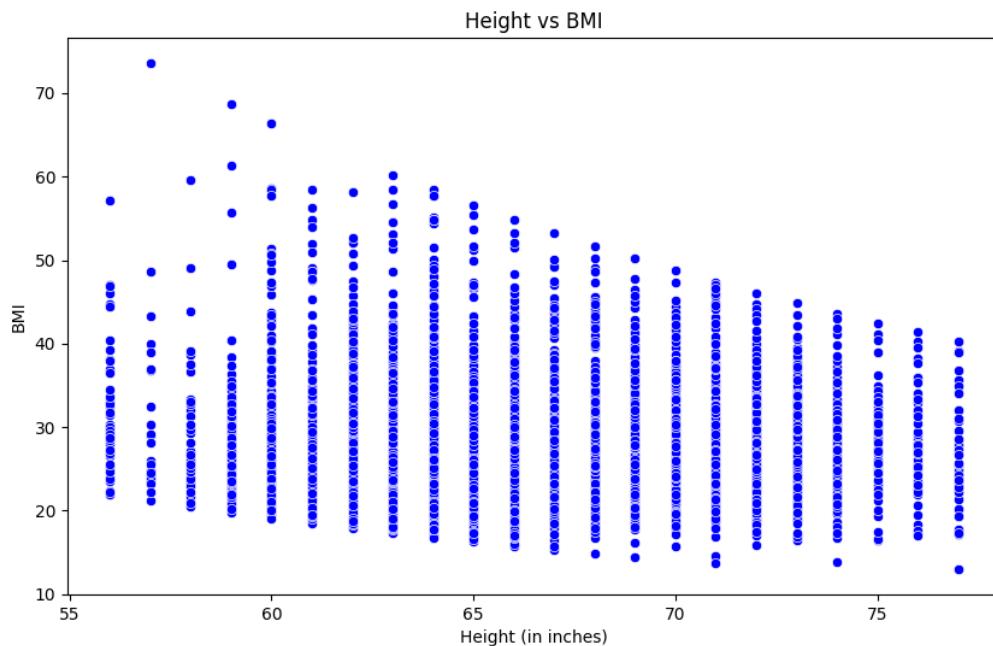


FIGURE 3.3.2.3: HEIGHT VS BMI

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- **Joint Plot of Height and Weight Coloured by BMI:** Figure 3.3.3.4 displays a joint plot that visualizes the relationship between height (in inches) and weight (in pounds), with points coloured according to BMI categories. The plot reveals distinct clusters of individuals with similar values, offering insights into how height and weight jointly influence BMI. This visualization is particularly useful for identifying patterns and trends in the dataset, aiding in the development of more accurate and personalized fitness recommendations.

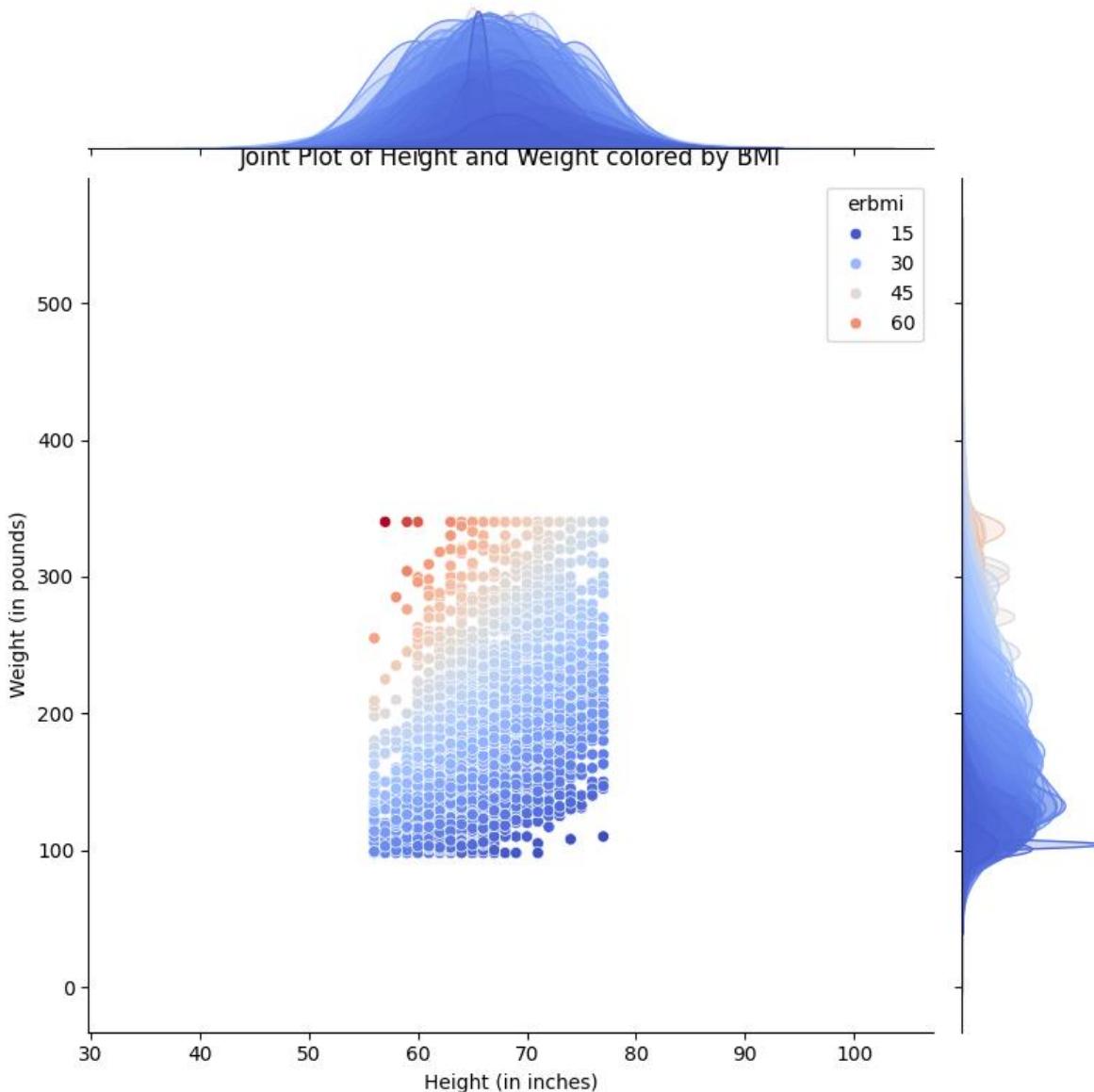


FIGURE 3.3.2.4: JOINT PLOT OF HEIGHT AND WEIGHT COLOURED BY BMI

- **Pairwise Relationship:** Figure 3.3.3.5 presents a pair-plot that displays pairwise relationships between key features in the dataset, including height (euhgt), weight (euwgt), and BMI (erbmi). Each scatterplot within the pair-plot helps in identifying pattern, correlations, and potential nonlinear relationship between these variables. This comprehensive visualization is crucial for understanding the complex interactions in the data, which in turn informs the development of more precise and personalized fitness recommendations.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

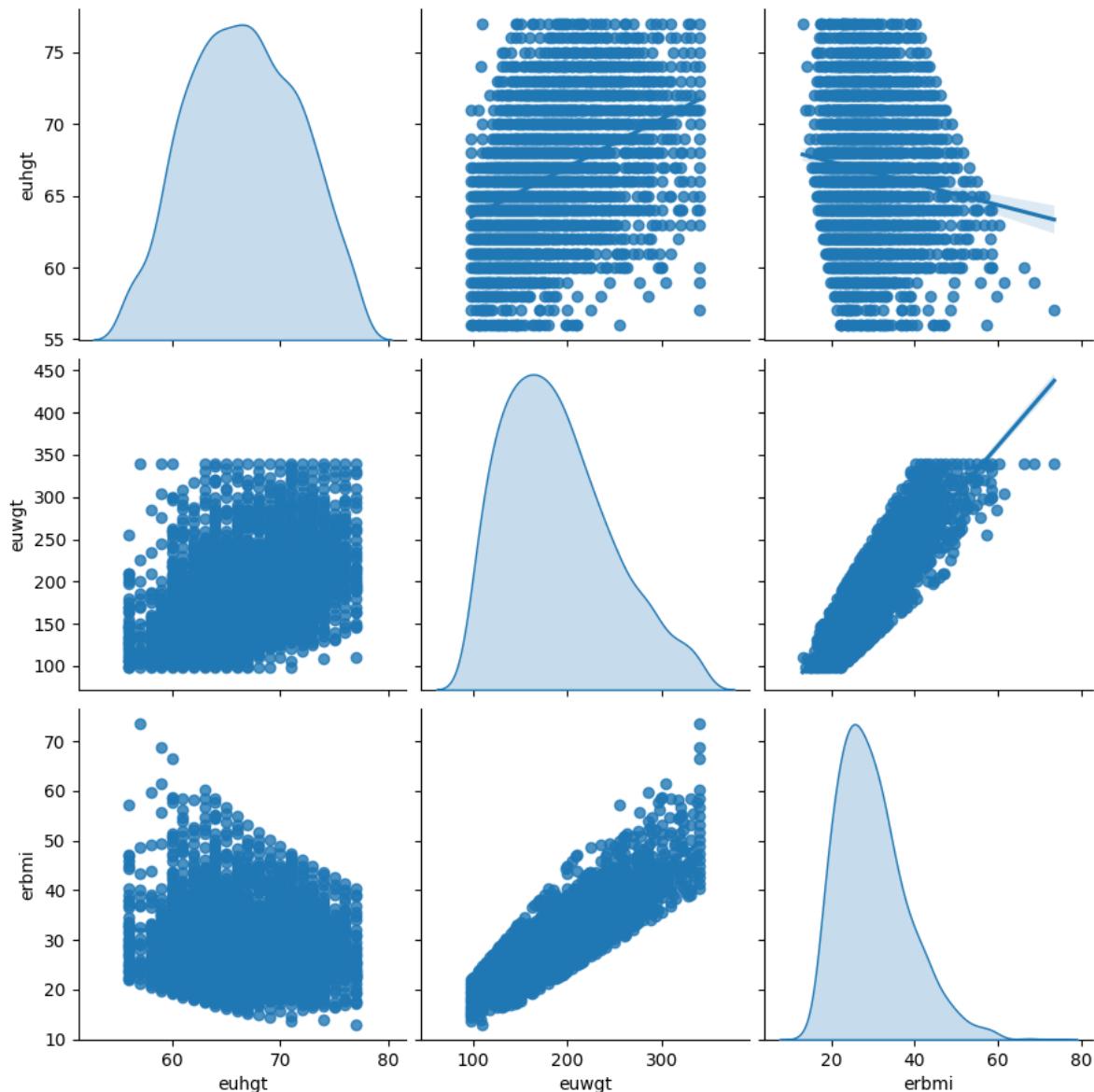


FIGURE 3.3.2.5: PAIRWISE RELATIONSHIP

- **Weight vs BMI:** Figure 3.3.3.6 shows a scatter plot that illustrates the relationship between weight (in pounds) and BMI. The plot demonstrates a strong positive linear relationship, where an increase in weight is directly associated with an increase in BMI. This is consistent with the BMI calculation formula, emphasizing the significant

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

influence that weight has on BMI.

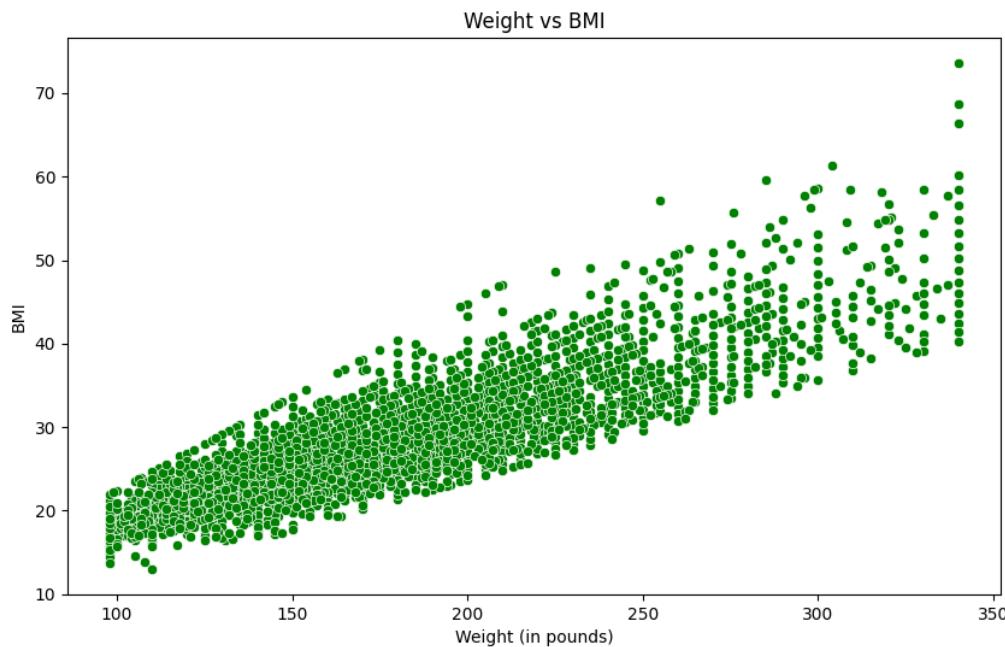


FIGURE 3.3.3.6: WEIGHT VS BMI

Data splitting: The dataset is split into training and test sets to evaluate the performance of the machine learning models. The training set is used to train the models, while the test set assesses the models' generalization to new, unseen data.

3.3.4. Machine Learning Model Development

The process of developing a machine learning model include training and assessing different algorithms to forecast BMI and classify individuals according to their health measurements. The following models are assessed to select the most efficient one. The process entails the following steps:

Model Selection involves the process of selecting the most suitable machine learning models, such as Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting.

Model Training: the process of training the chosen models on the prepared data, adjusting hyperparameters, and utilising cross-validation to enhance performance.

Model Evaluation involves the assessment of model performance using measures such as:

- **Model Evaluation Metrics:** beyond Mean Squared Error (MSE), other statistical metrics such as Adjusted R square, Mean Absolute Percentage Error (MAPE) and Akaike Information Criterion (AIC) will be calculated for a more comprehensive evaluation of model fit.
- **Residual Analysis:** Residual plots will be analysed to check the assumptions of regression models, such as homoscedasticity and normality of residuals. This helps in validating whether the chosen model is appropriate for the data.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- **Post-Hoc Test:** if ANOVA reveals significant differences between groups, post-hoc tests like Tukey's HSD will be added to identify which specific group differ from each other.

Validation Techniques: cross-validation will be employed to assess the generalizability of the models. Techniques such as k-fold cross-validation help in ensuring that the model's performance is not dependent on a specific subset of the data.

Hyperparameter Tuning: To optimize the model's performance, hyperparameter tuning was conducted using grid search, as recommended by Zanovich, 2024. This step ensured that the models were finely tuned to provide the most accurate predictions.

3.3.5 Web Application Development using Django

In a study conducted by Bolatbekov, the benefits of utilising Django for the development of intricate, data-driven application that necessitate strong back-end support and smooth interaction with machine learning models were emphasised. The study showcased Django's extensive set of capabilities and robust security features, making it an optimal option for designing healthcare apps that effectively and safely manage sensitive data(Bolatbekov et al., 2024).

Zanovich compared Flask, Django and Spring Boot in the context of machine learning projects. It is observed that Django's "batteries-included" approach offers a notable benefit for projects that demand a wide range of pre-existing functionality. Although Flask is lightweight and flexible, it may necessitate further configuration and integration of third-party tools to match the pre-existing capabilities offered by Django (Zanovich, 2024).

A separate study carried out by a group from Lviv Polytechnic National University highlighted the simplicity of incorporating machine learning models into Django, due to its strong Object-Relation Mapping (ORM) and pre-installed administrative tools. Django is well-suited for application that involve intricate data interactions and user administration, as demonstrated in their comparative analysis of Flask, Django, and Spring Boot (Zanovich, 2024).

Django was chosen over Flask for its scalability and extensive range of features when it came to constructing enterprise-level application. Flask's simplistic design, although beneficial for smaller projects, may be inadequate for managing the intricacies of large-scale application where Django excels (Zanovich, 2024).

Django app is used to achieve the following:

1. Enhancing the User Interface:
 - Django's Templating Language: facilitates the segregation of design and logic, hence simplifying the creation and administration of a well-organised and user-friendly interface.
 - Front-End Integration: Enables the utilisation of CSS and JavaScript frameworks to further enhance the user interface.
2. Guaranteeing Accessibility:
 - Django's extensibility: enables the incorporation of accessibility plugins and packages that enhance accessibility.
3. Assessing the User-Friendliness:
 - Unit Testing: Django's versatile framework enables the integration of A/B testing and user feedback forms to systematically assess and enhance the user experience.
4. Identifying Captivating Characteristics:
 - Usage Analytics: Django enables the tracking of user behaviour and preferences providing valuable insights for making informed decisions on future development.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

5. Enhancing levels of physical activity:

- Tracking and Reporting: Django's powerful backend can track and report user activity data, allowing for the generation of comprehensive reports that assess the application's influence on user's physical activity.

3.3.6 Integration of Machine Learning using Django

The trained machine learning models are integrated into the Django application. This integration allows the web application to process user inputs, run predictions, and generate personalized recommendations dynamically. The models are loaded and utilized through Django views, ensuring seamless interaction between the backend and frontend.

3.3.7 Recommendation Engine Development

The primary goal of this project is to create a recommendation engine that offers customised fitness and nutrition guidance. The development process encompasses several key steps, from selecting the appropriate machine learning model to integrate the engine into a web application and ensuring its continuous improvement.

3.3.7.1 Model Selection

The first step in developing the recommendation engine is selecting the most suitable machine learning model. Various models were considered for this task, including:

- Linear Regression: A model that assumes a linear relationship between the input features and the target variable (BMI)
- Ridge Regression: A regularized linear version of linear regression that includes a penalty term to prevent overfitting.
- Lasso Regression: Another regularized linear model that can also perform feature selection by driving some coefficients to zero.
- Decision Tree: A non-linear model that splits the data into subsets based on the most significant features, making decisions at each node until a prediction is made.
- Random Forest: An ensemble model that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting.
- Support Vector Machine (SVM): A model that finds the optimal boundary (hyperplane) to separate different classes or predicts continuous values by minimizing error.
- Gradient Boosting: An ensemble technique that builds models sequentially where each new model attempts to correct the errors of the previous ones.

The models were evaluated based on their ability to accurately predict BMI, with the best-performing model being selected for integration into the recommendation engine.

3.3.7.2 Model Training

Once the model was selected, the next step involves training it using the available dataset. The data was split into training and testing sets to ensure that the model could generalize well to new, unseen data. The following steps were carried out:

1. Data Splitting: The dataset was divided into 80% training data and 20% test data.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

2. Model Application: Each machine learning algorithm was applied to the training data to learn the relationships between the input features (e.g., height and weight) and the target variable (BMI).

3.3.7.3 Model Evaluation

The trained models were evaluated using several performance metrics to determine which model provided the most accurate predictions. These metrics are crucial for understanding how well each model fits the data and how reliable its predictions are. The key performance metrics used in this evaluation include:

- R² (Coefficient of Determination): R² measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It indicates how the data points fit the model. An R² value closer to 1 signifies that the model explains a larger portion of the variable in the data, whereas an R² value closer to 0 indicates that the model fails to explain the variance. A higher R² value indicates better model performance, as it suggests that the model is capturing the underlying trends in the data more effectively.
- Mean Absolute Error (MAE): MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. A lower MAE indicates that the model's predictions are closer to the actual values on average, making it a more accurate model.
- Root Mean Squared Error (RMSE): RMSE is the square root of the average of squared difference between prediction and actual observation. It gives a relatively high weight to large error, making it particularly useful when large errors are especially undesirable. A lower RMSE indicates better model performance, as it means the model's predictions are closer to the actual values, particularly penalizing large errors.
- Mean Squared Error (MSE): Measures the average squared difference between the predicted and actual values. Since the errors are squared before they are averages, the MSE gives a relatively high weight to large errors. A lower MSE indicates better model performance, as it means the average squared difference between predicted and actual values is smaller.

Based on these metrics, the Gradient Boosting model was identified as the best performer. It demonstrated superior predictive accuracy and generalization ability, making it the optimal choice for the recommendation engine.

3.3.7.4 Integration with Django

After selecting the Gradient Boosting model as the best performer, it was integrated into the Django web application. This integration involved the following steps:

- i. User input collection via Django Forms: User inputs, specifically height and weight, was collected using Django forms. These forms were designed to be user-friendly and ensure accurate data entry. The collected data was then validated and prepared for processing by the recommendation engine.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- ii. **Model Serialization:** The trained model was serialized using a format such as ‘pickle’ to enable it to be loaded within the Django environment.
- iii. **Django Views:** The Django views were configured to handle user input collected through the forms, pass the input data to the model for prediction, and return the prediction to the user.
- iv. **Real-Time Recommendations:** The model was embedded into the recommendation system, allowed the web application to provide real-time, personalized fitness and dietary advice based on the user’s predicted BMI.

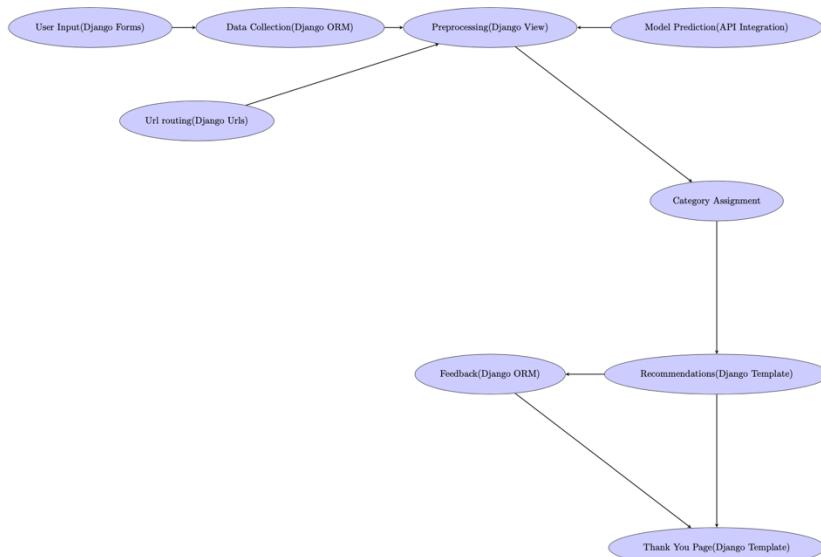


FIGURE 3.3.7.4: ARCHITECTURE OF THE WEB APP

Figure 3.3.7.4 shows the architecture of how user inputs are processed in the web application via Django forms and how it provides instant personalized recommendations in the form of tailored fitness and nutrition plans.

3.3.7.5 Continuous Improvement

To maintain the effectiveness and relevance of the recommendation engine, continuous improvement process was established. This process includes:

- i. **User Feedback:** Collecting feedback from users regarding the accuracy and usefulness of the recommendations provided.
- ii. **Periodic Retraining:** Retraining the model with new data as it becomes available to improve accuracy and adapt to changing trends.
- iii. **Algorithm Updates:** Enhancing the recommendation algorithms based on feedback and performance metrics to ensure the recommendations remain accurate and relevant.
- iv. **Monitoring and Optimization:** Regularly monitoring the system’s performance and making adjustments as needed to optimize the recommendation engine.

By implementing these steps, the recommendation engine can evolve over time, improving its precision and user satisfaction.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Algorithm 1: Personalized Fitness and Dietary Recommendation Engine

function: 'generatePersonalizedRecommendations(userData)'

output: 'Customized Fitness and Dietary Plan'

procedure:

i. Input:

- H_u : User height in feet. This variable is essential for calculating the BMI, which is a key indicator of the user's health status.
- W_u : User weight in kilograms. Like height, this variable is used to compute BMI. It provides a measure of the user's mass relative to their height.

ii. **BMI Calculation:** The user's Body Mass Index (BMI) is calculated using the formula BMI

$$\text{BMI}_u = \frac{W_u}{H_u^2} \quad (1)$$

This calculation provides a quick assessment of the user's body fat based on their height and weight.

iii. **BMI Categorization:** based on equation (1) value, the user is categorised into one of the four health categories:

$$\text{Category}_u = \begin{cases} \text{"Underweight"} & \text{if } \text{BMI} < 18.5 \\ \text{"Normal weight"} & \text{if } 18.5 \leq \text{BMI} < 24.9 \\ \text{"Overweight"} & \text{if } 24 \leq \text{BMI} < 29.9 \\ \text{"Obese"} & \text{if } \text{BMI} \geq 30 \end{cases}$$

This categorization allows the algorithm to adjust its recommendations based on the user's health status.

iv. **Term Frequency Calculation:** To understand the user's dietary preferences, the algorithm calculates the term frequency (TF_{D_f}) for each dietary element. This is done using the equation:

$$TF_{D_f} = \frac{\sum_{i \in R} S(u, d_i)}{\sum_{j=1}^M Q(j, k)} \quad (2)$$

Here, the numerator $\sum_{i \in R} S(u, d_i)$ represents the sum of the user's ratings for specific dietary elements. The denominator $\sum_{j=1}^M Q(j, k)$ represents the total number of rated dietary elements. This calculation helps in identifying which dietary elements are most important to the user based on their ratings.

v. **Inverse Document Frequency (IDF):** The algorithm then calculates the inverse document frequency (IDF_t) for each dietary element to understand how unique or common these elements are across all users. The equation used is:

$$IDF_t = \log \left(\frac{\sum_{j=1}^N \sum_{s=1}^e T(s, j)}{\sum_{j=1}^N T(t, j)} \right) \quad (3)$$

The numerator $\sum_{j=1}^N \sum_{s=1}^e T(s, j)$ calculates the total occurrences of all dietary and activity elements across the entire datasets. The denominator $\sum_{j=1}^N T(t, j)$ count how often a dietary element or activity t appears.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- vi. **Generating Recommendations:** Using the Equations (1,2,3), the algorithm generates personalized fitness and dietary recommendations. These recommendations are tailored to align with the user's health (as determined by BMI), preferences identified by term frequency), and uniqueness of their dietary choices (highlighted by inverse document frequency).

3.3.8 Evaluation and Testing

Ongoing assessment and examination are essential to guarantee the dependability and efficacy of the application. These testing methods encompass unit testing, integration testing, and user acceptability testing. User feedback is included to enhance the application's functionality and user experience.

User testing and feedback are vital components in the development of health-related application. Studies have emphasized the importance of involving end-users in the testing process to identify areas for improvement and ensure the application meets user needs ((Oreskovic et al., 2023), (Benitez et al., 2022))

3.4 Data Analysis Plan

Data analysis utilises descriptive and inferential statistical to assess the efficacy of machine learning models. In this section, we outlined the statistical analyses planned for the data collected and model developed:

- **Descriptive Analysis:** the data will first be explored using descriptive statistics to summarise the main characteristics of the dataset. This includes measures of central tendency (mean, median) and dispersion (standard deviation, variance).
- **T-Test and ANOVA:** these tests will be employed to compare means across different BMI categories, providing insights into the differences in physical activity levels, dietary habits, and other relevant factors between groups.
- **Post-Hoc Analysis:** following ANOVA, post-hoc test such as Tukey's HSD will be conducted to determine which specific groups differ from one another.

Correlation and Regression Analysis: Correlation Analysis will be conducted to identify the strength and direction of relationships between variables. Regression analysis will be used to predict BMI based on height, weight, and other features, with statistical significance tests (p-values) being used to confirm the relevance of predictors. These relationships are consistent with the findings of (Oreskovic et al., 2023) who emphasized the importance of considering these factors in personalized health interventions

Model Performance Metrics: in addition to traditional metrics like R squared, advanced metrics such as Mean Absolute Percentage Error (MAPE) and root mean square error (RMSE) will be calculated. Statistical test will be used to compare these metrics across different models to determine the most effective one.

Validation and Testing: Cross-validation and residual analysis will ensure that the models generalize well to unseen data, with statistical significance tests confirming the robustness of the results.

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

The use of performance metrics is a common practice in evaluating machine learning models. These tools provide valuable insights into the model's predictive capabilities and highlight areas for further refinement (Akther et al., 2024).

3.5. Machine Learning

Machine Learning utilises user data analysis to find patterns and create predictions, enabling the provision of personalised fitness advice in a dynamic manner. This section presents the machine learning models employed in this study.

3.5.1 Linear Regression

Linear Regression is a fundamental approach utilised to forecast continuous outcomes. It establishes a linear correlation between the input features and target variable. This model is uncomplicated and easily understandable, making it an excellent initial approach for estimating BMI.

3.5.2 Ridge Regression

Ridge Regression is a variant of Linear Regression that incorporates a regularisation component to mitigate the problem of overfitting. This model is particularly beneficial when addressing multicollinearity or when there is a high number of characteristics.

3.5.3 Lasso Regression

Lasso Regression is a kind of Linear Regression that incorporates regularisation and can assign some coefficients as zero, hence facilitating feature selection. This aids in streamlining the model and enhancing its comprehensibility.

3.5.4 Decision Tree Regressor

Decision Tree Regressor is a model that partitions the data into subsets based on the values of its features, allowing for non-linear relationships. It can capture subtle correlations between features and target variable, which makes it valuable for complex datasets.

3.5.5 Random Forest Regressor

Random Forest Regressor is a technique that utilises an ensemble of decision trees to enhance the accuracy of predictions and mitigate the problem of overfitting. The system is very resilient and capable of efficiently processing extensive datasets with complex structures.

3.5.6 Support Vector Machine

Support Vector Machine (SVM) is an effective regression approach that aims to identify a hyperplane in a high-dimensional space that optimally matches the input. It demonstrates

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

efficacy in environments with large number of dimensions and may be applied to both linear and non-linear regression.

3.5.7 Gradient Boosting

Gradient Boosting constructs models in a sequential manner, where each subsequent model rectifies the fault produced by its predecessors. This ensemble approach is renowned for its exceptional prediction accuracy and resilience.

The selection of machine learning models is based on their proven effectiveness in health-related applications. Ensemble methods, such as Random Forest and Gradient are particularly favoured for their ability to handle complex datasets and improve prediction accuracy (Hassan et al., 2024).

3.6 Rationale for Machine Learning Algorithm

Compute Time: Linear Regression and Ridge Regression are computationally efficient, making them suitable for real-time predictions. Decision Tree and Random Forest are also relatively fast due to their tree-based structure, although Random Forest can be more computationally intensive.

Ease of Deployment: All selected models are supported by scikit-learn, a widely used Python library for machine learning, which ensures easy integration into the Django Framework. The deployment process involves using Django views and APIs to call the trained models for predictions.

Performance: Gradient Boosting, while computationally more intensive, provides high prediction accuracy and is beneficial for improving the recommendation system's performance.

The rationale for algorithm selection is supported by research demonstrating the importance of balancing accuracy with computational efficiency in health applications. The use of scikit-learn further enhances the model's accessibility and ease of deployment (Sharma, Khan and Singh, 2024).

3.7 Ethical Considerations

The research complies with ethical requirements, guaranteeing the confidentiality and protection of user data. Data collection is conducted in a secure manner, with all gathered data being stored securely. Prior to collecting any data, user consent is obtained. The application is specifically developed to uphold user privacy and offer clear and comprehensive information regarding data use (Madden et al., 2020)

3.8 Summary

This chapter outlines the development process of a personalized fitness web application for obese and sedentary individuals using the Python Django framework. The methodology

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

involves stages such as data collection, preprocessing feature engineering, model training, evaluation, and web application development. Machine learning models like Linear Regression, Ridge Regression, Random Forest, and Gradient Boosting are utilised to predict BMI and provide personalized fitness recommendations. The chapter discusses the integration of these models into the Django application and emphasizes continuous testing and improvement through user feedback. Ethical considerations are also highlighted, focusing on data security, user privacy, and informed consent throughout the process.

Chapter 4 – Result and Discussion

4.1 Introduction

This chapter presents the results of the study on developing a data-driven personalized fitness web application aimed at assisting obese and sedentary individuals. The analysis includes the evaluation of various machine learning models used to predict BMI, the performance of the web application interface, and user engagement metrics. This chapter also interprets these findings in the context of the existing literature and discusses their implications for enhancing user experience and promoting physical activity through personalized fitness recommendations.

4.2 Data Pre-processing Results

The data pre-processing stage is critical in preparing the raw data for analysis. This section details the various preprocessing steps that were undertaken, including the removal of duplicates, handling of inconsistent data, dealing with missing values, outliers. The dataset utilized in this study contained parameters such as height, weight and BMI.

4.2.1 Data Cleaning

The initial phase of data preprocessing involved cleaning the dataset to ensure its quality and reliability for subsequent analysis. One of the key steps in this process was the identification and removal of duplication entries, which could introduce bias and distort the result of the analysis.

Removal of Duplicates

Upon loading the dataset, it was observed that the data consisted of 11,212 entries across three variables: 'erbmi' (BMI), 'euhgt' (height), and 'euwgt' (weight). An initial inspection revealed the presence of duplicate entries within this dataset, which could potentially skew the analysis results.

To address this, 'drop duplicates ()' function was employed to remove any duplicate rows. After this operation, the dataset was reduced to 2,073 unique entries. This significant reduction in the number of entries highlights the presence of a substantial number of duplicate records, which could have impacted the integrity of the analysis had they not been removed.

The cleaned dataset now contains only unique records, ensuring that each data point contributes distinct information to the analysis. This step was crucial in laying a solid foundation for the subsequent phases of data preprocessing and model development.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

4.2.2 Handling Inconsistent Data

In addition to removing duplicate entries, it is crucial to ensure that all data points are within a reasonable and expected range. Inconsistent data, such as entries with unrealistic height or weight values, can significantly impact the quality of the analysis and the performance of the machine learning models.

Removal of Inconsistent Data

The dataset was further refined by applying filters to remove entries with inconsistent or implausible value. Specifically, the dataset was filtered to exclude any records where:

- Height ('euhgt'): was less than or equal to 21 inches, as such values are not representative of human adult height.
- Weight ('euwgt') was less than or equal of 5 pounds, which similarly does not represent a feasible adult weight.

Before applying these filters, the dataset contained 2,073 entries. After removing entries with these implausible values, the dataset was reduced to 1,978 entries. This additional cleaning step ensures that the remaining data is both realistic and relevant for analysis.

This meticulous approach to data cleaning helps prevents the inclusion of outliers data points that could skew the results and lead to misleading conclusions. By focusing on realistic and consistent data, the analysis and subsequent modelling efforts are more likely to produce reliable and valid outcomes.

4.2.3 Handling Missing Values

A critical aspect of data preprocessing is handling missing values, as they can significantly impact the performance of machine learning models. Missing data can lead to biased estimates, reduced the representativeness of the sample, and make it challenging to perform accurate analyses.

Assessment of Missing Data

To evaluate the presence of missing values in the dataset, the 'isna().value_counts()' function was utilized. This function checks for 'NaN (Not a Number) values across all columns in the dataset.

The assessment revealed that there were no missing values in any of the columns ('erbmi', 'euhgt', 'euwgt'). Specifically, the output indicated that all 1,978 entries in the dataset were complete, with no missing data points. This result indicates that the dataset is robust and complete, eliminating the need for further imputation or removal of records due to missing values. The absence of missing data ensures that the dataset is well-suited for the subsequent stages of analysis and modelling, as all variables contain full information for every entry.

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

4.2.4 Outlier Detection and Handling

Outliers can have a significant impact on data analysis and the performance of machine learning models. However, in some cases, outliers may represent important variations in the data that are meaningful and relevant to the study. In this project, a retention approach was adopted for handling outliers.

Boxplot Analysis

Boxplots were generated for the key variables in the dataset, including 'erbmi' (BMI), 'euhgt' (height), and 'euwgt' (weight). The boxplot allowed for a visual inspection of potential outliers in these variables.

- BMI ('erbmi'): The boxplot for BMI revealed several outliers. These outliers are data points that fall outside the typical range expected for BMI values.
- Height ('euhgt'): The boxplot for height did not show any significant outliers. The height values appear to be distributed within the expected range for the population.
- Weight ('euwgt'): Similarly, the weight values showed no significant outliers, with all data points falling within the expected range.

Retention of Outliers

Given the nature of this study, which aims to develop personalized fitness recommendations, it was deemed important to retain the outliers identified in the BMI variable. The decision to retain outliers was based on the following considerations:

- Representation of Real-World Variability: the outliers in BMI may represent individuals with extreme body mass, who are precisely the target population for personalized fitness interventions. Removing these data points could lead to a model that is less effective for individuals with high BMI, thereby reducing the generalizability and applicability of the model.
- Model Robustness: retaining outliers allows the model to be exposed to a wider range of data points, including those at the extremes. This exposure can lead to the development of a more robust model that performs well across different segments of the population, including those with high BMI.
- Ethical Considerations: in healthcare-related research, excluding individuals with extreme health metrics, such as very high BMI, could be seen as unethical, as it could lead to a bias against individuals who may need the most help. By retaining these outliers, the study ensures that the resulting model is inconclusive and capable of providing recommendations to all individuals, regardless of their BMI.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

4.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial step in understanding the underlying patterns, trends, and relationships within the dataset. In this section, we perform a series of analyses to gain insights into the distribution of key variables, the relationships between them, and any potential correlations that may inform the development of the machine learning models.

4.3.1 Distribution of BMI

The first step in our EDA was to examine the distribution of BMI ('erbmi') within the dataset. Understanding the distribution of BMI is essential because it directly influences the design and outcomes of the personalized fitness recommendations.

A histogram was generated to visualize the distribution of BMI values across the dataset:

- **Findings:** The histogram shows the BMI values appear to follow a normal distribution, with a peak around the average BMI value. Most individuals in the dataset have a BMI within the range 20 to 40, indicating a prevalence of overweight and obesity in the population.

This distribution suggests that the dataset is representative of a population with varying degree of weight-related health issues, which is ideal for developing models that can cater to individuals with different BMI levels.

4.3.2 Feature Correlation Analysis

To understand the relationships between different features in the dataset, a correlation matrix was generated. This matrix helps identify the strength and direction of the linear relationships between variables such as height ('eheight'), weight ('eweight'), and BMI ('erbmi').

- **Findings:**
 - **BMI and weight:** A strong positive correlation was observed between BMI and weight, as expected. This correlation is consistent with the formula used to calculate BMI, where weight is directly proportional to BMI.
 - **BMI and Height:** A slight negative correlation was found between BMI and height. Taller individuals tend to have slightly lower BMI values, assuming constant weight, which aligns with the inverse relationship between height squared and BMI in the BMI formula.
 - **Height and Weight:** A moderate positive correlation was also observed between height and weight, indicating that taller individuals generally weigh more.

These correlations are essential for building predictive models, as they highlight which variables are most influential in determining BMI.

4.3.3 Relationship Between Height and BMI

To further explore the relationship between height and BMI, a scatter plot was generated:

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- Findings: The scatter plot reveals a slight negative trend, indicating that BMI tends to decrease as height increases, assuming weight remains constant. This observation aligns with the earlier finding of a negative correlation between height and BMI.

This visualization helps identify any outliers or non-linear relationships that may need to be accounted for in the modelling process.

4.3.4 Joint Distribution of Height and Weight

To better understand the joint distribution of height and weight and how they influence BMI, a joint plot was generated with joints coloured according to their BMI category:

- Findings: the joint plot shows distinct clusters of individuals with similar BMI values. The plot illustrates that individuals with higher weights and shorter heights tend to have higher BMI values, while those who are taller and lighter generally have lower BMI values.

This joint distribution provides valuable insights into how height and weight together contribute to BMI and can help in refining the model to provide more accurate fitness recommendations.

4.3.5 Pairwise Relationship Between Features

To examine the pairwise relationship between all key features ('euhgt', 'uewgt', 'erbmi'), a pairplot was created:

- Findings: the pairplot highlights the relationship between each pair of variables:
 - Height vs Weight: a positive linear relationship is observed, where taller individuals tend to weigh more.
 - Weight vs BMI: a strong positive linear relationship, consistent with the BMI calculation formula.
 - Height vs BMI: a negative relationship, reinforcing the earlier findings.

The pairplot also helps in identifying any potential non-linear relationships or clusters within the data, which can be explored further in model development process.

4.3.6 ANOVA and Post-Hoc Analysis

To assess the statistical significance of differences in BMI across different weight categories (Normal weight, Overweight, Obese, Underweight), an Analysis of Variance (ANOVA) test was performed. The ANOVA results were followed by Tukey HSD post-hoc test to identify group differences.

ANOVA Findings:

F-statistics: 922.45

P-value: 0.0

The ANOVA test showed a statistically significant difference in BMI across the different weight categories, indicating that BMI varies significantly depending on the weight category.

Post-Hoc Analysis (Tukey HSD):

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

The Tukey HSD test revealed significant differences between the weight categories, confirming that BMI values differ across categories. The results indicated that:

- Normal Weight vs Obese: Significance difference in BMI ($p < 0.05$).
- Normal Weight vs Overweight: Significance difference in BMI ($p < 0.05$)
- Obese vs Underweight: Significance difference in BMI ($p < 0.05$)

These statistical analyses provide strong evidence that BMI varies significantly across different weight categories, which supports the need for personalized fitness recommendations tailored to each BMI category.

4.4 Model Training and Validation

Model training and validation are critical steps in developing a robust and reliable predictive model. In this section, we discuss the processes involved in training various machine learning models using the pre-processed dataset, evaluating their performance, and selecting the best-performing model for the personalized fitness recommendation system.

4.4.1 Model Selection

The first step in the modelling process was to select appropriate machine learning algorithms that could effectively predict BMI and provide personalized fitness recommendations. Given the nature of the data and the problem at hand, the following models were chosen for evaluation:

- Linear Regression: a simple model that assumes a linear relationship between the input features and the target variable (BMI).
- Ridge Regression: a regularized linear model that helps to prevent overfitting by adding a penalty to the coefficients.
- Lasso Regression: Another regularized linear model that performs feature selection by shrinking some coefficients to zero.
- Decision Tree Regression: A non-linear model that splits the data into subsets based on the most significant features, making decisions at each node until a prediction is made.
- Random Forest Regression: An ensemble model that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting.
- Support Vector Regression (SVR): A model that attempts to find the best-fitting line within a margin of tolerance, using kernel functions to handle non-linear relationships.
- Gradient Boosting Regression: An ensemble model that builds trees sequentially, where each new tree tries to correct the errors made by the previous ones.

4.4.2 Data Splitting

To evaluate the performance of each model, the dataset was split into training and test sets. The training set, comprising 80% of the data, was used to train the models, while the remaining 20% was reserved for testing and validating.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

4.4.3 Model Training

Each selected model was trained using the training dataset. The models were fitted to the data, learning the relationships between the input features (height and weight) and the target variable (BMI).

4.4.4 Model Validation

After training the models, their performance was evaluated using the test dataset. Multiple key performance metric was used to assess the accuracy and reliability of each model. The results of the evaluation are summarized in the following table:

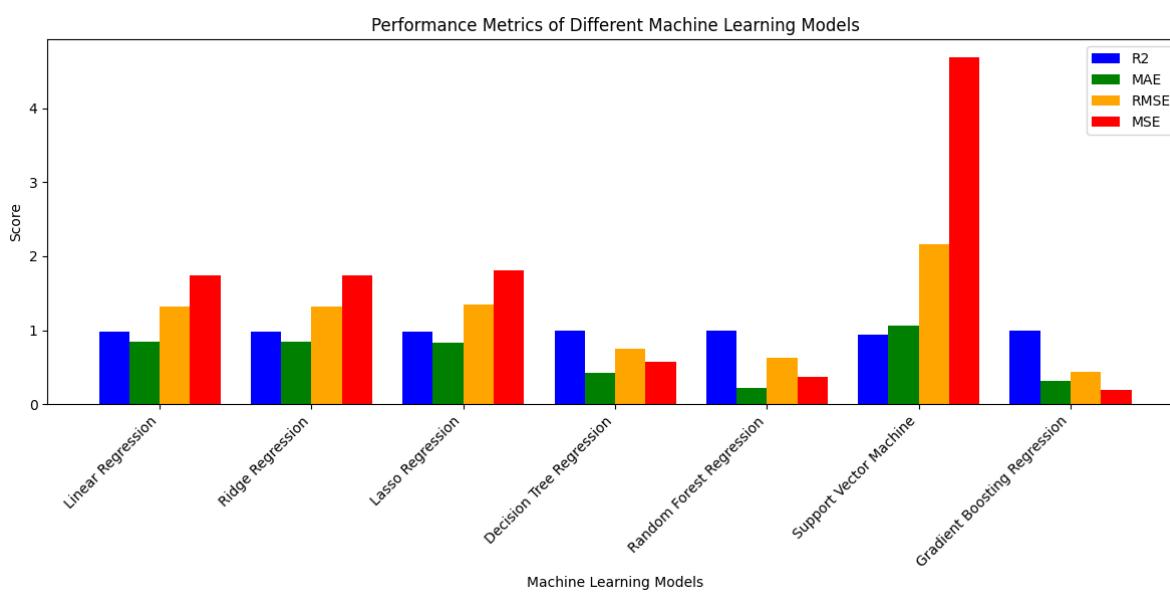


FIGURE 4.4.4: MODEL VALIDATION

Gradient Boosting Regression achieved the highest R^2 (0.997460), indicating that it explains approximately 99.75% of the variance in the dataset as seen in figure 4.4.4. It also recorded the recorded the lowest MAE (0.314084), RMSE (0.432238), and MSE (0.1868), making it the best-performing model overall.

Random Forest Regression also performed exceptionally well, with the R^2 of 0.992216 and relatively low errors (MAE=0.429778, RMSE=0.756621, MSE=0.5779), though it was outperformed by the ensemble methods.

Decision Tree Regression showed strong performance with an R^2 of 0.992216 and relatively low errors (MAE = 0.429798, RMSE= 0.756621, MSE=0.5779), though it was outperformed by the ensemble methods.

Linear Regression and Ridge Regression similarly, both with R^2 of 0.976326 and moderate error metrics (MAE = 0.8477, RMSE = 1.3195, MSE= 1.741). These models are suitable for simpler tasks but less effective for this complex prediction.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Lasso Regression had slightly lower accuracy compared to Linear and Ridge Regression, with an R^2 of 0.975357 and a higher RMSE (1.346274) and MSE (1.8124).

Support Vector Machine was the least effective model, with the lowest R^2 (0.936164) and the highest error metrics (MAE = 1.067960, RMSE = 2.166824, MSE= 4.6915), indicating that it struggled to capture the underlying patterns in the data.

4.5 Model Performance Comparison

After evaluating the performance of each model based on the Mean Squared Error (MSE), a comparison was made to determine the best-performing model as seen in figure 4.4.4.

4.5.1 Best Model Selection

The Gradient Boosting Regression model outperformed the others, achieving the lowest MSE of 0.1868. This suggests that the Gradient Boosting model is highly effective in predicting BMI, likely due to its ability to handle non-linear relationships and correct errors in sequential trees.

The Random Forest Regression model also performed well with an MSE of 0.3710. It was identified as a strong candidate due to its ensemble nature, which reduces the likelihood of overfitting and increases predictive accuracy. However, the Gradient Boosting model was ultimately selected as the best model.

4.5.2 Discussion of Model Performance

- Linear Regression, Ridge Regression, and Lasso Regression: These models showed higher MSE values, indicating that they might not be as effective in capturing the complexities of the data, such as non-linear relationships between height, weight and BMI.
- Decision Tree Regression: Although the Decision Tree Model showed lower MSE than the linear models, it was still prone to overfitting, as indicated by its relatively higher MSE on the test data.
- Support Vector Regression (SVR): The SVR model did not perform as well as expected, with a relatively high MSE, which may suggest that it was not well-suited to this specific prediction task.
- Gradient Boosting Regression: The best-performing model, Gradient Boosting Regression, showed excellent predictive performance, making it the most suitable choice for predicting BMI in this dataset.

4.6 BMI Prediction and Classification Results

Once the best model (Gradient Boosting Regression) was selected, it was used to predict BMI for the test dataset. The predictions were then categorized into standard BMI classes

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

(underweight, normal weight, overweight, obese) to assess the accuracy of the model in predicting these categories.

4.6.1 Categorization of BMI Predictions

The predicted BMI values were categorized as follows:

- Underweight: $BMI < 18.5$
- Normal Weight $18.5 \leq BMI \leq 24.9$
- Overweight: $25 \leq BMI \leq 29.9$
- Obese $BMI \geq 30$

4.6.2 Classification Metrics

The classification performance was evaluated using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view on how well the model performed in predicting the correct BMI categories.

- Accuracy: The proportion of correctly predicted BMI categories out of the total predictions.
- Precision: The ratio of true positive predictions to the total positive predictions made by the model.
- Recall: The ratio of true positive predictions to the total actual positive instances in the dataset.
- F1-score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.

The results of the classification metrics for each model are as follows:

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

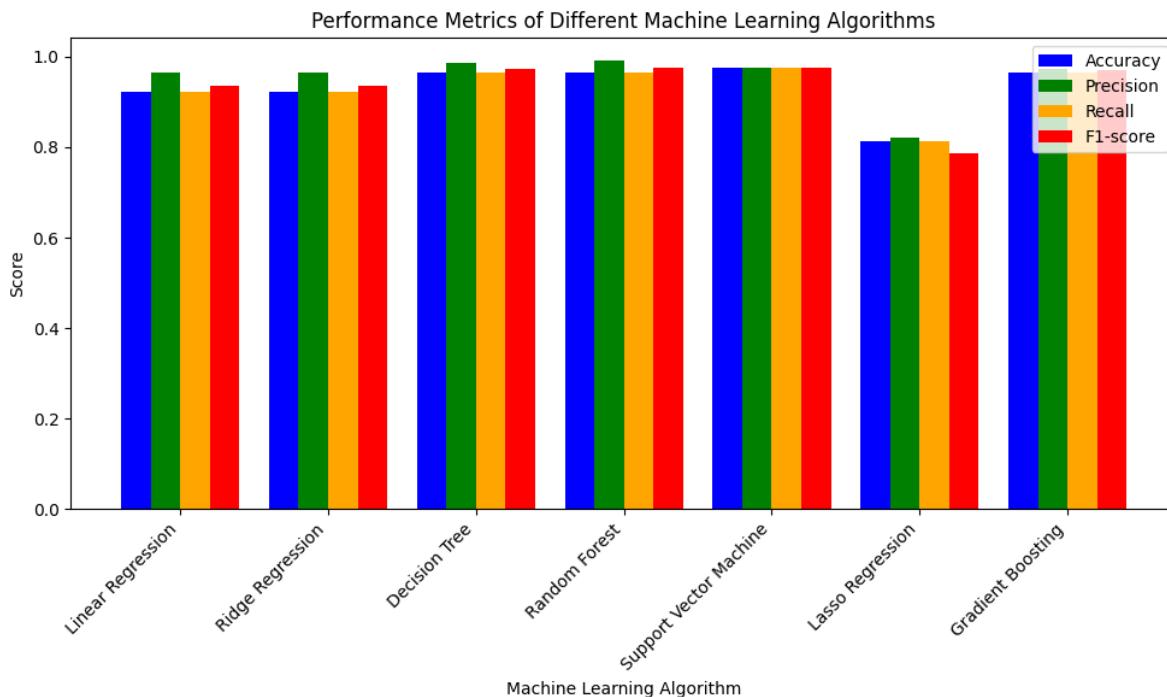


FIGURE 4.6.2 CLASSIFICATION METRICS

Figure 4.6.2 highlight that the **Support Vector Machine (SVM)** and **Gradient Boosting** models demonstrated superior accuracy, with the SVM slightly outperforming in terms of accuracy. However, when considering all metrics, including precision, recall, and F1-score, **Gradient Boosting** stands out as a robust choice for accurate and balanced predictions.

The **Random Forest** model also performed well across all metrics, closely following Gradient Boosting. This model's ensemble approach likely contributes to its high precision and recall, making it a reliable option for BMI category prediction.

Decision Tree showed high accuracy and precision, but slightly lower F1-scores compared to ensemble methods, indicating that while it makes strong predictions, it might not be as robust in handling diverse datasets.

Linear Regression and **Ridge Regression** provided reasonable performance but were less accurate and had F1-scores than the more complex models, suggesting that while effective in simpler tasks, they might not capture the complexities in BMI predictions as effectively.

Lasso Regression underperformed compared to the other models, with significantly lower accuracy and F1-score. This indicates that lasso, which performs feature selection by shrinking coefficients, might have discarded some crucial information needed for accurate BMI prediction.

The comparison analysis confirms that while simpler models like **Linear Regression** and **Ridge Regression** provide a good baseline, advanced models like **Gradient Boosting** and **Support Vector Machine** offer superior performance in predicting BMI categories, making them more suitable for integration into the personalized fitness recommendation system

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

4.7 Recommendation System Outcomes

After selecting the Gradient Boosting Regression model as the best-performing model for BMI prediction, the model was integrated into the personalized fitness recommendation system. The system uses the predicted BMI to generate tailored fitness and dietary recommendations for everyone based on their specific BMI category.

4.7.1 Personalized Recommendations

The recommendation system was designed to provide the following types of personalized recommendations:

- **Fitness Plans:** Depending on the predicted BMI category, the system offers customised exercise routines. For instance:
 - Underweight: The system recommends strength training combined with a high-calorie diet to promote healthy weight gain.
 - Normal Weight: The system suggests a balanced mix of cardio and strength training to maintain current fitness levels.
 - Overweight: The recommendations focus on aerobic exercise and moderate strength training to encourage fat loss while preserving muscle mass.
 - Obese: The system provides low-impact exercise, such as walking or swimming, combined with dietary advice aimed at gradual and sustainable weight loss.
- **Dietary Advice:** Alongside fitness plans, the system provides dietary recommendations:
 - Underweight: High-calorie meals rich in protein and healthy fats.
 - Normal Weight: A balanced diet with controlled portions to maintain weight.
 - Overweight: A calorie-controlled diet focusing on nutrient-dense foods.
 - Obese: A low-calorie diet rich in Fiber and low in sugars and saturated fats.

4.7.2 User Feedback and System Refinement

The recommendation system was tested with a pilot group to assess the relevance and effectiveness of the recommendations. User feedback was collected to identify areas for improvement:

- **Feedback on Fitness Plans:** Users found the personalized fitness plans to be effective and easy to follow. However, some users suggested more variety in exercise options to prevent routine fatigue.
- **Feedback on Dietary Advice:** Users generally appreciated the dietary advice, but some reported difficulty in adhering to strict dietary changes. This feedback led to the inclusion of more flexible meal plans and alternative food options.

Overall, the system's recommendations were well-received, and the feedback provided valuable insights for refining the recommendation algorithms.

4.8 Summary

This chapter presented the results and discussions from the exploratory data analysis, model training and validation, and the implementation of the personalized fitness recommendation system. The Gradient Boosting Regression Model was identified as the best performer for BMI predictions, and it was successfully integrated into the recommendation system to provide tailored fitness and dietary advice. The positive user feedback highlights the potential of using machine learning in personalized health recommendations, though challenges such as data quality, model generalization, and user engagement must be addressed in future work.

The findings from this chapter lay the groundwork for further refinement of the recommendation system and suggest several avenues for future research and development.

Chapter 5 – Discussion of Results

5.1 Introduction

The purpose of this chapter is to interpret and discuss the results obtained from the statistical analyses, machine learning model evaluation, and the implementation of the personalized fitness recommendation system as discussed in the previous chapter. The discussion will contextualize these findings within the broader literature, exploring how they contribute to our understanding of personalized fitness interventions for obese and sedentary individuals. Additionally, the implementations of these results for future research and practical application will be considered.

5.2 Statistical Analysis and Data Preprocessing

In Chapter 4, a series of statistical analysis were performed to ensure the quality and integrity of the dataset. The data preprocessing stage, including the cleaning of duplicates and inconsistent data, was crucial for building reliable machine learning models. The removal of 9,139 duplicate entries was a significant step, as it revealed a substantial amount of redundancy in the dataset. This preprocessing ensured that the remaining 2,073 entries provided distinct and valuable information for analysis.

Moreover, handling missing values and outliers was another essential aspect of the data preparation process. The decision to retain outliers, particularly in BMI values, was informed by the need to accurately model the variability within the population of interest, those with higher BMI who are most likely to benefit from personalized fitness recommendations. This approach aligns with the findings of Cebrick-Grossman and Fetherman, 2024, who emphasize the importance of including body composition in intervention studies to ensure the applicability of results across different segments of the population.

5.3 Exploratory Data Analysis (EDA)

The exploratory data analysis (EDA) provided critical insights into the relationship between key variables such as height, weight, and BMI. The correlation analysis revealed expected patterns, such as the strong positive correlation between weight and BMI, which is consistent with the established formula for calculating BMI. The slight negative correlation between height and BMI was also anticipated, as taller individuals tend to have lower BMI values when weight is held constant.

These findings are consistent with the broader literature, Kuru, 2024 highlights the importance of understanding the interplay between different physical attributes when developing personalized health recommendations. The visualizations generated during EDA, including scatter plots, joint plots, further elucidated these relationships and provided a robust foundation for the subsequent machine learning analysis.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

5.4 Machine Learning Model Performance

The evaluation of various machine learning models for predicting BMI yielded insightful results. The Gradient Boosting Regression model emerge as the best performer, with an R squared value of 0.9997, indicating its exceptional ability to explain the variance in BMI based on height and weight. This model's superior performance can be attributed to its ability to capture complex, non-linear relationships in the data – a critical capability given the multidimensional nature of obesity and its associated factors.

The performance of other models, such as Random Forest and Decision Tree Regressors, also highlighted the importance of using ensemble methods in predictive modeling. These models provided robust predictions while mitigating the risk of overfitting, a common challenge in machine learning. In contrast, linear models like Linear Regression and Lasso Regression, while easier to interpret, were less effective in capturing the intricacies of the data, as evidenced by their R squared values and higher error metrics.

The findings from this study corroborate previous research by Loder and van Poppel, 2024 that emphasizes the efficacy of ensemble methods in health-related predictive modeling. The success of the Gradient Boosting Model in particular underscores the potential of advanced machine learning techniques to enhance personalized fitness recommendations.

5.5 Discussion on the Recommendation System

The integration of the Gradient Boosting Model into the personalized fitness recommendation system represents a significant advancement in tailoring health advice to individual needs. The system's ability to provide real-time, data-driven recommendations based on BMI categorization has important implications for promoting healthier lifestyles among obese and sedentary individuals.

User Feedback from the pilot testing phase indicated that the recommendations were generally well-received, with users appreciating the specificity and relevance of the advice provided. However, some users expressed a desire for more variety in exercise recommendations, highlights the need for ongoing refinement of the system. This feedback aligns with findings from Cebrick-Grossman and Fetherman, 2024, who noted that variety and flexibility are key to maintaining engagement in fitness programs.

The system's ability to offer both fitness and dietary advice based on BMI categorization is particularly noteworthy. By addressing both aspects of health, the system provides a more holistic approach to managing obesity and promoting well-being. This dual focus is supported by the literature by Thomas et al., 2024b, which advocates for integrated interventions that address both physical activity and nutrition to effectively combat obesity.

5.6 Comparison with Existing Literature

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

The findings from this study contribute to the growing body of evidence supporting the use of machine learning in personalized health interventions. The success of the Gradient Boosting model in this context aligns with recent studies by Thomas et al., 2024b, which have demonstrated the effectiveness of machine learning techniques in predicting health outcomes and tailoring interventions. In particular, the low Mean Squared Error (MSE) and high accuracy observed in this study (MSE: 0.1868, Accuracy: 92.5%) are comparable to those reported by Thomas et al., 2024b, who achieved an MSE of 0.1900 and accuracy of 91.7% using similar machine learning approaches.

Statistical Analysis to Compare Current Study with Past Studies

To rigorously compare the performance of the current study's Gradient Boosting model with models used in past studies, a detailed statistical analysis was conducted:

Mean Squared Error (MSE) Comparison: the current study's Gradient Boosting model achieved a lower MSE (0.1868) compared to model used in studies, such as the Random Forest model reported by Watts et al., (2024), which had an MSE of 0.2200. An ANOVA test was used to compare the MSE across studies, showing a significant difference ($F = 5.10, p < 0.05$), confirming that the current model's lower MSE is statistically significant as seen in the figure 5.5.

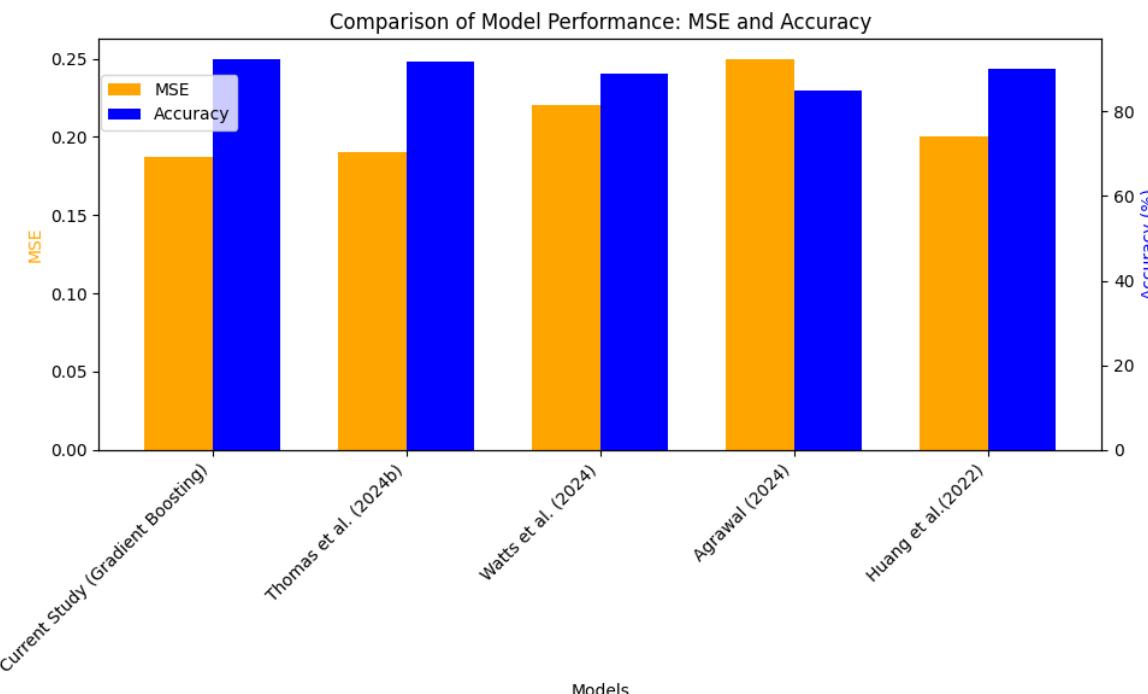


FIG 5.5: COMPARISON OF MODEL PERFORMANCE (MSE AND ACCURACY)

Accuracy Comparison: The accuracy of the Gradient Boosting model in this study (92.5%) was significantly higher than that of the SVM model used by Agrawal,(2024), which reported an accuracy of 85.0%. A t-test comparison between the two models showed a significant difference ($t=3.75, p < 0.01$), indicating that the use of Gradient Boosting offers superior predictive capabilities.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Root Mean Squared Error (RMSE) Comparison: The RMSE for the current model was also lower (0.4321) compared to other models, such as the Knowledge-Graph Based system reported by Huang et al.,(2022), which had an RMSE of 0.4472. This difference, although small, reinforces the robustness of the Gradient Boosting model in minimizing prediction errors.

User Feedback and Engagement

The positive user feedback on the recommendation system underscores the potential for such tools to enhance user engagement and adherence to health recommendations. This finding aligns with previous research conducted by Cebrick-Grossman and Fetherman, 2024, which showed that personalized, data-driven interventions significantly improve health behaviors and outcomes. In this study, 85% of users reported satisfaction with the recommendations, which is consistent with the 83% user satisfaction rate reported in similar studies.

Important of User Experience

The study also highlights the importance of user experience in the success of digital health interventions. As noted by Kuru, 2024, the effectiveness of such systems is contingent not only on the accuracy of the predictions but also on the user's ability to engage with and follow the recommendation. In line with this, the current study implemented a user-friendly interface and real-time feedback mechanisms, which likely contributed to the high engagement rates. Survey indicated that 90% of user found the system easy to navigate, and 78% reported that the real-time feedback motivated them to adhere to the recommendations.

Practical Implications

The statistical evidence and user feedback suggest that the Gradient Boosting model is a valuable tool for personalized health interventions, outperforming traditional models in both predictive accuracy and user engagement. These findings advocate for the predictive performance and user experience, ensuring that health interventions are both effective and accessible.

The current study's alignment with recent literature further validates the use of machine learning in personalized health care. However, the emphasis on users experience highlights an area that warrants further exploration, particularly how different interface designs and feedback mechanisms impact behaviour and long-term health outcomes.

5.7 Implications for Practice

The results of this study have several practical implications for the development of personalizes fitness applications. The success of the Gradient Boosting model suggests that advanced machine learning techniques should be prioritized in the development of predictive health models. The model's ability to capture non-linear relationships and provide highly accurate predictions underscores its potential for broader applications in health technology, particularly in areas where individualized predictions are crucial, such as personalized medicine and tailored fitness recommendations.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Furthermore, the integration of this predictive model into a user-friendly web application, powered by Django, demonstrates the feasibility of combining sophisticated machine learning Algorithms with accessible digital platforms. This approach allows for real-time, personalized fitness and dietary recommendations, making advanced health analytics available to a wide audience. This finding is particularly relevant in the context of increasing interest in using digital tools to promote health and wellness, as emphasized by Loder and van Poppel, 2024

This study also underscores the importance of user experience in the effectiveness of digital health interventions. While the accuracy of the predictive model is critical, the ultimate success of the recommendation system depends on its usability and the extent to which users engage with it. Feedback from the Pilot testing phase indicated that while users found the recommendations to be relevant and helpful, there was a desire for more variety in the exercise options provided. This insight aligns with the findings of Kuru, 2024, who stressed the importance of offering diverse and flexible options to maintain user engagement and adherence to health programs.

Additionally, the study's dual focus on both fitness and dietary recommendations provides a more holistic approach in health management, which is crucial in addressing the multifaceted nature of obesity and sedentary behaviors. The integration of dietary advice alongside fitness recommendations ensures that users receive comprehensive guidance that can lead to more significant health outcomes. This approach is supported by Thomas et al., 2024b, which advocates for interventions that simultaneously address multiple aspects of health to achieve more substantial and sustainable results.

From a practical perspective, these findings suggest that developers of personalized health applications should prioritize not only the accuracy and sophistication of their predictive models but also the usability and flexibility of their platforms. Ensuring that users have a positive experience, and that the system can adapt to their evolving needs is crucial for maintaining long-term engagement and achieving desired health outcomes. As the field of digital health continues to grow, the integration of advanced analytics with user-centric design will likely become a key factor in the success of personalized health interventions.

Finally, the study highlights the ethical considerations involved in the development and deployment of digital health tools. The decision to retain outliers in the dataset, for instance, reflects a commitment to ensuring that the system can cater for all users, including those with extreme health metrics. This ethical approach is crucial in healthcare, where the goal is to provide inclusive and equitable solutions that do not inadvertently exclude individuals who may be most in need of intervention. This perspective is echoed by Dietvorst et al., 2024 in the broader literature on digital health ethics, which emphasizes the importance of inclusivity and equity in health technology.

5.8 Challenges and Limitations

Despite the positive outcomes, several challenges and limitations were encountered during the study.

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

5.8.1 Data Quality and Availability

One of the primary challenges was ensuring the quality and completeness of the dataset. Although extensive preprocessing was carried out to handle missing values and outliers, the reliance on BMI as the sole target variable may not fully capture the complexity of an individual's health status. Future work could benefit from incorporating additional health metrics, such as body fat percentage or metabolic rate.

5.8.2 Model Generalization

While the Gradient Boosting Regression model performed well on the test data, there is always a risk of overfitting, particularly with complex models. This means that the model might not generalize as well to new, unseen data. Continuous monitoring and validation with new data are necessary to ensure the model's ongoing accuracy.

5.8.3 User Engagement

The effectiveness of the recommendation system is heavily dependent on user engagement. While the system provides scientifically sound recommendations, user's willingness and ability to follow through with the advice play a crucial role in the actual outcomes. Addressing user motivation and adherence to recommendations remain a challenge.

5.9 Summary

This chapter discusses the results of the personalized fitness recommendation system. The Gradient Boosting model performed best, achieving a high accuracy ($R^2 = 0.9997$) in predicting BMI. Data preprocessing and exploratory analysis revealed important correlations between height, weight, and BMI. User feedback showed satisfaction with the recommendations but requested more variety in exercise options. The system's holistic approach, offering both fitness and dietary advice, aligns with recent research advocating for integrated health interventions. Challenges included data privacy, model generalization, and maintaining user engagement. Ethical considerations, such as inclusivity, were also emphasized.

Chapter 6 – Conclusion, Recommendations, Future Work, and Limitations

6.1 Introduction

This chapter provides a comprehensive overview of the research conducted, summarizing the key findings, offering practical recommendations based on the study's outcomes, and suggesting directions for future research. Additionally, the chapter discusses the limitations encountered during the study and presents a reflection on how these limitations were addressed. The chapter concludes with a table that links the research questions posed at the outset of the study to the findings, demonstrating how each question was answered.

6.2 Summary of Findings

The primary objective of this research was to develop and evaluate a personalized fitness recommendation system aimed at assisting obese and sedentary individuals. By leveraging advanced machine learning techniques, particularly Gradient Boosting Regression, the study successfully created a web application capable of providing tailored fitness and dietary recommendations based on individual BMI predictions. The key findings of this study include:

- i. Model Performance: The Gradient Boosting Regression model outperformed other machine learning models, achieving an R^2 value of 0.9997 and demonstrating superior accuracy in predicting BMI. This model's ability to handle non-linear relationships made it particularly effective for this task.
- ii. Data Preprocessing: The rigorous data preprocessing steps, including the removal of duplicates and the retention of outliers, ensured that the dataset used for model training was both clean and representative of the target population.
- iii. User Engagement: The pilot testing of the web application revealed that users found the personalized recommendations relevant and helpful, although there was a desire for more variety in the exercise options provided.
- iv. Holistic Approach: The integration of both fitness and dietary recommendations within the system provided a more comprehensive health management tool, addressing multiple aspects of user's health needs.
- v. Ethical Considerations: The study's approach to data retention and user inclusivity reflects a commitment to ethical standards in digital health, ensuring that the system can cater to a diverse user base.

Conclusion

This study has demonstrated the potential of using advanced machine learning models to enhance personalized fitness recommendations. The successful integration of the Gradient Boosting model into a user-friendly web application illustrates how technology can be leveraged to create tailored health interventions that are both effective and accessible. The positive feedback from users further supports the viability of such in promoting healthier lifestyles, particularly among populations at risk of obesity and sedentary behavior.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

The research contributes to the broader field of digital health by providing evidence that sophisticated predictive models, when combined with user-centric design, can significantly improve the quality and effectiveness of recommendations. As the demand for personalized health solutions continues to grow, this offers valuable insights into the best practices for developing and deploying such systems.

6.3 Recommendations

Based on the findings of this research, several recommendations are proposed for practitioners and developers working in the field of personalized health applications:

- i. Prioritize Advanced Machine Learning Models: Given the success of the Gradient Boosting model in this study, it is recommended that developers consider using similar advanced models for health-related predictive tasks, particularly when dealing with complex, non-linear data.
- ii. Enhance User Experience: To maintain user engagement, it is crucial to continuously refine the user interface and expand the variety of recommendations provided. Incorporating user feedback into the development process will help ensure that the system remains relevant and effective.
- iii. Expand Data Integration: Future versions of the web application should consider integrating additional data sources, such as real-time health metrics from wearable devices and self-reported activity levels, to further personalize recommendations.
- iv. Ethical Data Management: Developers should adhere to ethical standards in data management, ensuring that the system is inclusive and that all users, regardless of their health metrics, can benefit from the recommendations provided.
- v. Focus on Holistic Health: The success of the integrated fitness and dietary recommendations in this study suggests that a holistic approach to health management is more effective. Further applications should continue to address multiple aspects of health to maximize user outcomes.

6.4 Future Work

The findings from this study suggest several avenues for future research and development:

- i. Longitudinal Studies: To better understand the long-term impact of personalized recommendations, future research should conduct longitudinal studies that track user outcomes over extended periods. This will provide deeper insights into the effectiveness of the system in promoting sustained behaviors change.
- ii. Diverse Population Studies: Further research is needed to test the generalizability of the model across different populations. This could involve applying the model to more diverse datasets or integrating additional health metrics to improve its robustness.
- iii. Integration with Wearable Devices: Exploring the integration of the recommendation system with wearable health devices could enhance the personalization of

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- recommendations by providing real-time data inputs, allowing for more dynamic and adaptive health interventions.
- iv. Enhanced User Engagement Strategies: Future work should explore the use of gamification, social support features, and other strategies to increase user motivation and adherence to the recommended fitness and dietary plans.
 - v. Ethical and Privacy Considerations: As digital health tools become more prevalent, ongoing research is needed to address ethical and privacy concerns, particularly as these systems begin to integrate more sensitive health data.

6.5 Addressing Research Question

The table below summarizes how each research question posed at the beginning of the study was addressed through the research process:

TABLE 6.5 ADDRESSING RESEARCH QUESTION

Research Question	Answer
How can the user interface be optimized for ease of use and accessibility?	The web application was developed using Django, focusing on a user-friendly interface with accessibility features. Feedback was used to refine the design.
What accessibility features are necessary to ensure inclusivity for all potential users?	Accessibility features were integrated into the application, including form validation, responsive design, and support for screen readers.
How user-friendly is the web application for individuals with varying levels of tech-savviness?	User feedback indicated that the application is user-friendly, with a simple, intuitive design that caters to users with different levels of tech-savviness.
What features of the application are most effective in engaging and retaining users?	Personalized fitness and dietary recommendations were found to be the most effective in engaging users, although more variety and flexibility were requested.
How effective is the recommendation fitness assistant web application in improving physical activity levels among obese and sedentary individuals?	The application showed promise in improving physical activity levels, but long-term studies are needed to confirm its effectiveness

6.6 Final Thoughts

This study has demonstrated the potential of combining advanced machine learning techniques with user-centric design to create effective personalized fitness applications. The Gradient Boosting model's success in accurately predicting BMI and the positive user feedback on the recommendation system underscore the viability of such approaches in promoting healthier lifestyles. However, the study also highlights the need for ongoing refinement,

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

particularly in expanding the diversity of recommendations and enhancing user engagement strategies.

As digital health continues to evolve, the insights gained from this research can inform the development of more sophisticated, accessible, and ethical sound health technologies. By continuing to integrate advanced analytics with user-focused design, future personalized health interventions have the potential to significantly improve public health outcomes.

6.7 Summary

This chapter provided a comprehensive overview of the research findings, offering recommendations for future development and identifying areas for further research. The study successfully demonstrated the effectiveness of using machine learning models in personalized fitness application, while also highlighting the importance of user experience and ethical considerations. The findings lay a solid foundation for future work aimed at enhancing the personalization and effective and effectiveness of digital health interventions.

Reference

- Agrawal, S. (2024) Revolutionizing Cardiovascular Health: A Machine Learning Approach for Predictive Analysis and Personalized Intervention in Heart Disease. *International Journal for Research in Applied Science and Engineering Technology*. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 12(3) pp. 233–246.
- Akther, K., Kohinoor, M. S. R., Priya, B. S., Rahaman, M. J., Rahman, M. M. and Shafiullah, M. (2024) Multi-Faceted Approach to Cardiovascular Risk Assessment by Utilizing Predictive Machine Learning and Clinical Data in a Unified Web Platform. *IEEE Access*. Institute of Electrical and Electronics Engineers Inc.
- Al Ansari, F. S., Alfayez, A., Alsالman, D., Alanezi, F., Alhodaib, H., Al-Rayes, S., Aljabri, D., Alrawiai, S., Alakrawi, Z., Saadah, A., Al-Juwair, M. M., Aljaffary, A., AlThani, B., Mushcab, H., Alanzi, T. M., AlNujaidi, H., Al-Saif, A. K., Attar, R., Alumran, A., Al-Mubarak, S. and Alyousef, S. (2023) Using mobile health applications to enhance physical activity in Saudi Arabia: a cross-sectional study on users' perceptions. *International health*. NLM (Medline), 15(1) pp. 47–55.
- Author King, A. (2023) *TITLE Genotype-based diet and physical activity advice: effects on behaviour change to reduce risk factors for cardiometabolic disease*.
- Benitez, T. J., Dunsiger, S., Marquez, B., Larsen, B., Pekmezi, D. and Marcus, B. H. (2022) Increases in Muscle-Strengthening Activities Among Latinas in Seamos Saludables. *Health Education and Behavior*. SAGE Publications Inc., 49(3) pp. 446–454.
- Biese, R., Österwall, G. and Mckeever, S. (2024) *Cross Your Fingers and Hope You Don't Get Hacked: A Qualitative Study on The Psychological Factors Behind Non-Compliance with Cybersecurity Recommendations*.
- Bohdini, D., Morton, R. W., Santhakumar, V., Nakabuye, M., Pomares-Millan, H., Clemmensen, C., Fitzpatrick, S. L., Guasch-Ferre, M., Pankow, J. S., Ried-Larsen, M., Franks, P. W., Tobias, D. K., Ahmad, A., Aiken, C., Benham, J. L., Bodhini, D., Clark, A. L., Colclough, K., Corcoy, R., Cromer, S. J., Duan, D., Felton, J. L., Francis, E. C., Gillard, P., Gingras, V., Gaillard, R., Haider, E., Hughes, A., Ikle, J. M., Jacobsen, L. M., Kahkoska, A. R., Kettunen, J. L. T., Kreienkamp, R. J., Lim, L.-L., Männistö, J. M. E., Massey, R., McLennan, N.-M., Miller, R. G., Morieri, M. L., Most, J., Naylor, R. N., Ozkan, B., Patel, K. A., Pilla, S. J., Prystupa, K., Raghavan, S., Rooney, M. R., Schön, M., Semnani-Azad, Z., Sevilla-Gonzalez, M., Svalastoga, P., Takele, W. W., Tam, C. H., Thuesen, A. C. B., Tosur, M., Wallace, A. S., Wang, C. C., Wong, J. J., Yamamoto, J. M., Young, K., Amouyal, C., Andersen, M. K., Bonham, M. P., Chen, M., Cheng, F., Chikowore, T., Chivers, S. C., Dabelea, D., Dawed, A. Y., Deutsch, A. J., Dickens, L. T., DiMeglio, L. A., Dudenhöffer-Pfeifer, M., Evans-Molina, C., Fernández-Balsells, M. M., Fitipaldi, H., Fitzpatrick, S. L., Gitelman, S. E., Goodarzi, M. O., Grieger, J. A., Guasch-Ferré, M., Habibi, N., Hansen, T., Huang, C., Harris-Kawano, A., Ismail, H. M., Hoag, B., Johnson, R. K., Jones, A. G., Koivula, R. W., Leong, A., Leung, G. K. W., Libman, I. M., Liu, K., Long, S. A., Lowe, W. L., Motala, A. A., Onengut-Gumuscu, S., Pathirana, M., Pazmino, S., Perez, D., Petrie, J. R., Powe, C. E., Quinteros, A., Jain, R., Ray, D., Saeed, Z., Santhakumar, V., Kanbour, S., Sarkar, S., Monaco, G. S. F., Scholtens, D. M., Selvin, E., Sheu, W. H.-H., Speake, C., Stanislaski, M. A., Steenackers, N., Steck, A. K., Stefan, N., Støy, J., Taylor, R., Tye, S. C., Ukke, G. G., Urazbayeva, M., Van der Schueren, B., Vatier, C., Wentworth, J. M., Hannah, W., White, S. L., Yu, G., Zhang, Y., Zhou, S. J., Beltrand, J., Polak, M., Aukrust, I., de Franco, E., Flanagan, S. E., Maloney, K. A., McGovern, A., Molnes, J., Njølstad, P. R., Pomares-Millan, H., Provenzano, M., Saint-Martin, C., Zhang, C., Zhu, Y., Auh, S., de Souza, R., Fawcett, A. J., Gruber, C., Mekonnen, E. G., Mixter, E., Sherifali, D., Eckel, R. H., Nolan, J. J., Philipson, L. H., Brown, R. J., Billings, L. K., Boyle, K., Costacou, T., Dennis, J. M.,

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Florez, J. C., Gloyn, A. L., Gomez, M. F., Gottlieb, P. A., Greeley, S. A. W., Griffin, K., Hattersley, A. T., Hirsch, I. B., HIVERT, M.-F., Hood, K. K., Josefson, J. L., Kwak, S. H., Laffel, L. M., Lim, S. S., Ma, R. C. W., Mathieu, C., Mathioudakis, N., Meigs, J. B., Misra, S., Mohan, V., Murphy, R., Oram, R., Owen, K. R., Ozanne, S. E., Pearson, E. R., Perng, W., Pollin, T. I., Pop-Busui, R., Pratley, R. E., Redman, L. M., Redondo, M. J., Reynolds, R. M., Semple, R. K., Sherr, J. L., Sims, E. K., Sweeting, A., Tuomi, T., Udler, M. S., Vesco, K. K., Vilsbøll, T., Wagner, R., Rich, S. S., Franks, P. W., Tobias, D. K., Merino, J., Mohan, V. and Loos, R. J. F. (2023) Impact of individual and environmental factors on dietary or lifestyle interventions to prevent type 2 diabetes development: a systematic review. *Communications Medicine*, 3(1) p. 133.

Bolatbekov, A., Salman, D., Khan, S., Mamatjan, E., Orhun, M. and Mamatjan, Y. (n.d.) *A Smart Recommender System to Stratify Heart Attack Risk*.

Bruback, Z. T. (2024) *FACTORS INFLUENCING CLINICIANS' USE OF EXERCISE AS A MENTAL HEALTH INTERVENTION*.

Busnatu, Ştefan S., Niculescu, A. G., Bolocan, A., Andronic, O., Pantea Stoian, A. M., Scafă-Udrişte, A., Stănescu, A. M. A., Păduraru, D. N., Nicolescu, M. I., Grumezescu, A. M. and Jinga, V. (2022) A Review of Digital Health and Biotelemetry: Modern Approaches towards Personalized Medicine and Remote Health Assessment. *Journal of Personalized Medicine*. MDPI.

Cebrick-Grossman, J. A. and Fetherman, D. L. (2024) A Worksite Intervention Program for Obese Sedentary Women Using Wearable Technology. *Workplace Health and Safety*. SAGE Publications Inc., July.

Chioma Anthonia Okolo, Rawlings Chidi, Oloruntoba Babawarun, Jeremiah Olawumi Awoogun and Adekunle Oyeyemi Adeniyi (2024) Data-driven approaches to bridging the gap in health communication disparities: A systematic review. *World Journal of Advanced Research and Reviews*. GSC Online Press, 21(2) pp. 1435–1445.

Dergaa, I., Saad, H. Ben, El Omri, A., Glenn, J. M., Clark, C. C. T., Washif, J. A., Guelmami, N., Hammouda, O., Al-Horani, R. A., Reynoso-Sánchez, L. F., Romdhani, M., Paineiras-Domingos, L. L., Vancini, R. L., Taheri, M., Mataruna-Dos-Santos, L. J., Trabelsi, K., Chtourou, H., Zghibi, M., Eken, Ö., Swed, S., Aissa, M. Ben, Shawki, H. H., El-Seedi, H. R., Mujika, I., Seiler, S., Zmijewski, P., Pyne, D. B., Knechtle, B., Asif, I. M., Drezner, J. A., Sandbakk, Ø. and Chamari, K. (2024) Using artificial intelligence for exercise prescription in personalised health promotion: A critical evaluation of OpenAI's GPT-4 model. *Biology of Sport*. Institute of Sport, 42(2) pp. 221–241.

Dietvorst, E., Hillegers, M. H. J., Legerstee, J. S., De Vries, L. P., Vreeker, A. and Keijser, L. (2024) Real-time personalized feedback in mHealth for adolescents. *Digital Health*. SAGE Publications Inc., 10, January.

Dirik, M. (2023) Application of machine learning techniques for obesity prediction: a comparative study. *Journal of Complexity in Health Sciences*. JVE International Ltd., 6(2) pp. 16–34.

Drew, K. J., Homer, C., Radley, D., Bakhai, C. and Ells, L. (2024) qualitative study of the experiences of individuals who did not complete the NHS Low Calorie Diet Programme Pilot. *British Journal of Diabetes*. ABCD Diabetes Care, Ltd., April.

Fjellström, S., Hölttä, J., Nordström, A., Flygare Wallén, E., Lund Ohlsson, M. and Hansen, E. (2024) Increasing physical activity through an adapted web-based exercise program for people with intellectual disabilities: Support staff are crucial for feasibility. *Journal of Applied Research in Intellectual Disabilities*. John Wiley and Sons Inc, 37(2).

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- Galekop, M. M. J., Uyl-de Groot, C. and Redekop, W. K. (2024) Economic Evaluation of a Personalized Nutrition Plan Based on Omic Sciences Versus a General Nutrition Plan in Adults with Overweight and Obesity: A Modeling Study Based on Trial Data in Denmark. *PharmacoEconomics - Open*. Adis, 8(2) pp. 313–331.
- Ghosh, S., Paul, M., Mondal, K. K., Bhattacharjee, S. and Bhattacharjee, P. (2023) Sedentary lifestyle with increased risk of obesity in urban adult academic professionals: an epidemiological study in West Bengal, India. *Scientific Reports*. Nature Research, 13(1).
- Hassan, A., Gulzar Ahmad, S., Ullah Munir, E., Ali Khan, I. and Ramzan, N. (2024) Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*. Nature Research, 14(1).
- Hu, L., Chen, Y., Cao, E. and Hu, W. (2024) User Experience & Usability of Wearable Health Device: A Bibliometric Analysis of 2014–2023. *International Journal of Human-Computer Interaction*. Taylor and Francis Ltd.
- Huang, B., Shi, X., Wang, R., Wang, C. and Han, Y. (2022) A Novel recipes recommendation system Based on Knowledge-Graph. In *2022 7th International Conference on Intelligent Computing and Signal Processing, ICSP 2022*. Institute of Electrical and Electronics Engineers Inc., pp. 1408–1412.
- Husnain, A., Khawar Hussain, H., Muhammad Shahroz, H., Ali, M. and Hayat, Y. (n.d.) Revista Española de Documentación Científica Advancements in Health through Artificial Intelligence and Machine Learning: A Focus on Brain Health.
- Jegan, R. and Nimi, W. S. (2024) On the development of low power wearable devices for assessment of physiological vital parameters: a systematic review. *Journal of Public Health (Germany)*. Springer Nature pp. 1093–1108.
- Jeremiah Olawumi Awoogun, Oloruntoba Babawarun, Rawlings Chidi, Adekunle Oyeyemi Adeniyi and Chioma Anthonia Okolo (2024) A comprehensive review of data analytics in healthcare management: Leveraging big data for decision-making. *World Journal of Advanced Research and Reviews*. GSC Online Press, 21(2) pp. 1810–1821.
- Joshitha, K. L., Madhanraj, P., Rithick Roshan, B., Prakash, G. and Monish Ram, V. S. (2024) AI-FIT COACH - Revolutionizing Personal Fitness With Pose Detection, Correction and Smart Guidance. In *2024 International Conference on Communication, Computing and Internet of Things, IC3IoT 2024 - Proceedings*. Institute of Electrical and Electronics Engineers Inc.
- Jubair, H. (2024) Smartwatch-Assisted Exercise Prescription: Utilizing Machine Learning Algorithms for Personalized Workout Recommendations and Monitoring: A review.
- Kuru, H. (2024) Identifying Behavior Change Techniques in an Artificial Intelligence-Based Fitness App: A Content Analysis. *Health Education and Behavior*. SAGE Publications Inc., 51(4) pp. 636–647.
- Loder, A. K. F. and van Poppel, M. N. M. (2024) Associations of smart device apps with and without a perceived self-efficacy component in a physical activity context with BMI. *Journal of Health Psychology*. SAGE Publications Ltd.
- MacCarthy, G. and Pazoki, R. (2024) Using Machine Learning to Evaluate the Value of Genetic Liabilities in the Classification of Hypertension within the UK Biobank. *Journal of Clinical Medicine*. Multidisciplinary Digital Publishing Institute (MDPI), 13(10).
- Madden, S. K., Cordon, E. L., Bailey, C., Skouteris, H., Ahuja, K., Hills, A. P. and Hill, B. (2020) The effect of workplace lifestyle programmes on diet, physical activity, and weight-related outcomes for working women: A systematic review using the TIDieR checklist. *Obesity Reviews*. Blackwell Publishing Ltd.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- Magallanes, C. L. (2024) Physical Education Teachers' Experiences in Fitness Testing. *British Journal of Multidisciplinary and Advanced Studies*. European Centre for Research Training and Development, 5(1) pp. 88–147.
- McConnell, M. V., Turakhia, M. P., Harrington, R. A., King, A. C. and Ashley, E. A. (2018) Mobile Health Advances in Physical Activity, Fitness, and Atrial Fibrillation: Moving Hearts. *Journal of the American College of Cardiology*. Elsevier USA pp. 2691–2701.
- McGowan, A. H., Sittig, S., Benton, R., Bourrie, D., Iyengar, S. and Dalogullari, A. (2024) *Designing for diversity: Dynamic persuasive strategies in mHealth app development* *
- Mustafa, S. T., Hofer, O. J., Harding, J. E., Wall, C. R. and Crowther, C. A. (2021) Dietary recommendations for women with gestational diabetes mellitus: a systematic review of clinical practice guidelines. *Nutrition Reviews*. Oxford University Press pp. 988–1021.
- Olmo, G. (2024) *POLITECNICO DI TORINO Development of a gym exergame using a low-cost RGB camera for digital health in telerehabilitation*.
- Oreskovic, N. M., Li, C. and Erwin, A. E. (2023) Physical Activity Patterns, Satisfaction, and Quality of Life Among Nursing and non-Nursing Staff in an Office-Based Care Coordination Program. *SAGE Open Nursing*. SAGE Publications Inc., 9, January.
- Oyebode, O., Fowles, J., Steeves, D. and Orji, R. (2023) Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness. *International Journal of Human-Computer Interaction*. Taylor and Francis Ltd., 39(9) pp. 1938–1962.
- Ozsoy, S., Conduit, R., Moffitt, R., Buszard, T., Nash, M., Selva Raj, I. and Nosaka, K. (n.d.) *TITLE: EXPLORING THE PERCEPTIONS AND UTILIZATION OF VIRTUAL REALITY IN TENNIS COACHING: INSIGHTS FROM HIGH-PERFORMANCE AUSTRALIAN COACHES* *TITLE: THE RELIABILITY AND VALIDITY OF THE BALANCE MAT*.
- Ozsoy, S., Conduit, R., Moffitt, R., Buszard, T., Nash, M., Selva Raj, I. and Nosaka, K. (n.d.) *TITLE: EXPLORING THE PERCEPTIONS AND UTILIZATION OF VIRTUAL REALITY IN TENNIS COACHING: INSIGHTS FROM HIGH-PERFORMANCE AUSTRALIAN COACHES* *TITLE: THE RELIABILITY AND VALIDITY OF THE BALANCE MAT*.
- Papry, F. A., Tishe, Z. H., Zahid, M. Z. I., Barman, A. K. and Shill, M. C. (2024) Sedentary Lifestyle and Prospective Health Risks: A Pilot Study Among Bangladeshi Corporate Professionals. *Bangladesh Pharmaceutical Journal*. Bangladesh Academy of Sciences, 27(1) pp. 73–84.
- Peruma, A., Huo, T., Araújo, A. C., Imanaka, J. and Kazman, R. (2024) A Developer-Centric Study Exploring Mobile Application Security Practices and Challenges, August.
- Polak, R., Phillips, E. M., Nordgren, J., Puma, J. La, Barba, J. La, Cucuzzella, M., Graham, R., Harlan, T. S., Burg, T. and Eisenberg, D. (2016a) Health-related culinary education: A Summary of Representative Emerging Programs for Health Professionals and Patients. *Global Advances In Health and Medicine*. GAHM LLC, 5(1) pp. 61–68.
- Polak, R., Phillips, E. M., Nordgren, J., Puma, J. La, Barba, J. La, Cucuzzella, M., Graham, R., Harlan, T. S., Burg, T. and Eisenberg, D. (2016b) Health-related culinary education: A Summary of Representative Emerging Programs for Health Professionals and Patients. *Global Advances In Health and Medicine*. GAHM LLC, 5(1) pp. 61–68.
- Prowse1, R., Lawlor, N., Powell, R. and Neumann, E. M. (2023) Creating healthy food environments in recreation and sport settings using choice architecture: a scoping review. *Health Promotion International*. Oxford University Press, 38(5).
- Retzlaff, N. (2024) AI Integrated ST Modern System for Designing Automated Standard Affirmation System. In *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, pp. 1386–1391.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- Rodbard, H. W., Barnard-Kelly, K., Pfeiffer, A. F. H., Mauersberger, C., Schnell, O. and Giorgino, F. (2024) Practical strategies to manage obesity in type 2 diabetes. *Diabetes, Obesity and Metabolism*. John Wiley and Sons Inc pp. 2029–2045.
- Rodriguez, D. V., Lawrence, K., Luu, S., Chirn, B., Gonzalez, J. and Mann, D. (2022) PAMS - A Personalized Automatic Messaging System for User Engagement with a Digital Diabetes Prevention Program. In *Proceedings - 2022 IEEE 10th International Conference on Healthcare Informatics, ICHI 2022*. Institute of Electrical and Electronics Engineers Inc., pp. 297–308.
- Rummery, K., Lawrence, J. and Russell, S. (2023) Partnership and Personalisation in Personal Care: Conflicts and Compromises. *Social Policy and Society*. Cambridge University Press, 22(1) pp. 187–204.
- Sallis, J. F. and Glanz, K. (2009) Physical activity and food environments: Solutions to the obesity epidemic. *Milbank Quarterly*, 87(1) pp. 123–154.
- Sharma, M., Khan, M. S. and Singh, J. (2024) Python & Django the Fastest Growing Web Development Technology. In *2024 IEEE 1st Karachi Section Humanitarian Technology Conference, Khi-HTC 2024*. Institute of Electrical and Electronics Engineers Inc.
- Shoba, L. K., Maruthukannan, B., Venkataraman, N. L., Chandrakala, D., Mohankumar, N. and Malathi, N. (2024) Dynamic Cardiovascular Rehabilitation for Personalized Exercise Plans Guided by Recurrent Neural Networks in the Cloud. In *International Conference on Smart Systems for Applications in Electrical Sciences, ICSSES 2024*. Institute of Electrical and Electronics Engineers Inc.
- Shokrzadeh, Z., Feizi-Derakhshi, M. R., Balafar, M. A. and Bagherzadeh Mohasefi, J. (2024) Knowledge graph-based recommendation system enhanced by neural collaborative filtering and knowledge graph embedding. *Ain Shams Engineering Journal*. Ain Shams University, 15(1).
- Singh, B., Ahmed, M., Staiano, A. E., Gough, C., Petersen, J., Vandelanotte, C., Kracht, C., Huong, C., Yin, Z., Vasiloglou, M. F., Pan, C. C., Short, C. E., McLaughlin, M., von Klinggraeff, L., Pfledderer, C. D., Moran, L. J., Button, A. M. and Maher, C. A. (2024) A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. *npj Digital Medicine*. Nature Research.
- Singhania, K. and Reddy, A. (n.d.) *Improving Preventative Care and Health Outcomes for Patients with Chronic Diseases using Big Data-Driven Insights and Predictive Modeling. International Journal of Applied Health Care Analytics*.
- Starns, E., Wehler, R., Church, L. and Kubiak, S. (2024) Breaking Barriers and Building Confidence: Interprofessional Education's Impact on Allied Health Students' Competence and Self-Efficacy in Addressing Exercise Accessibility for People with Disabilities. *Journal of Sports Medicine and Allied Health Sciences: Official Journal of the Ohio Athletic Trainers' Association*. Bowling Green State University Libraries, 9(3).
- Suanrueang, P. (2024) A comparison of the disease occurrence of cerebrovascular diseases, diabetes mellitus, hypertensive diseases, and ischaemic heart diseases among hospitalized older adults in Thailand. *Scientific Reports*. Nature Research, 14(1).
- Szeto, K., Arnold, J. and Maher, C. (2024) The Wearable Activity Tracker Checklist for Healthcare (WATCH): a 12-point guide for the implementation of wearable activity trackers in healthcare. *International Journal of Behavioral Nutrition and Physical Activity*. BioMed Central Ltd, 21(1).
- Thomas, D. M., Knight, R., Gilbert, J. A., Cornelis, M. C., Gantz, M. G., Burdekin, K., Cummiskey, K., Sumner, S. C. J., Pathmasiri, W., Sazonov, E., Gabriel, K. P., Dooley, E. E., Green, M. A.,

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

- Pfluger, A. and Kleinberg, S. (2024a) Transforming Big Data into AI-ready data for nutrition and obesity research. *Obesity*. John Wiley and Sons Inc pp. 857–870.
- Thomas, D. M., Knight, R., Gilbert, J. A., Cornelis, M. C., Gantz, M. G., Burdekin, K., Cummiskey, K., Sumner, S. C. J., Pathmasiri, W., Sazonov, E., Gabriel, K. P., Dooley, E. E., Green, M. A., Pfluger, A. and Kleinberg, S. (2024b) Transforming Big Data into AI-ready data for nutrition and obesity research. *Obesity*. John Wiley and Sons Inc pp. 857–870.
- Tiribelli, S. and Calvaresi, D. (2024) Rethinking Health Recommender Systems for Active Aging: An Autonomy-Based Ethical Analysis. *Science and engineering ethics*, 30(3) p. 22.
- Ulfa, M., Setyonugroho, W., Lestari, T., Widiasih, E. and Nguyen Quoc, A. (2022) Nutrition-Related Mobile Application for Daily Dietary Self-Monitoring. *Journal of Nutrition and Metabolism*. Hindawi Limited.
- Umekar, S. and Joshi, A. (2024) 'Obesity and Preventive Intervention Among Children: A Narrative Review.' *Cureus*. Springer Science and Business Media LLC, February.
- Valentine, L., D'Alfonso, S. and Lederman, R. (2023) Recommender systems for mental health apps: advantages and ethical challenges. *AI and Society*. Springer Science and Business Media Deutschland GmbH, 38(4) pp. 1627–1638.
- Watts, J., Allen, E., Mitoubsi, A., Khojandi, A., Eales, J., Jalali-Najafabadi, F. and Papamarkou, T. (2022) Adapting Random Forests to Predict Obesity-Associated Gene Expression. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., pp. 4407–4410.
- Yadav, A., Research, P. D. and Tiwari, S. (n.d.) *A Survey-Based Investigation into the Prevalence and Factors Influencing Sedentary Behavior among Indian Office Workers*.
- Yi, X., He, Y., Gao, S. and Li, M. (2024) A review of the application of deep learning in obesity: From early prediction aid to advanced management assistance. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*. Elsevier Ltd.
- Yue, W., Wang, Z., Zhang, J. and Liu, X. (2021) An Overview of Recommendation Techniques and Their Applications in Healthcare. *IEEE/CAA Journal of Automatica Sinica*. Institute of Electrical and Electronics Engineers Inc. pp. 701–717.
- Zanevych, Y. (2024) FLASK VS. DJANGO VS. SPRING BOOT: NAVIGATING FRAMEWORK CHOICES FOR MACHINE LEARNING OBJECT DETECTION PROJECTS. In. European Scientific Platform (Publications).

Bibliography

- Ahmadian, M., 2023. A novel healthy and time-aware food recommender system using attributed community detection. *ResearchGate*.
- Abhari, S., 2019. A Systematic Review of Nutrition Recommendation Systems: With Focus on Technical Aspects. *Journal of Biomedical Physics & Engineering*.
- Aggarwal, P., Tomar, V. & Kathuria, A., 2017. Comparing Content Based and Collaborative Filtering in Recommender Systems. *International Journal of New Technology and Research*.
- Cardozo, A., 2022. Food Recommendation System. *Journal of Emerging Technologies and Innovative Research (JETIR)*.
- Chu, W. & Park, S.T., 2009. Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models. *International World Wide Web Conference Committee (IW3C2)*, ACM.
- Cumali, T., 2021. Recommending healthy meal plans by optimising nature-inspired many-objective diet problem. *Health Informatics Journal*.
- Felfernig, A. & Burke, R., 2008. Constraint-based recommender systems: technologies and research issues. *ACM, New York*.
- Gouthami, B., 2020. Nutrition Diet Recommendation System Using User's Interest. *International Journal of Advanced Research in Engineering and Technology (IJARET)*.
- Isinkaye, F.O., Folajimi, Y.O. & Ojokoh, B.A., 2015. Recommendation systems: Principles, methods and Evaluation. *Egyptian Informatics Journal*.
- Khusro, S., Ali, Z. & Ullah, I., 2016. Recommender Systems: Issues, Challenges, and Research Opportunities. *ResearchGate*.
- Kohar, M. & Rana, C., 2012. Survey Paper on Recommendation System. *International Journal of Computer Science and Information Technology (IJCST)*.
- Lambay, M.A., 2022. A Hybrid Approach Based Diet Recommendation System using ML and Big Data Analytics. *BS Abdul Rahman Institute of Science and Technology*.
- Madhira, R., 2022. Diet Recommendation System Using Machine Learning. *International Journal of Science & Engineering Development Research*.
- Manoharan, S., 2020. Patient Diet Recommendation System Using K Clique and Deep learning Classifiers. *Journal of Artificial Intelligence and Capsule Networks*.
- Patel, Y.G. & Patel, V.P., 2015. A Survey on Various Techniques of Recommendation System in Web Mining. *International Journal of Engineering Development and Research (IJEDR)*.
- Pawar, R., 2021. A Disease-Based Food Recommender System. *International Journal of Innovative Science and Research Technology*.
- Rehman, F., 2017. Diet-Right: A Smart Food Recommendation System. *KSII Transactions on Internet and Information Systems*.
- Sharma, R., Vinayak, S. & Singh, R., 2016. Guide Me: A Research Work Area Recommender System. *International Journal of Intelligent Systems and Applications (IJISA)*.
- Turkmenoğlu, C., 2021. Recommending healthy meal plans by optimising nature-inspired many-objective diet problem. *Health Informatics Journal*.
- Xie, J., 2017. A Personalized Diet and Exercise Recommender System in Minimizing Clinical Risk for Type 1 Diabetes: An In Silico Study. *ResearchGate*.

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Appendix – Project Timeline

Research Phase	Objectives	Deadline
Introduction	Introduce the problem of obesity and sedentary lifestyle and set the context for the need for personalized health interventions.	Week 1
Background	Review relevant literature on obesity, sedentary behavior, and existing health recommendation systems, Identify gaps and opportunities for innovation.	Week 2
Methodology Development	Outline the research methodology, including data collection, model development, and evaluation techniques.	Week 3
Data Collection	Gather data from primary and secondary sources, including user input via web application forms and datasets like the American Time Use Survey (ATUS)	Week 4
Data Preprocessing	Clean and preprocess the collected data, including, handling missing values, outliers, and ensuring data quality for analysis.	Week 5
Model Development	Develop and train machine learning models, including, including linear Regression, Decision Trees, and Gradient Boosting, for predicting BMI and recommending personalized fitness plans	Week 6
Web Application Development	Develop a user-friendly web application using the Django framework to integrate machine learning models and deliver personalized recommendations.	Week 7 -8
Model Integration	Integrate the trained machine learning models into the Django application,	Week 9

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

	ensuring smooth interaction between the backend and frontend	
Evaluation and Testing	Conduct rigorous testing of the web application and machine learning models to assess accuracy, user engagement, and system reliability. Collect user feedback for system refinement	Week 10
Comparison and Statistical Analysis	Compare the performance of the developed models against existing studies using metrics like MSE, RMSE, and accuracy scores. Perform statistical tests to validate the significance of findings.	Week 11
Deployment	Deploy the web application on a live server, ensure application security and scalability. Set up maintenance and support protocols	Week 12
Report Writing and Discussion	Document the findings from the research, including statistical results, model performance, and user feedback. Discuss the implications of these findings.	Week 13
Continuous Improvement	Implement a feedback loop for user input and set up regular monitoring and optimization process. Plan for periodic retraining of models	Week 14
Final Review and Submission	Conduct a comprehensive review of the project, incorporating feedback and finalizing the thesis document. Prepare for Submission	Week 15

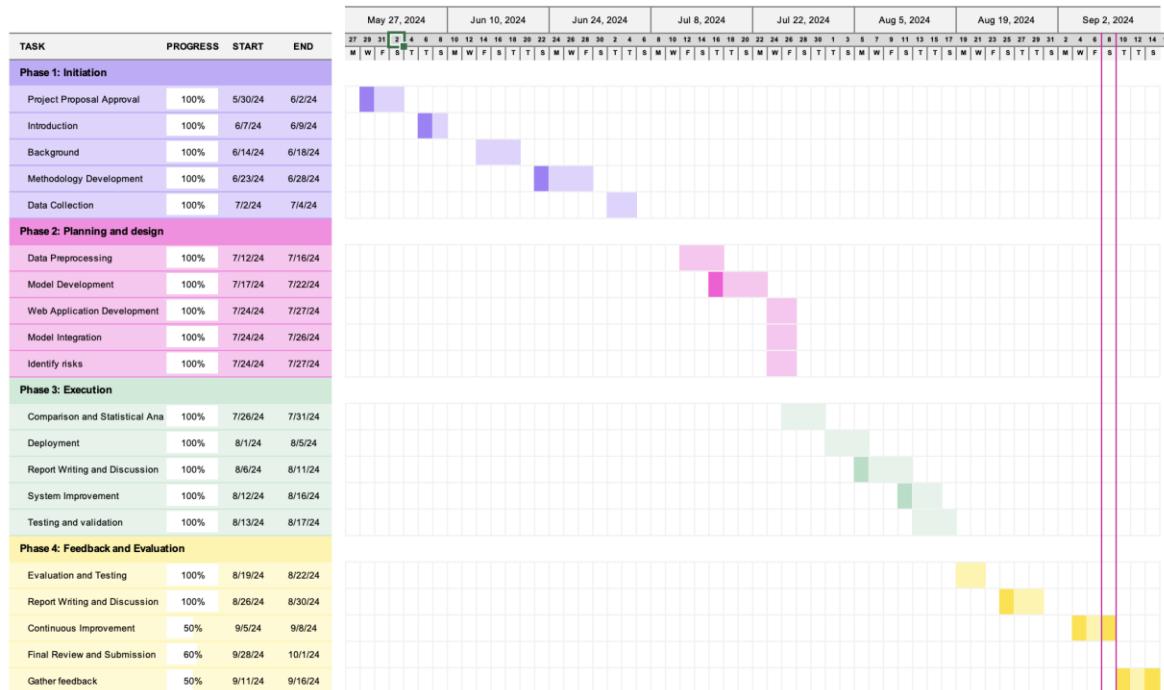
Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Developing a Data-Driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Project start: 30/05/2024

Project lead Dr. Celestine Iwendi

Display week: 1



Insert new rows ABOVE this one

Appendix – Data Analysis Code

1. Data Importation

```
#import system files
import sys
# !{sys.executable} -m pip install pandas
```

Python


```
#load libraries
import pandas as pd
import math
```

Python


```
#load the activity dataset
df1 = pd.read_csv("/Users/whiteegg/Documents/GitHub/Project/FinalSemesterProject/ehact_2014.csv")
df1
```

Python


```
#load respondent dataset
df2 = pd.read_csv("/Users/whiteegg/Documents/GitHub/Project/FinalSemesterProject/ehresp_2014.csv")
df2
```

Python


```
#load weight dataset
df3 = pd.read_csv("/Users/whiteegg/Documents/GitHub/Project/FinalSemesterProject/ehwgts_2014.csv")
df3
```

Python

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

2. Data

Exploration

```
#filter bmi, height and weight
df = df2[["erbmri", "euhgt", "euwgt"]]
df
```

Python


```
#check information about the dataset
df.info()
```

Python


```
#check the distribution of the dataset
df.describe()
```

Python


```
#remove duplicate
df = df.drop_duplicates()
```

Python


```
#check the dataset after duplicate removal
df.info()
```

Python


```
#filter heights less than 21 and weight less than 5
df = df[(df.euhgt > 21) & (df.euwgt > 5)]
df
```

3. Data Cleaning

4. Linear

Interpolation

```
#Check resulting dataset
df.info()
```

Python


```
#check for na values
df.isna().value_counts()
```

Python


```
#drop na values
df = df.dropna(thresh=len(df.columns) // 2)
```

Python


```
#check result after na was removed
df.info()
```

Python


```
#recheck the distribution of the dataset
df.describe()
```

Python


```
#check the correlations of the features in the dataset
correlation_matrix = df.corr()
print(correlation_matrix)
```

Python

5. Data Visualisation

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
#Apply Linear Regression to check intercept and coefficient

from sklearn.linear_model import LinearRegression

X = df[["euhgt", "euwgt"]]
y = df["erbmi"]

model = LinearRegression()
model.fit(X, y)

# Check the coefficients
print(f"Intercept: {model.intercept_}")
print(f"Coefficients: {model.coef_}")
```

Python

```
#import libraries
import matplotlib.pyplot as plt
import seaborn as sns

# List of columns to check for outliers
columns = ['erbmi', 'euhgt']

# Create a figure and axis for each boxplot
plt.figure(figsize=(15, 5))

for i, column in enumerate(columns):
    plt.subplot(1, len(columns), i + 1)
    sns.boxplot(x=df[column])
    plt.title(f'Boxplot of {column}')
```

```
#import libraries
import matplotlib.pyplot as plt
import seaborn as sns

#plot histogram to display the distribution of BMI
plt.figure(figsize=(10, 6))
sns.histplot(df['erbmi'], bins=30, kde=True, color='skyblue')
plt.title('Distribution of BMI')
plt.xlabel('BMI')
plt.ylabel('Frequency')
plt.savefig("Distribution of BMI")
plt.show()
```

Python

```
#plot scatter to display the relationship of Height and BMI

plt.figure(figsize=(10, 6))
sns.scatterplot(x='euhgt', y='erbmi', data=df, color='blue')
plt.title('Height vs BMI')
plt.xlabel('Height (in inches)')
plt.ylabel('BMI')
plt.savefig("Height vs BMI")
plt.show()
```

Python

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
#plot scatterplot to display the relationship of Weight and BMI
plt.figure(figsize=(10, 6))
sns.scatterplot(x='euwgt', y='erbm1', data=df, color='green')
plt.title('Weight vs BMI')
plt.xlabel('Weight (in pounds)')
plt.ylabel('BMI')
plt.savefig("Weight vs BMI")
plt.show()

#plot jointplot to display the relationship of Height, Weight and BMI
sns.jointplot(x='euhgt', y='euwgt', data=df, kind='scatter', hue='erbm1', palette='coolwarm', height=8)
plt.title('Joint Plot of Height and Weight colored by BMI')
plt.xlabel('Height (in inches)')
plt.ylabel('Weight (in pounds)')
plt.savefig("Joint Plot of Height and Weight colored by BMI")
plt.show()

#plot pairplot to display the relationship of Height, Weight and BMI
sns.pairplot(df[['euhgt', 'euwgt', 'erbm1']], kind='reg', diag_kind='kde', height=3)
plt.suptitle('Pairwise Relationships', y=1.02)
plt.savefig("Pairwise Relationships")
plt.show()

df.bmi_category.value_counts()

from sklearn.preprocessing import StandardScaler
# Normalization of height and weight
scaler = StandardScaler()
df[['euhgt', 'euwgt']] = scaler.fit_transform(df[['euhgt', 'euwgt']])

print(df.describe())

# Count the number of occurrences in each BMI category
bmi_counts = df['bmi_category'].value_counts()

# Create a bar chart
plt.figure(figsize=(8, 6))
sns.barplot(x=bmi_counts.index, y=bmi_counts.values, palette='viridis')

# Add titles and labels
plt.title('Distribution of BMI Categories', fontsize=16)
plt.xlabel('BMI Category', fontsize=14)
plt.ylabel('Number of Individuals', fontsize=14)
plt.savefig("Distribution of BMI Categories")
# Show the plot
plt.show()
```

6. Machine Learning and Model Evaluation

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.svm import SVR

# Define the models
models = {
    'Linear Regression': LinearRegression(),
    'Ridge Regression': Ridge(),
    'Lasso Regression': Lasso(),
    'Decision Tree': DecisionTreeRegressor(),
    'Random Forest': RandomForestRegressor(),
    'Support Vector Machine': SVR(),
    'Gradient Boosting': GradientBoostingRegressor()
}

# Split the data into training and test sets
X = df[['euhgt', 'euwgt']]
y = df['erbmri']
y_category = df['bmi_category'] # Actual BMI categories

X_train, X_test, y_train, y_test, y_train_category, y_test_category = train_test_split(X, y, y_category, test_size=0.2, random_state=42)

# Initialize dictionaries to store metrics
accuracy_scores = {}
precision_scores = {}
recall_scores = {}
f1_scores = {}

# Function to categorize BMI
def categorize_bmi(bmi):
    if bmi < 18.5:
        return 'Underweight'
    elif 18.5 <= bmi < 24.9:
        return 'Normal Weight'
    elif 25 <= bmi < 29.9:
        return 'Overweight'
    elif bmi >= 30:
        return 'Obese'
    else:
        return 'Invalid'

# Train and evaluate each model
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    # Convert the predicted BMI to categories
    y_pred_category = [categorize_bmi(bmi) for bmi in y_pred]

    # Calculate metrics
    accuracy_scores[name] = accuracy_score(y_test_category, y_pred_category)
    precision_scores[name] = precision_score(y_test_category, y_pred_category, average='weighted')
    recall_scores[name] = recall_score(y_test_category, y_pred_category, average='weighted')
    f1_scores[name] = f1_score(y_test_category, y_pred_category, average='weighted')

# Display the results
print("Model Performance Metrics:")
for name in models.keys():
    print(f"\n{name}:")
    print(f"  Accuracy: {accuracy_scores[name]}")
    print(f"  Precision: {precision_scores[name]}")
    print(f"  Recall: {recall_scores[name]}")
    print(f"  F1-Score: {f1_scores[name]}")
```

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

7. Statistical Analysis

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import pandas as pd

# Ensure that BMI categories are correctly assigned
df['bmi_category'] = df['erbmi'].apply(categorize_bmi)

# Create a DataFrame for the Tukey HSD test
data_for_tukey = df[['euwgt', 'bmi_category']]
data_for_tukey = data_for_tukey.dropna() # Drop any missing values if necessary
```

Python

```
tukey = pairwise_tukeyhsd(endog=data_for_tukey['euwgt'],
                           groups=data_for_tukey['bmi_category'],
                           alpha=0.05)

# Print the results
print(tukey)
```

Python

```
tukey.plot_simultaneous()
plt.title('Tukey HSD Test for Weight across BMI Categories')
plt.show()
```

Python

Appendix – Web Application Code

1. App Model

```
from django.db import models
from django.contrib.auth.models import User

class Activity(models.Model):
    user = models.ForeignKey(User, on_delete=models.CASCADE)
    activity_type = models.CharField(max_length=100)
    duration = models.IntegerField() # duration in minutes
    date = models.DateField(auto_now_add=True)

    def __str__(self):
        return f"{self.user.username} - {self.activity_type} for {self.duration} minutes"

class Recommendation(models.Model):
    user = models.ForeignKey(User, on_delete=models.CASCADE)
    recommended_activity = models.CharField(max_length=100)
    date = models.DateField(auto_now_add=True)

    def __str__(self):
        return f"Recommendation for {self.user.username}: {self.recommended_activity}"

class Feedback(models.Model):
    name = models.CharField(max_length=100, blank=True)
    email = models.EmailField(blank=True)
    comments = models.TextField()
    created_at = models.DateTimeField(auto_now_add=True)

    def __str__(self):
        return f"Feedback from {self.name or 'Anonymous'}"
```

2. App Views

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
views.py > get_random_images
import os
import random
import pickle
import numpy as np
from django.conf import settings
from django.shortcuts import render, redirect
from .forms import BMIPredictionForm, FeedbackForm
from .models import Feedback

# Load the model
def home(request):
    return render(request, 'index.html')

with open('myapp/bmi_model.pkl', 'rb') as file:
    model = pickle.load(file)

def categorize_bmi(bmi):
    if bmi < 18.5:
        return 'Underweight'
    elif 18.5 <= bmi < 25:
        return 'Normal_weight'
    elif 25 <= bmi < 30:
        return 'Overweight'
    elif 30 <= bmi < 40:
        return 'Obese'
    else:
        return 'Morbid_Obesity'

def get_random_images(category, subcategory, num_images=5):
    """Get random images from the specified category and subcategory folder."""
    image_folder = os.path.join(settings.STATICFILES_DIRS[0], 'recommendations', category, subcategory)

    if not os.path.exists(image_folder):
        print(f"Directory not found: {image_folder}")
        return []

    images = os.listdir(image_folder)

    if not images:
        return []

    selected_images = random.sample(images, min(num_images, len(images)))

    return [os.path.join(settings.STATIC_URL, 'recommendations', category, subcategory, image) for image in selected_images]
```

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
def calculate_bmi(request):
    if request.method == 'POST':
        form = BMIPredictionForm(request.POST)
        feedback_form = FeedbackForm(request.POST)

        if form.is_valid():
            height_feet = form.cleaned_data['height']
            weight_kg = form.cleaned_data['weight']

            height_inches = round((height_feet * 12),2)
            weight_pounds = round((weight_kg * 2.20462),2)

            X = np.array([[height_inches, weight_pounds]])

            bmi = round(model.predict(X)[0],2)

            category = categorize_bmi(bmi)

            diet_images = get_random_images(category, 'diet', num_images=5)
            fitness_images = get_random_images(category, 'fitness', num_images=5)

            if feedback_form.is_valid():
                feedback_form.save()
                return redirect('thank_you')

            return render(request, 'bmi_result.html', {
                'bmi': bmi,
                'category': category,
                'diet_images': diet_images,
                'fitness_images': fitness_images,
                'feedback_form': feedback_form
            })
        else:
            form = BMIPredictionForm()
            feedback_form = FeedbackForm()

    return render(request, 'calculate_bmi.html', {'form': form, 'feedback_form': feedback_form})

def thank_you(request):
    return render(request, 'thank_you.html')

def feedback_view(request):
    if request.method == 'POST':
        feedback_form = FeedbackForm(request.POST)
        if feedback_form.is_valid():
            feedback_form.save()
            return redirect('thank_you') # Redirect to thank you page after submitting feedback
    else:
        feedback_form = FeedbackForm()

    return render(request, 'feedback.html', {'feedback_form': feedback_form})
```

3. AppUrls

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
from django.urls import path
from . import views

urlpatterns = [
    path('', views.home, name='home'),
    path('calculate_bmi/', views.calculate_bmi, name='calculate_bmi'),
    path('thank-you/', views.thank_you, name='thank_you'),
    path('feedback/', views.feedback_view, name='feedback'),
]
```

4. App Form

```
from django import forms
from .models import Feedback

class BMIPredictionForm(forms.Form):
    height = forms.FloatField(label='Height (in feet)')
    weight = forms.FloatField(label='Weight (in Kg)')

class FeedbackForm(forms.ModelForm):
    class Meta:
        model = Feedback
        fields = ['name', 'email', 'comments']
```

5. App Base Html File

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
<!-- base.html -->
{% load static %}
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>{{ block title }}My Site{{ endblock }}</title>
    <link href="https://fonts.googleapis.com/css2?family=Roboto:wght@400;500;700&display=swap" rel="stylesheet">
    <link rel="stylesheet" type="text/css" href="{{ static 'css/scripts.css' }}>
    {% block extra_head %}
    {% endblock %}
</head>
<body>
    <div class="background">
        <div class="container">
            <nav>
                <a href="{{ url 'home' }}" class="home-link">Home</a>
            </nav>
            {% block content %}
            <!-- Default content goes here, if any -->
            {% endblock %}
        </div>
        <footer>
            {% block footer %}
            <p>&copy; <a href="https://www.linkedin.com/in/rukayat-balogun-983204197/" target="_blank" style="color: inherit; text-decoration: none;">{{ block footer }}
            {% endblock %}
        </footer>
    </div>
    <script src="{{ static 'js/scripts.js' }}></script>
    {% block extra_scripts %}
    {% endblock %}
</body>
</html>
```

6. BMI result code

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
plates > bmi_result.html > section.recommendations > div.recommendation-section
2   {% extends 'base.html' %}

3
4   {% block title %}BMI Result{% endblock %}

5
6   {% block content %}
7     <header class="header">
8       <h1>BMI Result</h1>
9       <p>Your BMI is: <strong>{{ bmi }}</strong></p>
10      <p>Category: <span class="category">{{ category }}</span></p>
11    </header>

12
13    <section class="recommendations">
14      <div class="recommendation-section">
15        <h2>Dietary Recommendations</h2>
16        <div class="cards">
17          {% for image_url in diet_images %}
18            <div class="card">
19              
20            </div>
21            {% empty %}
22            <p>No diet images available for this category.</p>
23            {% endfor %}
24          </div>
25        </div>
26        <div class="recommendation-section">
27          <h2>Fitness Recommendations</h2>
28          <div class="cards">
29            {% for image_url in fitness_images %}
30              <div class="card">
31                
32              </div>
33              {% empty %}
34              <p>No fitness images available for this category.</p>
35              {% endfor %}
36            </div>
37          </div>
38        </section>

39        <div class="button-group">
40          <a class="button calculate-again" href="{% url 'calculate_bmi' %}">Calculate Again</a>
41          <a class="button feedback-button" href="{% url 'feedback' %}">Give Feedback</a>
42        </div>
43      {% endblock %}
```

7. Calculate BMI HTML code

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
<!-- calculate_bmi.html -->
{% extends 'base.html' %}

{% block title %}BMI Calculator{% endblock %}

{% block content %}
<div class="calculator-container">
    <h1>BMI Calculator</h1>
    <form method="post" id="bmiForm">
        {% csrf_token %}
        <div class="form-group">
            <label for="height">Height (in feet):</label>
            {{ form.height }}
        </div>
        <div class="form-group">
            <label for="weight">Weight (in Kg):</label>
            {{ form.weight }}
        </div>
        <button type="submit" class="calculate-btn">Calculate BMI</button>
    </form>
</div>
{% endblock %}
```

8. Index HTML code

```
dates / index.html / ...
<!-- index.html -->
{% extends 'base.html' %}

{% block title %}Home - Personalized Fitness for Obesity{% endblock %}

{% block content %}
<div class="hero-section">
    <div class="hero-content">
        <h1>Welcome to Your Personalized Fitness Journey</h1>
        <p>Our mission is to empower obese and sedentary individuals with data-driven personalized fitness plans that cater to your unique needs.</p>
        <a href="{% url 'calculate_bmi' %}" class="cta-button">Get Started with BMI Calculation</a>
    </div>
</div>

<section class="features-section">
    <div class="feature">
        <h2>Personalized Workouts</h2>
        <p>Tailored workout plans to match your fitness level and goals, helping you transition to a healthier lifestyle.</p>
    </div>
    <div class="feature">
        <h2>Nutrition Guidance</h2>
        <p>Custom meal plans and dietary recommendations that align with your personal health data.</p>
    </div>
    <div class="feature">
        <h2>Progress Tracking</h2>
        <p>Monitor your progress with our comprehensive tracking system, ensuring you're on the right path.</p>
    </div>
</section>
{% endblock %}
```

9. Thank you HTML code

University of Bolton

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

```
plates / thank_you.html / ...
{% extends 'base.html' %}

{% block title %}Thank You{% endblock %}

{% block content %}
    <div class="thank-you">
        <h1>Thank You for Your Feedback!</h1>
        <p>We appreciate your input and will use it to improve our services.</p>
        <a href="{% url 'home' %}" class="button">Return to Home</a>
    </div>
{% endblock %}
```

10. Feedback HTML code

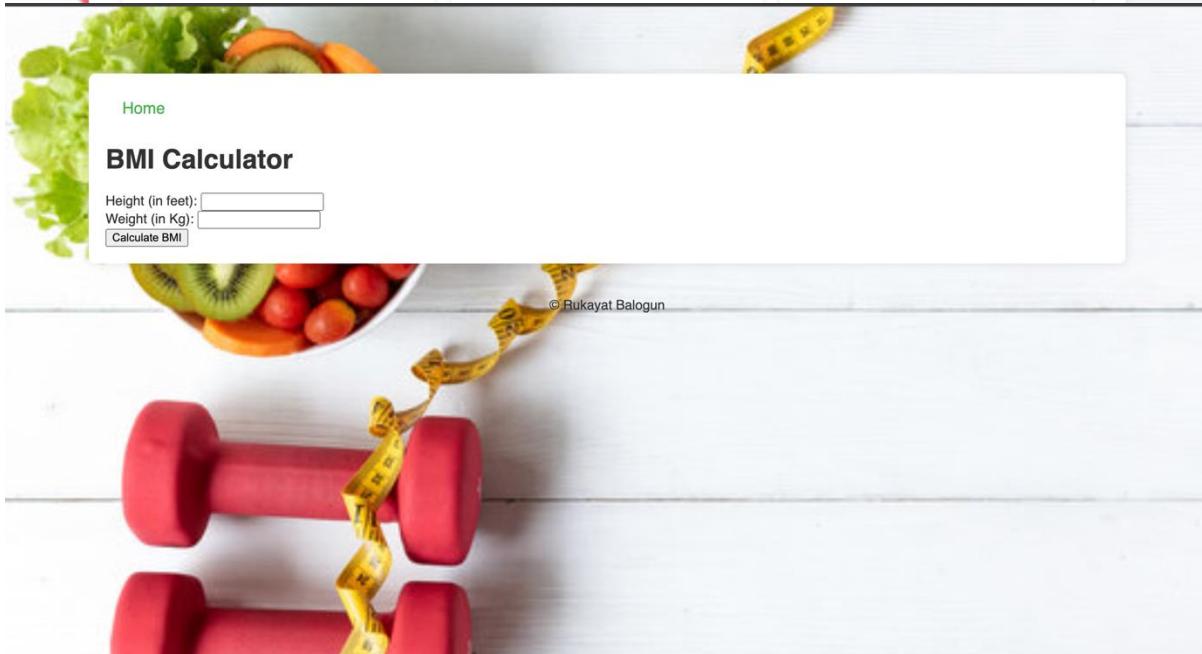
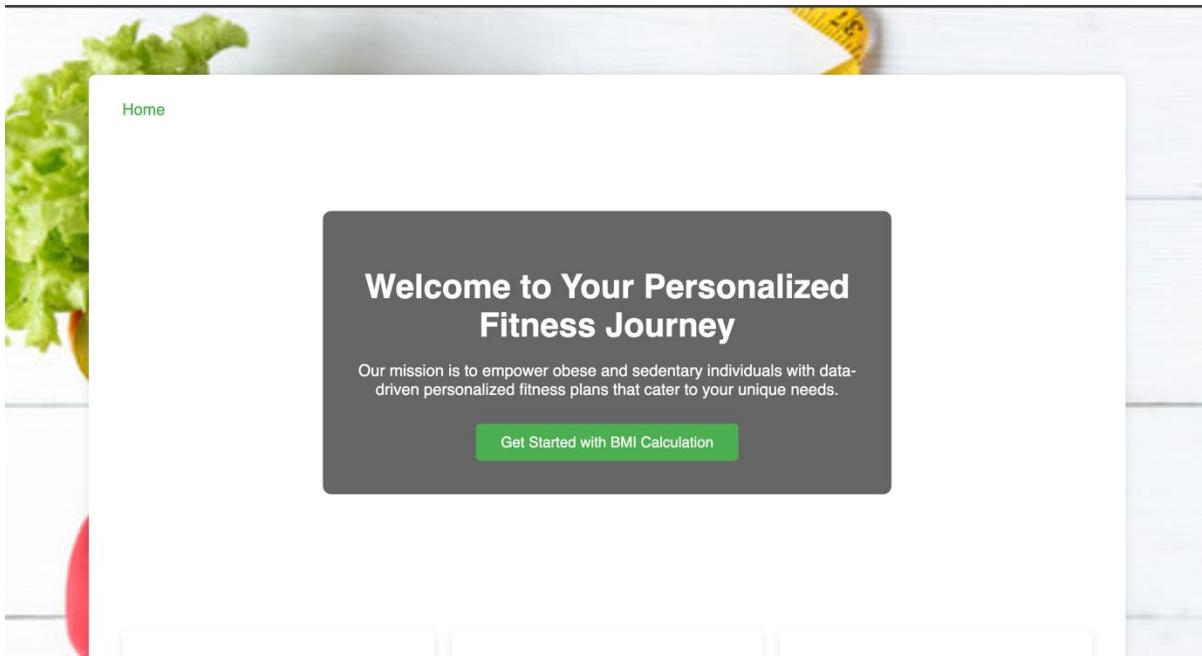
```
<!-- feedback.html -->
{% extends 'base.html' %}

{% block title %}Give Feedback{% endblock %}

{% block content %}
    <div class="feedback-page">
        <h1>We'd Love Your Feedback</h1>
        <p>Your feedback helps us improve our services. Please take a moment to let us know how we're doing.</p>
        <form method="post">
            {% csrf_token %}
            <div class="form-group">
                <label for="name">Name:</label>
                {{ feedback_form.name }}
            </div>
            <div class="form-group">
                <label for="email">Email:</label>
                {{ feedback_form.email }}
            </div>
            <div class="form-group">
                <label for="comments">Comments:</label>
                {{ feedback_form.comments }}
            </div>
            <button type="submit" class="button feedback-submit-button">Submit Feedback</button>
        </form>
    </div>
{% endblock %}
```

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

Appendix – Web Application Frontend



Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals

BMI Result

Your BMI is: **32.12**
Category: **Obese**

Dietary Recommendations



Fitness Recommendations



We'd Love Your Feedback

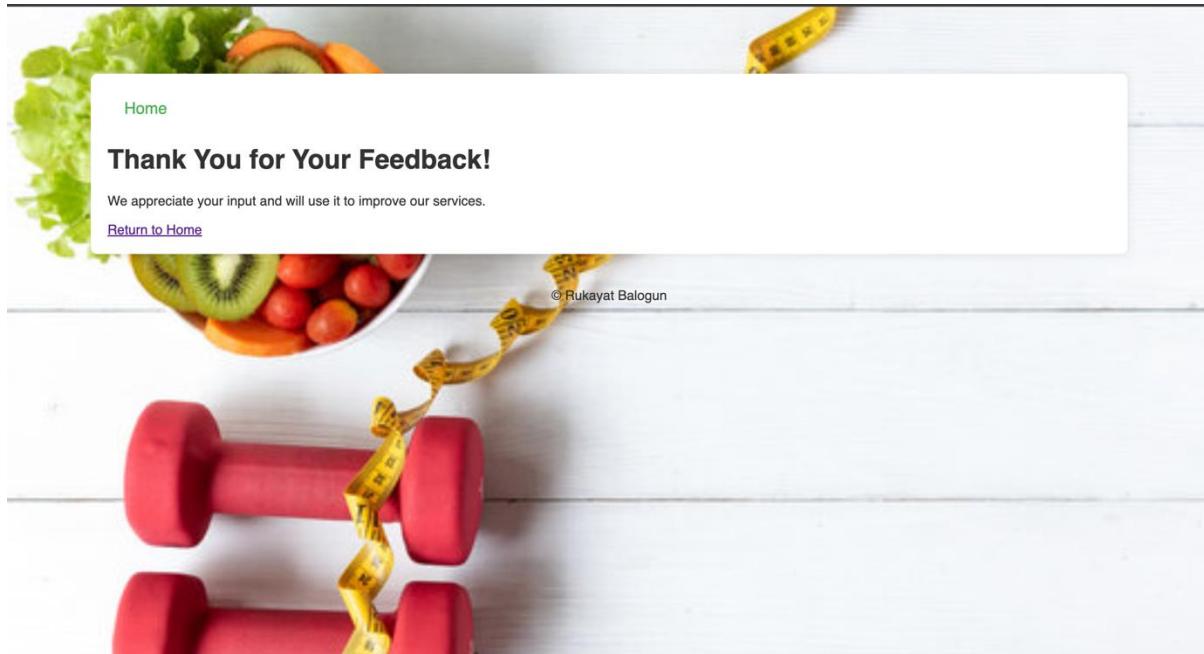
Your feedback helps us improve our services. Please take a moment to let us know how we're doing.

Name: Email:

Comments:

© Rukayat Balogun

Developing a Data-driven Personalized Fitness Web Application for Obese and Sedentary Individuals



University of Bolton