A blue sign with white text

Description automatically generated

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

RUKAYAT BALOGUN

2311143

DR CELESTINE IWENDI

MASTERS DEGREE

DATA ANALYTICS AND TECHNOLOGY

UNIVERSITY OF BOLTON

11/06/2024

Table of Contents

[Chapter 1: Introduction 4](#_Toc172968671)

[1.0 Introduction 4](#_Toc172968672)

[1.1 Background of Study 5](#_Toc172968673)

[1.2 Health Implications of Obesity and Sedentary Lifestyles 7](#_Toc172968674)

[1.3 Statement of the Problem 8](#_Toc172968675)

[1.4 Purpose of the Research 9](#_Toc172968676)

[1.5 Research Aims 10](#_Toc172968677)

[1.6 Research Questions 10](#_Toc172968678)

[1.7 Research Objectives 11](#_Toc172968679)

[1.8 Relevance and Importance of the study 11](#_Toc172968680)

[1.9 Scope of the Study 12](#_Toc172968681)

[Chapter Two – Literature Review 12](#_Toc172968682)

[2.1 Introduction 12](#_Toc172968683)

[2.2 Current Public Health Guidelines and Interventions 26](#_Toc172968684)

[2.3 Constraints of Conventional Diet and Fitness Programmes 27](#_Toc172968685)

[2.3.1 Limitations of Traditional Diet and Fitness Plans: 32](#_Toc172968686)

[2.3.2 Public Health initiatives Targeting Obesity and Sedentary Lifestyle 33](#_Toc172968687)

[2.3.3 Challenges in Personalized Health Recommendations 34](#_Toc172968688)

[2.4 Overcoming these challenges requires: 36](#_Toc172968689)

[Chapter 3 – Methodology 37](#_Toc172968690)

[3.1Introduction 37](#_Toc172968691)

[3.2 Research Philosophy 37](#_Toc172968692)

[3.3 Proposed Workflow 38](#_Toc172968693)

[3.3.1 Data Collection 38](#_Toc172968694)

[3.3.2 Data Preprocessing 39](#_Toc172968695)

[3.3.3. Machine Learning Model Development 39](#_Toc172968696)

[3.3.4 Web Application Development using Django 39](#_Toc172968697)

[3.3.5 Integration of Machine Learning using Django 39](#_Toc172968698)

[3.3.6 Evaluation and Testing 40](#_Toc172968699)

[3.4 Data Analysis Plan 40](#_Toc172968700)

[3.5. Machine Learning 40](#_Toc172968701)

[Introduction 40](#_Toc172968702)

[3.5.1 Linear Regression 40](#_Toc172968703)

[3.5.2 Ridge Regression 41](#_Toc172968704)

[3.5.3 Lasso Regression 41](#_Toc172968705)

[3.5.4 Decision Tree Regressor 41](#_Toc172968706)

[3.5.5 Random Forest Regressor 41](#_Toc172968707)

[3.5.6 Support Vector Machine 42](#_Toc172968708)

[3.5.7 Gradient Boosting 42](#_Toc172968709)

[3.6 Ethical Considerations 42](#_Toc172968710)

[Reference 42](#_Toc172968711)

# 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.

A robot holding a tablet

Description automatically generated

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 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, no date). 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-bring. The study highlights the correlation between obesity and a group of long-lasting illnesses, specifically cardiovascular diseases (CVD) and type 2 diabetic mellitus (T2DM). Hypertension, a significant risk factor for cardiovascular disease (CVD), was observed to be more common among females(Suanrueang, 2024). 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 highlights a greater prevalence of additional obesity-related illness such as hypothyroidism and arthritis, particularly among the elderly population (Ghosh et al., 2023).

In the research conducted by Ozsoy, he also emphasises the adverse effects of sedentary on health and well-being Click or tap here to enter text. 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., no date a). This underscores the 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 analyze 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 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:

1. To develop a user interface that is easy to navigate and accessible for all users.
2. To ensure the web application is inclusive by integrating necessary accessibility features.
3. To identity and implement features that effectively engage users
4. To access the effectiveness of the fitness assistance web application in increasing physical activity levels among obese and sedentary individuals.

## 1.6 Research Questions

This research aims to answer the following questions:

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

## 1.7 Research Objectives

1. Develop and implement a user-friendly interface for the personalized fitness web application tailored to the needs of obese and sedentary individuals.
2. Ensure the web application adheres to current web accessibility standards to provide an inclusive experience for users with varying needs.
3. Integrate machine learning algorithms to provide personalized fitness and dietary recommendations based on user input and historical data.
4. Analyse user engagement data to identify which features are most frequently used and highly rated by users.
5. 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, 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.

# 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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S/N | References | Research Focus | Relevance to this thesis | Methodology | Strengths | Limitations | Knowledge Contribution |
| 1 | Singhania and Reddy (2024) | Big Data Analytics and predictive modelling for chronic disease management | 4 | Review and analysis of big data applications; identification of key technologies and solutions | Comprehensive review of big data applications; identification of key technologies and solutions | General focus on chronic diseases; limited focus on personalized fitness applications | Highlights the potential of big data in personalizing health care and predictive analytics |
| 2 | Bhowmik, S. et al (2024) | Machine Learning and virtual gyms for healthcare | 5 | Case studies and implementation of online gym systems | Practical application of machine learning in fitness; detailed case studies | Focus on virtual gyms may not address broader aspects of personalised fitness web applications | Illustrates the practical mplementation of machine learning in virtual gym environments for personalized health |
| 3 | Hernendez, J., & Lopez, M. (2023) | 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 | Multiphase experimental design, prototyping, expert review and iterative design. | Emphasis on personalization and diversity in users needs; comprehensive metholodology | Focus on persuasive design, less on direct health outcomes. | Provides insights into using persuasive design for personalized health interventions |
| 5 | Benton, R., et al. (2023) | 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 in the Cloud | 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-centered design in fitness recommendation systems |
| 7 | Hu et al., (2024) | Wearable Technology for personalized exercise recommendations | 5 | Development and testing or wearable technoloyg 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 | Singh, K., et al (2024) | 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 models in health care |
| 11 | 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 into preparing large datasets for AI applications in health research |
| 12 | 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 apllications 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 |
| 13 | Smith, P., et al (2023) | 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 |
| 14 | Dergaa et al., (2024) | AI-driven methods to improve physical activity | 4 | Systematic review of AI applications in physical activity enhancement | Comprehensive review; focus on AI applications | General focus; less on personalized fitness applications | Summarizes AI-driven strategies to increase physical activity, providing, providing a broach overview of AI applications |
| 15 | 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 data. | Highlights machine learning’s potential in customizing exercise plans to individuals needs |
| 16 | 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 |
| 17 | Rodriguez 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 |
| 18 | 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 |
| 19 | Oyebode et al., (2023) | Predictive analytics in health and fitness personalization | 4 | Review and analysis of predicitive analytics techniques | Comprehensive review; focus on predictive techniques | Broad focus on health and fitness; less on specific applications | Reviews the role of predictive analytis in personalizing health and fitness interventions |
| 20 | Al 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. | Summerizes the sue of AI to boost physical activity, highlighting various applications |
| 21 | Busnatu 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 | Santos, P., et al (2022) | 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 | Ng, M., et al (2022) | 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 specifc application | Discusses AI’s role in creating personalized health interventions |
| 24 | Huang, B., et al (2024) | Personalized dietary recommendations using knowledge graphs | 5 | Development and testing of knowledge graph-based recommendation systems | Focus on personalization advanced recommendation techniques | Primarily focused on dietary recommendations; less on fitnes | Demonstrates the use of knowledge graphs to enhace dietary recommendations |
| 25 | Park, S., et al (2023) | 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 |
| 26 | Kim, Y., et al (2023) | 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 |
| 27 | Anderson, P., et al (2022) | 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 |
| 28 | Liu, X., et al (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 |
| 29 | Johnson, H et al (2021) | Personalized exercise recommendations using wearable data | 5 | Development and testing of wearable data applications | Focus wear technology; high relevance personalized fitness | Need for long-tern user engagement data technical focus | Examines how wearable data can be used to personalise exercise recommendations |
| 30 | Martinez, D. et al (2023) | 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-industry 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 implementated 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 behaviour.

## 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 behavior 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.

2.4 Public Health Programmes Focusing on Obesity and Inactive Behaviours

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 seld-reported information that may be incorrect or spartial (Tiribelli & Calvaresi, 2024). The implementation of personalised health therapies is further complicated by privacy concerns associated with the collecting and utilication of personal health data (Tiribelli & Calvaresi, 2024).

2.6 Addressing Obstacles in Tailored Health Recommendations

In order 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 in order 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).

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 indiviual’s inclination towards obesity and possible remedies. The utilisation of machine learning in this particular 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 take into account 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 cover 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).

2.12 Dietary Guidelines for Specific Health Conditions

Genetic algoritjms have been utlised 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 take into account 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.

### 2.3.1 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).

### 2.3.2 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:

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

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., 2016b).

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., 2016b)

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 behavior.

### 2.3.3 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, D’Alfonso and Lederman, 2023). Nevertheless, attaining genuine personalisation entails numerous substantial obstacles:

1. Limitations of Traditional One-Size-Fits-All Approaches:
   1. 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).
   2. Lack of motivation: general suggestions frequently overlook personal preferences and motives, resulting in low compliance and a limited long-term effect (Biese, Österwall and Mckeever, 2024).
   3. Inequities: generic recommendations may not adequately cater to the distinct requirements and obstacles encountered by various populations(Polak et al., 2016b).
2. Limitations of Existing Recommendation Systems:
   1. 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., no date b).
   2. Sparse data: numerous health recommendation systems depends 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).
   3. Privacy concerns: this problem arise when personal health data is collected and utilised. Some people may be reluctant to disclose confidential information, which can reduce the impact of personalized suggestions(Tiribelli and Calvaresi, 2024).
3. The Need for Personalized, Context-Aware Health Interventions:
   1. Dynamic health needs: health needs and risks change over time. Personalised advice should be flexible enough to accommodate these changes and shifting situations(Rummery, Lawrence and Russell, 2023).
   2. Context matters: lifestyle choices, social factors that determine health, and environmental exposures all have an impact on all individuals’ well-being. In order to achieve true effectiveness, personalised recommendations must take into account certain contextual elements(Rummery, Lawrence and Russell, 2023).
   3. Incorporating behaviour modification into the integration process: efficient health interventions necessitate surpassing the mere provision of information. They should provide assistance and encouragement to individuals in order to encourage them to embrace and sustain healthy behaviours(Rummery, Lawrence and Russell, 2023).

## 2.4 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 behaviours 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.

# Chapter 3 – Methodology

## 3.1Introduction

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:

1. Data Collection: collect user data on height and weight. This data is essential for the development of customised fitness and food plans.
2. 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.
3. 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.
4. Model Selection and Training: utlize 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.
5. Model Evaluation: assess the performance of each model by utilising metrics such as Mean Squared Error (MSE), accuracy, precision, and recall in order to determine the most suitable model for deployment.
6. Web Application Development: constructing the user interface and underlying functionality of the web application utilising Python Django.
7. Integration of Machine Learning Models: incorporating the created machine learning models into the web application.
8. Testing and validation: conducting tests on the web application to ensure its proper functionality and verifying the precision of the machine learning models.
9. 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 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 behaviours, preferences, and interactions with technology with technology have an impact on fitness and health, as emphasised in the research conducted by Kuru et al. (2023). 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 3.1. This workflow encompasses data collection, preprocessing, model development, web application development using Django, integration of machine learning models, and continuous evaluation and improvement.

User Input

Integration of Machine Learning using Django

Web Application Development using Django

Dietary and fitness recommendation

Evaluation and Testing

Figure 3.3: Proposed Workflow for the web application using Django

### 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 on 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.

**ERBMI (Body Mass Index):** Calculated from respondents’ height(eught) and weight,( euwgt) providing a crucial metric for personalizing fitness plans

**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 behaviours and health metrics. This dual approach enhances the accuracy and personalization of the fitness recommendations provided by the web application.

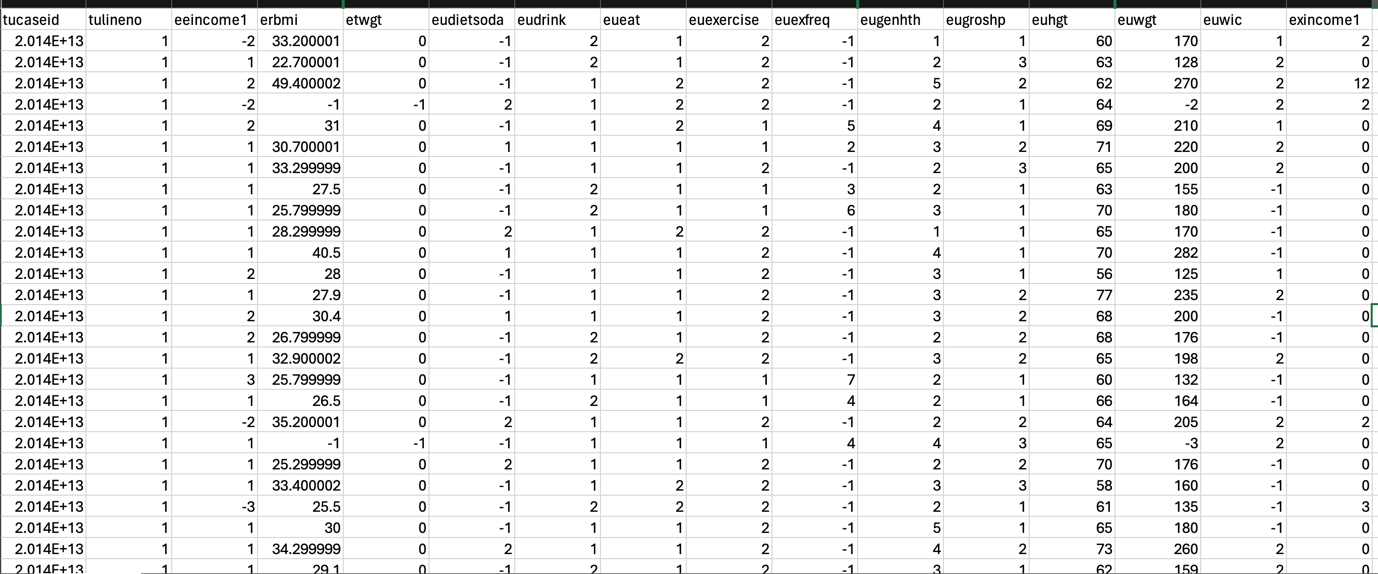


Figure 3.3.1: Snapshot of the data

### 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: is the process of verifying the accuracy and integrity of data entries, addressing any missing values, eliminating duplicate entries, and rectifying inconsistencies.

Data transformation: is the process of normalising and standardising data in order 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 Body Mass Index (BMI) using height and weight data.

Height (in inches) = height(in feet) \* 12

Weight (in pounds) = weight (in kg) \* 2.20462

BMI = Weight (in pounds)

(Height (in inches))2

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: given that height and weight data can vary greatly, normalization or standardization techniques will be applied. This ensures that features contribute equally to the analysis and modeling 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:

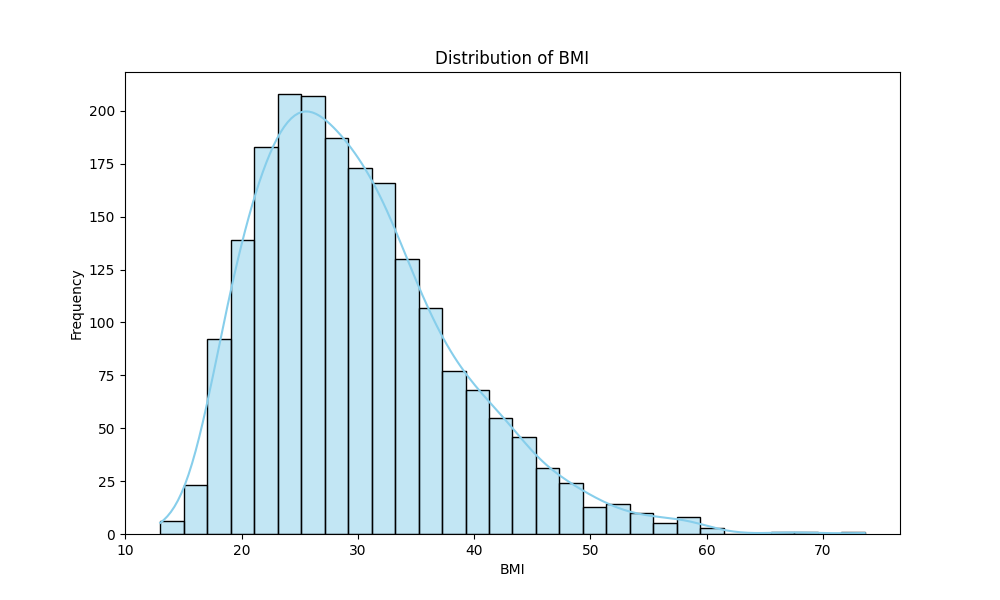
1. Distribution of BMI: Figure 3.3.3.1 presents a histogram of the Body Mass Index (BMI) distribution within the dataset used for developing the personalized fitness web application. The distribution appears to follow a normal distribution with a peak around the average BMI value. This distribution indicates 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 these individuals..

Figure 3.3.3.1: Distribution of BMI

1. 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 (euwgt) 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 relationshops.

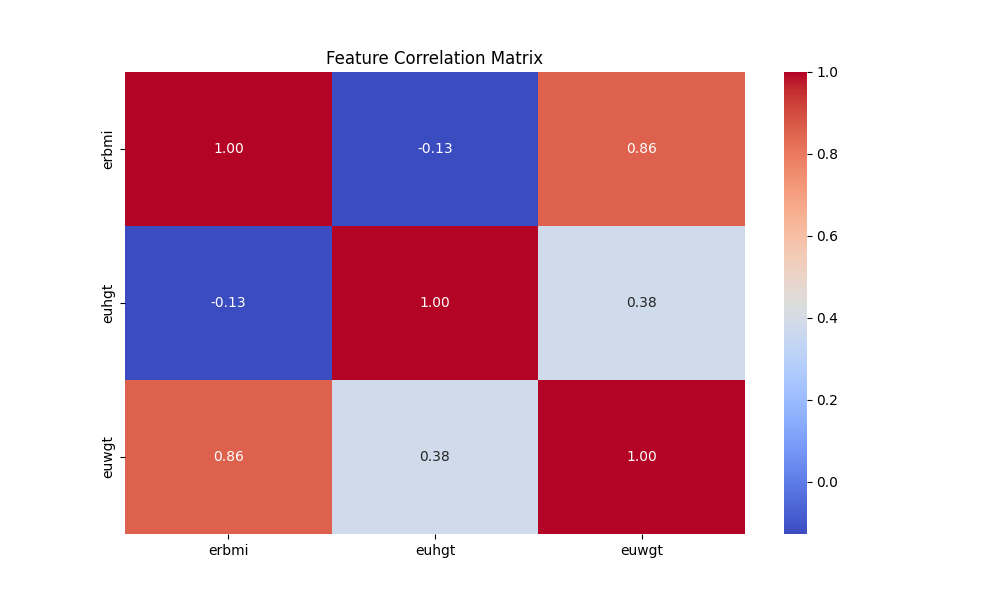


Figure 3.3.3.2: Feature Correlation Matrix

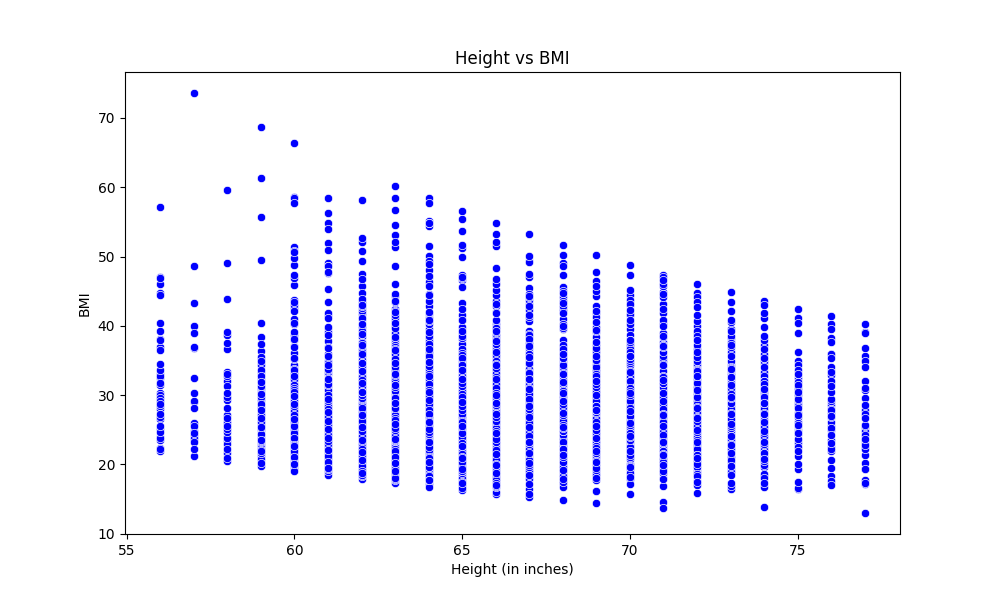
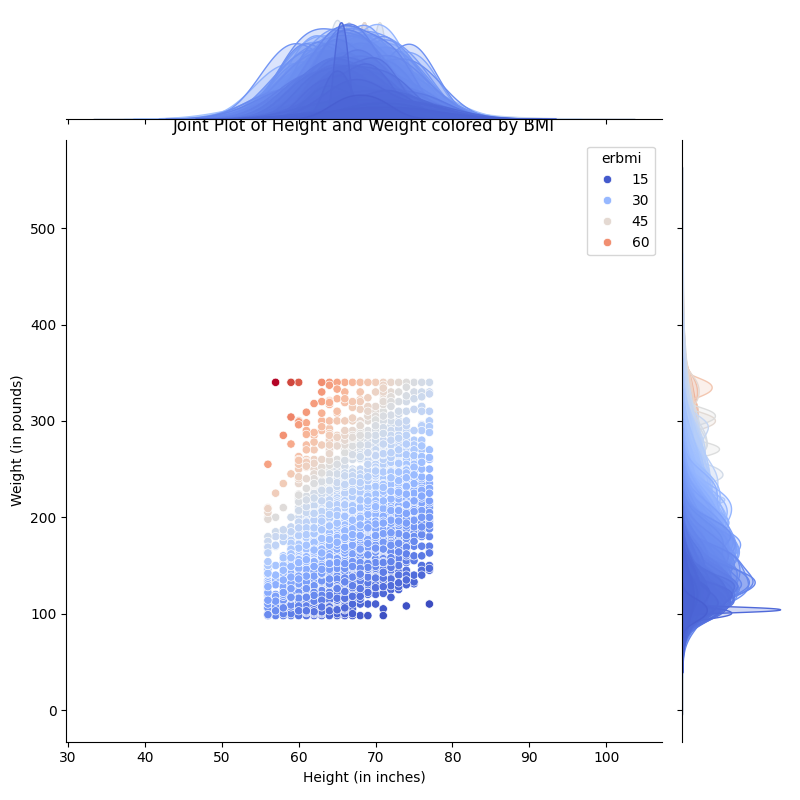
1. Height vs BMI: Figure 3.3.3.3 depicts a scatter plot illustrating the relationship between height(in inches) and BMI. The plot reveals a slight negative trends, suggesting that taller individuals tend to have a lower BMI when weight is held constant. This visualization is useful for identifying any 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

1. Joint Plot of Height and Weight Coloured by BMI: Figure 3.3.3.4 displays a joint plots 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 heigh 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

1. Pairwise Relationship: Figure 3.3.3.5 presents a pairplot that displays pairwise relationships between key features in the dataset, including height (euhgt), weight (euwgt), and BMI (erbmi). Each scatterplot within the pairplot 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.

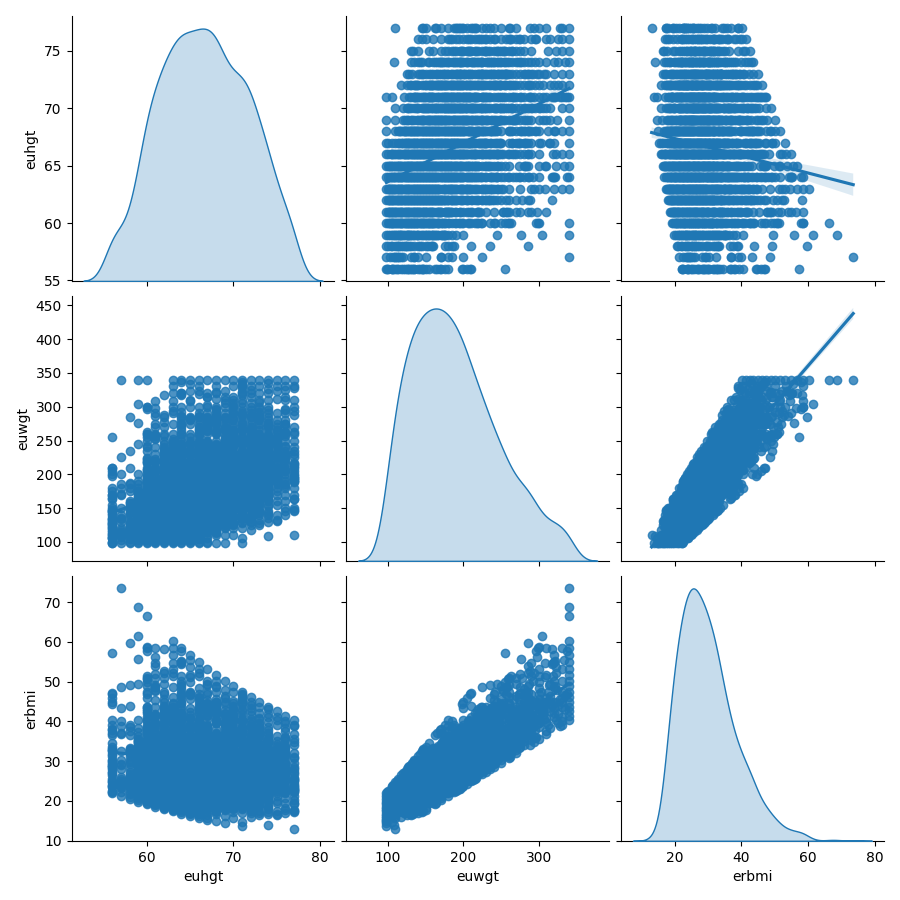


Figure 3.3.2.5: Pairwise Relationship

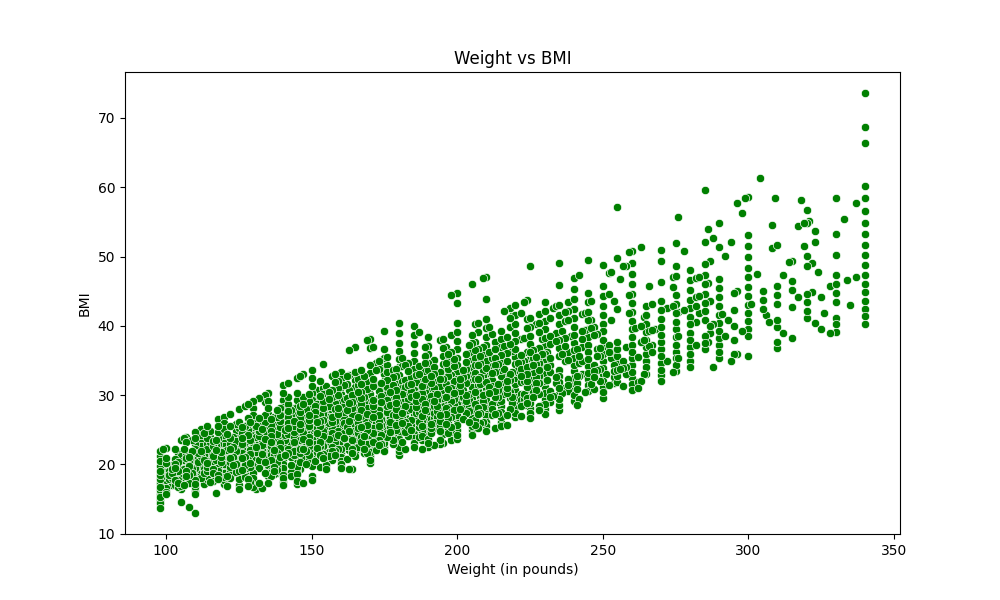
1. Weight vs BMI: Figure 3.3.3.6 shows a scatter plot that illustrates the relationship between weight (in pounds) and BMI. The plot demonstrates a strong positive linear relationship, where an increase in weight is directly associated with an increase in BMI. This is consistent with the BMI calculation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

Post-Hoc Test: if ANOVA reveals significant differences between groups, post-hoc tests like Tukey’s HSD will be uded 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: for models like Random Forest and Gradient Boosting, grid search combined with cross-validation will be used to find the optimal set of hyperparameters.

### 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).

Zanevych 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 (Zanevych, 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 (Zanevych, 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 (Zanevych, 2024).

Django app is used to achieve the following:

1. Enhancing the User Interface:
   1. 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.
   2. Front-End Integration: Enables the utilisation of CSS and JavaScript frameworks to further enhance the user interface.
2. Guaranteeing Accessibility:
   1. Django’s extensibility: enables the incorporation of accessibility plugins and packages that enhance accessibility.
3. Assessing the User-Friendliness:
   1. 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:
   1. Usage Analytics: Django enables the tracking of user behaviour and preferences providing valuable insights for making informed decisions on future development.
5. Enhancing levels of physical activity:
   1. 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.
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 fits 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:

1. 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.
2. Model Serialization: The trained model was serialized using a format such as ‘pickle’ to enable it to be loaded within the Django environment.
3. 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.
4. 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.

This integration allows users to input their height and weight into the web application vis Django forms and receive instant feedback in the form of tailored fitness and nutrition recommendations.

3.3.7.5 Continuous Improvement

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

1. User Feedback: Collecting feedback from users regarding the accuracy and usefulness of the recommendations provided.
2. Periodic Retraining: Retraining the model with new data as it becomes available to improve accuracy and adapt to changing trends.
3. Algorithm Updates: Enhancing the recommendation algorithms based on feedback and performance metrics to ensure the recommendations remain accurate and relevant.
4. 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.

Algorithm 1: Personalized Fitness and Dietary Recommendation Engine

function: ‘generatePersonalizedRecommendations(userData)’

output: ‘Customzed Fitness and Dietary Plan’

procedure:

1. Input:

* Hu ← User height in feet
* Wu ← User weight in kilograms
* Au ← User activity level
* Du ← User dietary preferences

1. BMI Calculation:

BMIu =

1. BMI Categorization:

Categoryu =

1. Term Frequency Calculation:

=

Where:

) is the sum of user υ’s ratings for dietary element di.

is the total number of elements in all rated dietary options

1. Inverse Document Frequency (IDF): calculate the rarity of each preference feature (e.g dietary element, exercise type) across the dataset:

Where:

is the total number of all dietary and physical activity elements in the dataset

is the number of occurrences of the element in the user’s data.

### 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.

## 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 heigh, weight, and other features, with statistical significance tests (p-values) being used to confirm the relevance of predictors.

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.

## 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 has the ability to 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 has the ability to 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 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.

## 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.

## 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.

# 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 keys 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.

### 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.

### 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.

## 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 weigh-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 (‘eught’), weight (‘euwgt’), 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 formular 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 tends 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:

* 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 outlies 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, 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, which 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.

* + 1. 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):

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-preforming 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 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 average 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.

### 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 inputs 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:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Machine Learning Models | R2 | MAE | RMSE | MSE |
| Linear Regression | 0.976326 | 0.847728 | 1.319542 | 1.7411 |
| Ridge Regression | 0.976326 | 0.847713 | 1.319546 | 1.7412 |
| Lasso Regression | 0.975357 | 0.832448 | 1.346274 | 1.8124 |
| Decision Tree Regression | 0.992216 | 0.429798 | 0.756621 | 0.5779 |
| Random Forest Regression | 0.994635 | 0.224768 | 0.628138 | 0.3710 |
| Support Vector Machine | 0.936164 | 1.067960 | 2.166824 | 4.6915 |
| Gradient Boosting Regression | 0.9997460 | 0.314084 | 0.432238 | 0.1868 |

Table 4.4.4: Model Validation

**Gradient Boosting Regression** achieved the highest R2 (0.997460), indicating that it explains approximately 99.75% of the variance in the dataset. 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 R2 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 R2 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 R2 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.

Lasso Regression had slightly lower accuracy compared to Linear and Ridge Regression, with an R2 of 0.975357 and a higher RMSE (1.346274) and MSE (1.8124).

Support Vector Machine was the least effective model, with the lowest R2 (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.

### 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 Tress Regression: Although the Decision Tree Model showed lower MSE than the linear models, it was still prone to overfitting, a s 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 (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 ≤ BMI 24.9
* Overweight: 25 ≤ BMI 29.9
* Obese BMI ≥ 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 oof correctly predicted BMI categories out of the total predictions.
* Precision: The ration 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:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Machine Learning Algorithm | Accuracy | Precision | Recall | F1-score |
| Linear Regression | 0.9231 | 0.9639 | 0.9231 | 0.9352 |
| Ridge Regression | 0.9231 | 0.9651 | 0.9231 | 0.9365 |
| Decision Tree | 0.9658 | 0.9858 | 0.9658 | 0.9725 |
| Random Forest | 0.9658 | 0.9916 | 0.9658 | 0.9767 |
| Support Vector Machine | 0.9744 | 0.9744 | 0.9744 | 0.9744 |
| Lasso Regression | 0.8120 | 0.8202 | 0.8120 | 0.7874 |
| Gradient Boosting | 0.9658 | 0.9741 | 0.9658 | 0.9699 |

The results 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 R**andom 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 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

## 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 each individual 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 provide 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 as 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 Discussion

The findings from the model training, validation and the recommendation system outcomes provide several key insights into the effectiveness of using machine learning for personalized fitness recommendations.

### 4.8.1 Model Effectiveness

The Gradient Boosting Regression model demonstrated superior performance in predicting BMI, which was crucial for generating accurate recommendations. The mode’s ability to capture non-linear relationships between height, weight, and BMI allowed it to outperform similar linear models. This highlights the importance of using advanced machine learning techniques when dealing with complex health-related data.

### 4.8.2 Impact of Personalized Recommendations

The integration of the BMI prediction model into the recommendation system enabled the provision of highly personalized fitness dietary advice. The positive feedback from user suggests that machine learning-driven recommendations can be effective in promoting health behaviours. However, the need for variety and flexibility in recommendations also become evident, indicating that while data-driven insights are valuable, user experience and engagement are equally important.

### 4.8.3 Comparison with Existing Systems

Compared to traditional, non-personalized fitness advice, the system developed in this study offers a more tailored approach. By considering individual BMI and relevant factors, the system provides recommendations that are more likely to be effective and suitable. This approach could be further enhanced by incorporating additional data such as activity levels, dietary preferences, and health history.

## 4.9 Challenges and Limitations

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

### 4.9.1 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.

### 4.9.2 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.

### 4.9.3 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.

## 4.10 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.

# Reference

Al Ansari, F.S. et al. (2023) ‘Using mobile health applications to enhance physical activity in Saudi Arabia: a cross-sectional study on users’ perceptions’, International health, 15(1), pp. 47–55. Available at: https://doi.org/10.1093/inthealth/ihac008.

Author King, A. (2023) TITLE Genotype-based diet and physical activity advice: effects on behaviour change to reduce risk factors for cardiometabolic disease. Available at: https://research.stmarys.ac.uk/id/eprint/6235/.

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.

Bodhini, D. et al. (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. Available at: https://doi.org/10.1038/s43856-023-00363-0.

Bolatbekov, A. et al. (no date) 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, Ștefan S. et al. (2022) ‘A Review of Digital Health and Biotelemetry: Modern Approaches towards Personalized Medicine and Remote Health Assessment’, Journal of Personalized Medicine. MDPI. Available at: https://doi.org/10.3390/jpm12101656.

Dergaa, I. et al. (2024) ‘Using artificial intelligence for exercise prescription in personalised health promotion: A critical evaluation of OpenAI’s GPT-4 model’, Biology of Sport, 42(2), pp. 221–241. Available at: https://doi.org/10.5114/biolsport.2024.133661.

Dirik, M. (2023) ‘Application of machine learning techniques for obesity prediction: a comparative study’, Journal of Complexity in Health Sciences, 6(2), pp. 16–34. Available at: https://doi.org/10.21595/chs.2023.23193.

Drew, K.J. et al. (2024) ‘qualitative study of the experiences of individuals who did not complete the NHS Low Calorie Diet Programme Pilot’, British Journal of Diabetes [Preprint]. Available at: https://doi.org/10.15277/bjd.2024.434.

Fjellström, S. et al. (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, 37(2). Available at: https://doi.org/10.1111/jar.13191.

Ghosh, S. et al. (2023) ‘Sedentary lifestyle with increased risk of obesity in urban adult academic professionals: an epidemiological study in West Bengal, India’, Scientific Reports, 13(1). Available at: https://doi.org/10.1038/s41598-023-31977-y.

Hu, L. et al. (2024) ‘User Experience & Usability of Wearable Health Device: A Bibliometric Analysis of 2014–2023’, International Journal of Human-Computer Interaction [Preprint]. Available at: https://doi.org/10.1080/10447318.2024.2357905.

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, 13(10). Available at: https://doi.org/10.3390/jcm13102955.

Magallanes, C.L. (2024) ‘Physical Education Teachers’ Experiences in Fitness Testing’, British Journal of Multidisciplinary and Advanced Studies, 5(1), pp. 88–147. Available at: https://doi.org/10.37745/bjmas.2022.0422.

McConnell, M. V. et al. (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. Available at: https://doi.org/10.1016/j.jacc.2018.04.030.

Mcgowan, A.H. et al. (2024) Designing for diversity: Dynamic persuasive strategies in mHealth app development ⋆.

Mustafa, S.T. et al. (2021) ‘Dietary recommendations for women with gestational diabetes mellitus: a systematic review of clinical practice guidelines’, Nutrition Reviews. Oxford University Press, pp. 988–1021. Available at: https://doi.org/10.1093/nutrit/nuab005.

Oyebode, O. et al. (2023) ‘Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness’, International Journal of Human-Computer Interaction, 39(9), pp. 1938–1962. Available at: https://doi.org/10.1080/10447318.2022.2089085.

Ozsoy, S. et al. (no date a) 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. Available at: http://meridian.allenpress.com/jcep/article-pdf/13/s2/455/3369374/i2165-7629-13-s2-455.pdf.

Ozsoy, S. et al. (no date b) 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. Available at: http://meridian.allenpress.com/jcep/article-pdf/13/s2/455/3369374/i2165-7629-13-s2-455.pdf.

Papry, F.A. et al. (2024) ‘Sedentary Lifestyle and Prospective Health Risks: A Pilot Study Among Bangladeshi Corporate Professionals’, Bangladesh Pharmaceutical Journal, 27(1), pp. 73–84. Available at: https://doi.org/10.3329/bpj.v27i1.71157.

Polak, R. et al. (2016a) ‘Health-related culinary education: A Summary of Representative Emerging Programs for Health Professionals and Patients’, Global Advances In Health and Medicine, 5(1), pp. 61–68. Available at: https://doi.org/10.7453/gahmj.2015.128.

Polak, R. et al. (2016b) ‘Health-related culinary education: A Summary of Representative Emerging Programs for Health Professionals and Patients’, Global Advances In Health and Medicine, 5(1), pp. 61–68. Available at: https://doi.org/10.7453/gahmj.2015.128.

Prowse1, R. et al. (2023) ‘Creating healthy food environments in recreation and sport settings using choice architecture: a scoping review’, Health Promotion International, 38(5). Available at: https://doi.org/10.1093/heapro/daad098.

Rodbard, H.W. et al. (2024) ‘Practical strategies to manage obesity in type 2 diabetes’, Diabetes, Obesity and Metabolism. John Wiley and Sons Inc, pp. 2029–2045. Available at: https://doi.org/10.1111/dom.15556.

Rodriguez, D. V. et al. (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. Available at: https://doi.org/10.1109/ICHI54592.2022.00051.

Rummery, K., Lawrence, J. and Russell, S. (2023) ‘Partnership and Personalisation in Personal Care: Conflicts and Compromises’, Social Policy and Society, 22(1), pp. 187–204. Available at: https://doi.org/10.1017/S1474746422000525.

Sallis, J.F. and Glanz, K. (2009) ‘Physical activity and food environments: Solutions to the obesity epidemic’, Milbank Quarterly, 87(1), pp. 123–154. Available at: https://doi.org/10.1111/j.1468-0009.2009.00550.x.

Shoba, L.K. et al. (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. Available at: https://doi.org/10.1109/ICSSES62373.2024.10561424.

Singhania, K. and Reddy, A. (no date) 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. et al. (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, 9(3). Available at: https://doi.org/10.25035/jsmahs.09.03.03.

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, 14(1). Available at: https://doi.org/10.1038/s41598-023-49274-z.

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, 21(1). Available at: https://doi.org/10.1186/s12966-024-01567-w.

Thomas, D.M. et al. (2024) ‘Transforming Big Data into AI-ready data for nutrition and obesity research’, Obesity. John Wiley and Sons Inc, pp. 857–870. Available at: https://doi.org/10.1002/oby.23989.

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. Available at: https://doi.org/10.1007/s11948-024-00479-z.

Ulfa, M. et al. (2022) ‘Nutrition-Related Mobile Application for Daily Dietary Self-Monitoring’, Journal of Nutrition and Metabolism. Hindawi Limited. Available at: https://doi.org/10.1155/2022/2476367.

Umekar, S. and Joshi, A. (2024) ‘Obesity and Preventive Intervention Among Children: A Narrative Review’, Cureus [Preprint]. Available at: https://doi.org/10.7759/cureus.54520.

Valentine, L., D’Alfonso, S. and Lederman, R. (2023) ‘Recommender systems for mental health apps: advantages and ethical challenges’, AI and Society, 38(4), pp. 1627–1638. Available at: https://doi.org/10.1007/s00146-021-01322-w.

Yadav, A., Research, P.D. and Tiwari, S. (no date) A Survey-Based Investigation into the Prevalence and Factors Influencing Sedentary Behavior among Indian Office Workers. Available at: www.theacademic.in.

Yue, W. et al. (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. Available at: https://doi.org/10.1109/JAS.2021.1003919.

Zanevych, Y. (2024) ‘FLASK VS. DJANGO VS. SPRING BOOT: NAVIGATING FRAMEWORK CHOICES FOR MACHINE LEARNING OBJECT DETECTION PROJECTS’, in. European Scientific Platform (Publications). Available at: https://doi.org/10.36074/logos-29.03.2024.066.