ASSESSING COMMERCIAL WEARABLES IN PREDICTING PHYSICAL ACTIVITY: A CASE STUDY OF APPLE AND FITBIT

Ayodeji Shadare

(10611378)

Dissertation submitted in partial fulfilment of the requirements for the degree of

Master of Science in Data Analytics

at Dublin Business School

Supervisor: Bolu Adeagbo

Word Count: 10,090

8th January 2024

DECLARATION

I declare that this dissertation that I have submitted to Dublin Business School for the award of Master of Science in Data Analytics is the result of my own investigations, except where otherwise stated, where it is clearly acknowledged by references. Furthermore, this work has not been submitted for any other degree.

Signed: Ayodeji Shadare
Student Number: 10611378
Date: 8th January, 2023

ACKNOWLEDGEMENTS

I am immensely grateful to everyone who has been a part of my journey in completing this dissertation. Your support and encouragement have been instrumental in making this academic pursuit possible.

Firstly, I would like to express my deepest gratitude to my supervisor, Bolu Adeagbo. Your attentive guidance and expertise have been invaluable throughout this research process. Your insights and constructive criticisms, especially in helping me design the dashboard, have greatly enriched the quality of this dissertation, and I am truly thankful for the mentorship you provided.

I also would like to dedicate this study to my aunt Mrs Patience Lawson for her continuous support both morally and financially and to my Parents, whose love and encouragement have been my pillars of strength, I extend my heartfelt thanks. I am grateful for the sacrifices you have made to see me through this academic journey. I would also like to acknowledge my siblings for their understanding and support during the challenging phases of this research. Your encouragement was a source of inspiration.

Ayodeji Shadare

(10611378)

TABLE OF CONTENTS

DECLARATION 2

ACKNOWLEDGEMENTS 3

ABSTRACT 6

CHAPTER ONE 7

INTRODUCTION 7

1.1 Background 7

1.2 Research Problem and Justification 8

1.3 Research Question 9

1.4 Research Questions 10

1.5 Research Hypothesis 10

1.6 Significance of the Study 10

1.7 Structure of this Report 11

CHAPTER TWO 13

LITERATURE REVIEW 13

2.1 Preamble 13

2.2 Physical Activity 13

2.3 Commercial Wearable Devices 14

2.4 Artificial Intelligence (AI) 14

2.5 Machine Learning (ML) 15

2.6 Predicting Physical Activity with Commercial Wearable Devices 16

2.7 Literature Review: Wearable Technology and Physical Activity Monitoring (Tokucoglu, 2018) 17

2.8 Literature Review: Machine Learning Approaches for Physical Activity Recognition (Bianchim et al., 2023) 19

2.9 Research Gap 20

CHAPTER THREE 21

IMPLEMENTATION 21

3.1 Preamble 21

3.2 Data Understanding 21

3.2.1 Dimension of the data 22
3.2.2 Data types of the features 23
3.2.3 Duplicate Rows 24
3.2.4 Missing Values 24
3.3 Data Cleaning 25
3.4 Data Visualization 25
3.4.1 Univariate Analyses 26
3.4.2 Bivariate Analysis 30
3.4.4 Multi Variate Analysis 31
3.5 Data Splitting 32
3.6 Feature Engineering 32
3.6 Dashboard design and Build 34
CHAPTER FOUR 36
RESULTS 36
4.1 Cross Validation 36
4.2 Machine Learning 38
4.3 Interpretable Machine Learning 41
4.4 Stack Model for Apple and FitBit 44
4.4 Dashboard Insights 44
CHAPTER FIVE 48
DISCUSSION AND CONCLUSION 48
5.1 Summary of Findings 48
5.2 Discussions 49
5.3 Implication of Findings/Recommendation 51
5.4 Limitation/Future Research 52
Reference List 54

ASSESSING COMMERCIAL WEARABLES IN PREDICTING PHYSICAL ACTIVITY: A CASE STUDY OF APPLE AND FITBIT

ABSTRACT

This study explored the predictive accuracy of Apple Watch and Fitbit in tracking physical activity, employing advanced machine learning models, feature engineering, ensemble learning, and interpretable machine learning. Using Microsoft Power BI, an interactive dashboard was also constructed to analyze user demographics, physiological metrics, and activity patterns. The machine learning models, including Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, LightGBM, XGBoost, CatBoost, and Artificial Neural Network (ANN) were comprehensively evaluated using multiple metrics. Interpretability was enhanced through Shapley values, unravelling the contribution of features to classification results. Stacking models reveal insights into their performance compared to individual models. The result generally showed that the single LightGBM model was better compared to other models and stacking. Furthermore, the dashboard insights provide a detailed exploration of user engagement across different activities, revealing variations in heart rates, distances, and calorie expenditure. This study contributes unique insights into wearable technology.

Key words: Wearable Technology; Predictive Accuracy; Physical Activity; Apple Watch; Fitbit; Machine Learning

CHAPTER ONE

INTRODUCTION

1.1 Background

The emergence of commercial wearable devices has marked a significant breakthrough in promoting physical activity within the general populace. These devices, equipped with activity-tracking capabilities, represent a pioneering innovation in technology designed to be worn directly on the body (Amit et al., 2020). Described as portable electronic gadgets, they can either be directly worn on the body or seamlessly integrated into an individual's clothing or accessories, as articulated by Zhang, Sun, and Chen (2022). Among the well-known commercially available wearable devices are Apple Watch, Fitbit, Samsung Gear, and others, primarily falling into the categories of smartwatches and fitness trackers. It is essential to note that this study focuses specifically on these categories, excluding other wearable device classifications such as smart clothing, body-mounted sensors, augmented reality headsets, hearing aids, virtual reality headsets, and smart jewelry.

Aside from being physical gadgets, wearable technology facilitates robust functionalities through data exchange, cloud computing, and application assistance. Advancements in technology have addressed previous challenges related to the bulkiness and transportation issues of wearable devices. Simpler and lighter components are now employed to create fundamental parts like sensors. The evolution of Web 2.0, combined with the maturity of big data, cloud technology, data mining, and integrated technologies, has significantly elevated the reputation and capabilities of wearable technology. Presently, these devices are capable of continuous and non-invasive measurement of various physiological parameters such as temperature, blood pressure, sweat biomarkers, electrodermal activities, electromyography, electrocardiography, and electroencephalogram (Than, Zan, and Chen, 2020). The increasing capabilities of these wearable devices have a potentially profound impact on global health and healthcare. With their ability to continuously monitor vital health parameters, they stand to revolutionize health practices and interventions on a global scale.

The proliferation of commercial wearable devices represents a pivotal juncture in bridging the gap between technology and personal health management. Despite their promise to foster physical activity and health monitoring, the accuracy and reliability of these devices in assessing and predicting physical activity levels have drawn substantial scrutiny. Numerous studies have highlighted discrepancies in the precision of these wearables, raising concerns about their efficacy in providing accurate data on physical activity. Factors such as placement on the body, sensor quality, individual variations, and the algorithms used for activity detection contribute to these discrepancies (Nweke et al., 2018; Arvidsson, Fridolfsson and Börjesson, 2019; Wang, Cang and Yu, 2019). Such variations can lead to disparities in the recorded data, affecting the overall reliability of these devices in quantifying and analyzing physical activity. Moreover, while wearable devices have evolved to encompass a wide array of health-monitoring capabilities, there is an imminent need for a comprehensive evaluation of their accuracy in predicting physical activity levels. The potential of these devices to positively influence health outcomes relies heavily on their capacity to provide precise and consistent data, a facet that necessitates rigorous examination.

This study seeks to address these gaps in the literature by conducting a meticulous evaluation specifically focused on Apple Watch and Fitbit. This research aims to provide valuable insights into the reliability and accuracy of commercially available wearables by analyzing and comparing the performance of these two prominent devices in predicting physical activity. The findings will not only contribute to academic discourse but also offer practical implications for individuals, healthcare practitioners, and the fitness industry reliant on these devices for activity tracking and health management. In doing so, this study endeavours to shed light on the veracity of data provided by these wearables, aiming to further refine their capabilities and establish their credibility in accurately predicting physical activity, an imperative step toward leveraging technology for improved personal health and well-being.

1.2 Research Problem and Justification

In spite of the potential of wearing devices and the level of development attained, the accuracy of commercial wearable devices has been called into question (Raykar and Shet, 2020). According to Mondal et al. (2020), selectivity, precision, and calibration still have a great deal of room for development. Therefore, it is possible that wearable wristwatch measurements will not be useful. For example, fall and sleep detection results are found to not be accurate enough for practical use (Liu et al., 2020; Xiong et al., 2020). It is necessary to conduct additional research on wearable technology with various degrees of accuracy for ordinary and everyday uses (Jackson et al., 2019; Xuan et al., 2020). This could explain why not many institutions in the healthcare sector use wearable devices for patient care or safety as asserted by Krey (2020). They are insufficiently precise or dependable to serve as a basis for making treatment decisions in clinical settings (Kanady et al., 2020). Therefore, the utility of consumer wearable technology needs to be confirmed by additional research. Apple Watch was chosen because it has generally been reported to have the highest accuracy in studies by Hardekopf and Fountaine (2017) and Cosoli and Scalise (2019). It is crucial to note that commercial wearable devices use proprietary software algorithms (typically machine learning based), making it impossible to determine how the measurement data were actually acquired (Cosoli, Spinsante and Scalise, 2020). The data (accessible directly on the device, or through a smartphone or internet applications) cannot be accurately compared due to the absence of generally accepted test protocols and techniques, which is a problem for researchers who want to evaluate the performances of various devices.

1.3 Research Question

The aim of this study is to investigate the accuracy of commercial wearable devices in predicting physical activity, using Apple watch and Fitbit data. The study will concentrate on the following research objectives in order to accomplish this:

1) The first objective is to focus on comparing advanced machine learning models (model development), including deep learning architectures, to achieve higher accuracy and robustness in predictive tasks. This includes Decision Tree, Logistic Regression, Naïve Bayes, Xgboost, Lightgbm, Catboost and Artificial Neural Networks (ANN).
2) The second objective is to develop and implement advanced feature engineering and selection techniques in Python to enhance model performance, including identification of the most important features and domain-specific feature creation.
3) The third objective is to investigate and implement ensemble learning methods and model stacking strategies to combine diverse models for superior predictive performance, exploring techniques like boosting, bagging, and stacking.
4) The fourth objective is to make machine learning models more interpretable and transparent, allowing for a better understanding of the decision-making process, which is essential for both practical application and ethical considerations. This is to innovate

methods to improve interpretability and explainability of complex models, ensuring transparency and understanding of how these models make decisions.

5) The fifth objective is to utilize Microsoft Power BI to design and construct a comprehensive and visually intuitive dashboard based on the dataset. This dashboard aims to distill actionable insights from the dataset in relation to Apple and Fitbit wearables, providing a user-friendly interface for the exploration and analysis of key metrics related to physical activity.

1.4 Research Questions

The study will concentrate on the following research questions in order to accomplish this:

1) How accurate is Apple Watch in predicting sedentary behaviour and mild, moderate and strenuous physical activity?
2) How does Apple watch compare with Fitbit for predicting physical activity?
3) Which Machine Learning models are the most accurately in predicting physical activity with Apple Watch data?
4) What features can aid in forecasting for future studies?

1.5 Research Hypothesis

H1: Apple watch will accurately predict moderate and vigorous physical activity.

H2: Fitbit watch will accurately predict light and sedentary behaviour.

1.6 Significance of the Study

The significance of this proposed study lies in its profound potential to elevate the field of machine learning and data analysis to new heights, impacting multiple crucial dimensions. At the forefront, this research endeavours to propel the current understanding of machine learning methodologies to an advanced level. It seeks to provide unique insights into model performance and interpretability by concentrating on the latest techniques and harnessing the latest technologies. These advancements are anticipated to carve a path toward the forefront of cutting-edge machine learning research and application. The practical implications of this study extend far beyond academia. In this regard, industries heavily reliant on data-driven decision-making, such as healthcare, finance, autonomous systems, and e-commerce, stand to benefit significantly. The development of enhanced machine learning models could potentially revolutionize these sectors, optimizing operations and refining decision-making processes.

A paramount aspect of this study revolves around addressing the ethical and transparent application of artificial intelligence. The emphasis on transparency, fairness, and the reduction of biases in machine learning models aligns closely with the current efforts to ensure ethical practices in AI applications. These endeavours are pivotal in fostering trust and reliability in the application of machine learning models. Moreover, the investigation into efficient deployment strategies for machine learning models is of immense practical importance. Bridging the gap between model development and real-world implementation is often a bottleneck in various applications. Successful integration can pave the way for smoother deployment processes, offering practical solutions in diverse scenarios. Academically, the study's outcomes are poised to contribute substantially. The insights, methodologies, and advancements in machine learning techniques could lay the groundwork for future research, providing invaluable resources for scholars, researchers, and practitioners in the field. On a more practical level, this study's outcomes hold significance for professionals seeking to enhance their data analysis and machine learning skills, starting with the researcher. Furthermore, it has the potential to serve as an educational resource, offering new methodologies and best practices for both learning and teaching machine learning using Python.

1.7 Structure of this Report

Each chapter of this report contributes significantly to the study, progressing from the initial identification of the problem to the proposal and development of advanced solutions, ending with insightful conclusions and directions for future research. The structure of this research report follows a sequential and comprehensive approach, comprising the following five key chapters:

- **Chapter 1: Introduction and Problem Identification**

This opening chapter sets the stage by providing an introduction to the study. It articulates the context and significance of the research while identifying the existing gaps in the field of machine learning analysis. The chapter will offer a detailed examination of the problem statements, setting the foundation for subsequent explorations.

- **Chapter 2: Literature Review**

Building upon the introduction, this chapter extensively explores existing policy frameworks, industry reports, and relevant literature in the domain of advanced machine learning and Python-based data analysis. It critically evaluates prior studies, methodologies, and best practices to provide a comprehensive background for the study.

- **Chapter 3: Implementation**

This chapter describes the research plan, methodologies, and design employed in the study. It focuses on the methodical approach adopted for developing and refining advanced machine learning models. Additionally, it highlights the strategies used for enhancing model interpretability, scalability, and considerations in the analysis.

- **Chapter 4: Results**

Focused on the core findings of the study, this chapter presents the proposed solutions and key features developed. It discusses the innovative machine-learning techniques, models, and strategies implemented using Python. This chapter aims to detail the advancements made in model performance, interpretability, handling unstructured data, and ethical considerations.

- **Chapter 5: Conclusions and Future Work**

The concluding chapter summarizes the key findings and insights derived from the study. It discusses the implications of the research outcomes, offering a reflection on the significance of the study. Furthermore, this chapter outlines potential areas for future work and advancements in the field, paving the way for subsequent research endeavours.

CHAPTER TWO

LITERATURE REVIEW

2.1 Preamble

With the advent of wearable technology, vital indications such as heart rate, step count, and other factors like calories elevation, may now be monitored in real-time (Sunny et al., 2022). These tools make it possible to monitor the aforementioned physiological indicators longitudinally and continuously. Increased activity is expected to translate into significant favourable health effects, both physical and mental. Tracking, measuring, and recording one's physical activity can be a technique to monitor and encourage oneself to participate in regular physical activity (Germini et al., 2022). The majority of physical activity tracking used to be done manually by the individual or an outside assessor using records. There is the possibility to develop novel measures of physical activity and sedentary behaviour utilizing technologies that are already widely used by the public, even though specific measures from commercial wearable devices (i.e., Heart rate steps) are known to involve measurement error (Fuller et al., 2021). This study aims to investigate if movement types may be predicted using information from commercial wearables like the Apple Watch.

2.2 Physical Activity

Physical inactivity is a serious public health issue on a global scale. According to a recent study, 81 percent of teenagers who are enrolled in school and roughly 23 percent of adults are not meeting the recommended levels of physical activity (World Health Organization, 2015). Governmental agencies have tried to increase these figures by putting into place programs that encourage physical activity. Behavior change is a tried-and-true technique to enhance physical activity, even though the successful promotion of physical activity is a difficult multifaceted issue (Conn, Hafdahl and Mehr, 2011). Guidelines for physical activity have been defined using metrics from commercial wearable devices, such as 10,000 steps per day and 100 steps per minute for moderate to strenuous activity (Fuller et al., 2020). The number of steps taken by different devices varies, according to research, and the relevance of these measurements may differ by device type and brand (Case et al., 2015).

The American Heart Association (AHA) suggests physical activity as one of the "Life's Simple 7" lifestyle recommendations to maintain health since it is inversely connected with poor cardiovascular outcomes and all-cause mortality (Bayoumy et al., 2021; Blond et al., 2019). The evaluation of physical activity levels has historically been subjective and, if recorded at all, only done so during clinic visits. This method's limitations include insufficient information, recall bias, and the inability to impartially evaluate physical activity in a real-world setting. Common statements like "I walk for 30 minutes five times a week" exclude crucial details such physical activity intensity, distance, and idle time. Therefore, the subjective reporting of physical activity levels will become outdated because digital health trends like wearables and smartphones can precisely and objectively measure physical activity and energy expenditure using a variety of sensors.

2.3 Commercial Wearable Devices

Consumers are starting to control their health with consumer-grade software and hardware as technological advancements continue to become increasingly integrated in daily life. Consumer-grade linked electronic gadgets known as "smart wearables" can be worn as an accessory or sewn into clothing and are of the highest quality (Bayoumy et al., 2021). They come in a variety of forms, such as smartwatches, rings, and wristbands, to name a few, and all of them feature many advanced sensors and powerful processing capacity. These gadgets serve as a motivational tool that can increase physical activity rather than just being a fitness accessory. Wearables incorporate a range of behavioural change psychological strategies (Lewis et al., 2020). These methods, also known as features, include social support (socialization), incentives, recognition, alerts, and giving feedback. The features that are most crucial to facilitating a change in physical activity are perhaps social support, personalization, prompts, and activity tracking.

2.4 Artificial Intelligence (AI)

Artificial intelligence (AI) is already capable of complex tasks including picture categorization, speech recognition, translation, object identification, driving, gaming, finance, and even legal decision-making (Vatandsoost and Litkouhi, 2019). By using sophisticated algorithms and data from electronic health records (EHRs), clinical research results published in PubMed, millions of patient imaging results, hundreds of biomarkers, and data retrieved from EHRs, AI is able to identify diseases (Krittanawong, 2018). As per Ray, Nagarajan, and Minu, (2022), artificial intelligence is defined as a program that is capable of doing tasks often associated with the human intellect. Artificial intelligence is a theory and the development of computerized systems intended to carry out tasks that would ordinarily be completed utilizing human intelligence and senses including hearing, vision, language comprehension, and decision-making (McGrow, 2019). The recent improvements in the Artificial intelligence technologies across healthcare, brings about certain speculations on substituting the human physicians in the future. But practically, AI medical technologies cannot replace human physicians, but they can help them get better outcomes and greater accuracy. Healthcare data availability is a key factor supporting the development of AI solutions in the medical area (Manne and Kantheti, 2021). Recently, there has been a lot of work done by scientists and researchers to use AI to medical diagnosis, treatment, and care. AI appears to be quickly transforming the face of healthcare. Medical physics foresees that AI will significantly advance healthcare as a whole. Some scientists anticipate that AI will have an impact on medical physics research and practice, while others think that this is implausible given AI's current state of development and practical constraints (Vatandsoost and Litkouhi, 2019).

2.5 Machine Learning (ML)

The phrase "Machine Learning" is attributed to Arthur Samuel of IBM, who in 1959 suggested that it would be able to educate computers to learn what they need to know about the environment and how to do jobs for themselves (Simon 1983). In the branch of computer science known as "machine learning," algorithms and statistical techniques are used to recognize patterns and draw conclusions from data without the use of explicit programming. It is a method for creating software that can learn from data and adjust itself to perform tasks like prediction better (Dash et al., 2022). The premise that computers may learn without being programmed to carry out specific tasks gave rise to the application of AI known as machine learning (Panesar and Panesar, 2020). Techniques used in machine learning include Bayesian approaches, reinforcement learning, neural networks, explanation-based, inductive logic programming, natural language processing, and decision trees. Examples of cutting-edge ML models include

Xgboost, Lightgbm, Logistic Regression, Support Vector Machine, Naive Bayes, Random Forest.

The machine learning enables software applications to predict outcomes more precisely and particularly without having to be explicitly programmed. The creation of algorithms that can both receive input data and utilize statistical analysis to predict an output while updating outputs with fresh data is the main goal of machine learning (Saleh, 2019). Machines learn from prior instances and historical trends, and based on their prior knowledge, a model that can be used to predict future values can be developed. Machine learning can be used to help find answers to problems based on the study of data when the data and questions are too big to be resolved organically (Bansal, Singh and Kaur, 2020). It can facilitate quicker searches for crucial items. Because machines learn more quickly and can even outperform humans in some fields, complex issues are easily handled. As a result, its demand is always increasing (Bansal, Singh and Kaur, 2020). Machine learning is becoming more important as big data and cloud computing does since it uses computational power to solve various problems.

2.6 Predicting Physical Activity with Commercial Wearable Devices

When employed in a research intervention, wearables have proven to be reliable and effective, according to researchers. Despite the fact that these tools may not be the best for measuring exercise, they are nonetheless capable of accurately monitoring steps and sedentary behaviour (Nelson et al., 2016; Evenson, Goto and Furberg, 2015). The wearer receives an accurate assessment of their physical activity patterns thanks to this. Enhancements in all physical activity metrics result from this validity. According to research conducted by Brickwood et al., (2019) wearing a wearable increased daily step count, moderate and vigorous activity levels, and overall energy expenditure much more than comparator groups. Engagement with these gadgets depend on a number of variables. These variables include tracking precision, social utility, aesthetics, and the device's physical design (Harrison et al., 2015). A diary study of people using wearable technology for the first time revealed that the way the data was recorded, handled, visualized, and used had an impact on the participants' involvement (Rapp and Cena, 2016). Although it has been demonstrated that the wearables are beneficial in interventions with a variety of demographics, they don't seem to be as successful as conventional behavioural interventions

(Lewis et al., 2015; Sypes, Newton and Lewis, 2019). Additionally, improvements in functionality and usability could boost wearables' effectiveness by boosting engagement.

According to a study conducted by Fuller et al., (2020), Consumer wearable device reviews from nine different brands (Apple Inc., Fitbit, Garmin, Mio, Misfit, Polar, Samsung, Withings, and Xiaomi) were included in this systematic review of 158 publications. The reviews focused on the accuracy and precision of the devices in measuring heart rate, energy expenditure, and step count. In addition to highlighting the findings of the inter- and intra-device reliability of the nine consumer wearable companies, this review looked at the validity of consumer wearable devices in both free-living and laboratory situations. Fitbit received the most attention across the research, while Xiaomi and Mio received the least. Apple, Fitbit, and Garmin were roughly 50% accurate on average, with Apple and Samsung having the highest validity for step count. No brand was within the allowed energy expenditure accuracy criteria. Steps, heart rate, and calorie inter-device reliabilities were all quite high. Only step count had enough data to determine intra-device dependability, and the results revealed a lot of variation. No one gadget or brand stood out as the "gold standard" in fitness wearables, nor was there one that involved a thorough evaluation across all metrics. The market for electronic devices is now dominated by applications for basic biophysical signal monitors, such as heart rate monitors, temperature monitors, and motion monitors. Apple Watch has a heart rate tracking feature that is as accurate as clinical monitors for patients and can also identify irregular heartbeats and inform users (Wang, 2020). Many different sectors are also covered by other wearable systems. According to a study, it was found that data acquired from commercial wearable devices can predict movement types in 82% of Apple watch (Fuller et al., 2021). In this following sections, special review is carried out for two very related studies. These studies were very critical in helping the research navigate this research domain.

2.7 Literature Review: Wearable Technology and Physical Activity Monitoring (Tokucoglu, 2018)

Tokucoglu (2018) conducted a comprehensive review focusing on the utilization of wearable technology for monitoring physical activity and its applications in healthcare. The study emphasized the necessity of objective and accurate measurements for evaluating both healthy individuals and patients, highlighting the role of wearable sensors and related technologies in this context.

Key research highlights from this study include:

1. **Wearable Technology System:** The study provided an in-depth analysis of the components required for a wearable technology system, encompassing sensors, communication hardware, software, and analyzing equipment. It classified wearable devices into head-mounted gadgets and bodily placed instruments, detailing recent advances in microelectronics and e-textile production.

2. **Wearable Technology Devices:** The historical evolution of wearable devices, from watches to modern smartwatches, was outlined. The study discussed commercially available devices capable of measuring vital signs continuously, such as blood pressure, heart rate, and oxygen concentration. It also touched upon the integration of flexible sensors into garments, known as "e-textile," facilitating electrocardiographic and electromyographic data collection.

3. **Potential Uses in Medicine:** Jones et al. explored potential medical applications of wearable technologies, emphasizing their role in monitoring congestive heart failure, dysrhythmia, asthma, balance problems, Parkinson's disease, sleep disorders, and various chronic conditions. The study highlighted the importance of motion sensors and posture detection devices in evaluating efficacy of treatment and enhancing patient outcomes.

4. **Handicaps of Wearable Technologies:** The study acknowledged challenges associated with wearable technologies, including potential discomfort for users, data accuracy concerns, and the ethical implications related to privacy and information usage. Safety and privacy issues were identified as crucial considerations, posing questions about authorized personnel, decision-making competence, and financing.

The research by Tokucoglu (2018) aligns with the present study's objectives of investigating the predictive capabilities of commercial wearables. The emphasis on the diverse applications of wearable technology in healthcare, particularly in monitoring specific medical conditions, resonates with the current focus on predicting movement types. The challenges identified, such as user discomfort and data accuracy, inform the potential limitations of the current study and underscore the need for thoughtful consideration of ethical and practical aspects. Overall, the insights contribute to the understanding of the broader landscape of wearable technology in healthcare, providing a foundation for the present study's exploration of predicting movement types using commercial wearables.

2.8 Literature Review: Machine Learning Approaches for Physical Activity Recognition (Bianchim et al., 2023)

The study by Bianchim et al. (2023) aimed to develop and validate machine learning models for predicting physical activity (PA) intensities in children and adolescents with cystic fibrosis (CF). The research utilized accelerometers from different brands and placements on both wrists and waist. Three machine learning classifiers (K-Nearest Neighbour, Random Forest, and eXtreme Gradient Boosted Decision Tree) were employed to identify input signal patterns for each PA type and intensity. The study involved 35 children and adolescents with CF and 28 healthy youth, and a 10-fold cross-validation assessed the performance of the classifiers.

The key findings from this study include:

1. **Prediction Accuracy:** The machine learning models achieved high accuracy (97–100%) for predicting different PA intensities, with sensitivity and specificity greater than 95%. Notably, ActiGraph GT9X on the dominant wrist and waist and GENEActiv on the dominant wrist failed to predict vigorous intensity PA activities.
2. **Influence of Health Condition:** Children with CF exhibited higher energy expenditure during certain activities compared to healthy peers, emphasizing the need for CF-specific approaches in predicting PA levels.
3. **Accelerometer Brand and Placement:** The study found that GENEActiv demonstrated higher accuracy than ActiGraph, and overall, all placements (wrist and waist) yielded comparable accuracy across classifiers. However, there were instances where specific

placements and brands failed to predict certain PA intensities, highlighting the importance of considering these factors in PA measurement.

4. **Machine Learning Algorithms:** The XGBoost and k-NN algorithms achieved higher sensitivity and classification accuracy compared to Random Forest, suggesting their potential superiority in predicting PA intensities.
5. **Implications for CF Treatment:** The study's findings have potential implications for clinical practice, indicating that machine learning algorithms could be used to identify daily patterns of PA in youth with CF, contributing to the design of PA interventions and specific recommendations for this population.

The study provides novel insights into the use of machine learning models with raw accelerometry data to predict PA intensities in children and adolescents with CF, emphasizing the importance of considering health conditions, accelerometer brand, and placement for accurate predictions. This study also provides insights into methodologies, challenges, and potential benefits. Consideration of factors such as accelerometer brand, placement, and the unique characteristics of the target population, as discussed in the study, could enhance the design and interpretation of the current work.

2.9 Research Gap

Wearable technology and mobile cloud computing have advanced quickly as a result of the shift to individualized, evidence-based healthcare, which will relieve pressure on healthcare systems, save costs, and improve health outcomes. However, the majority of wearable wrist and textile technology that is currently commercially available is at the level of general health and fitness. The existing body of scientific literature on the application of wrist and textile monitoring technology to support clinical judgments in physical activity suggests that more work is needed to get the technology into clinical practice.

CHAPTER THREE

IMPLEMENTATION

3.1 Preamble

In this chapter, the methodology employed for the classification of physical activity in the dataset is described. The dataset utilized for this analysis is sourced from Kaggle.com and comprises a set of 20 distinct features. These features include diverse user-related attributes, including age, gender, height, weight, heart rate, device specifications, the activity variable (serving as the target variable), and additional parameters assessing aspects of body metabolism. The implementation chapter is structured into two pivotal aspects. The first aspect is dedicated to the application of machine learning techniques for the classification of physical activity, wherein various machine learning algorithms are deployed to discern patterns and relationships within the dataset. Subsequently, the second aspect focuses on the design and implementation of a comprehensive dashboard, employing Microsoft Power BI as the primary tool. This dashboard is devised to facilitate the visual representation and interpretation of additional insights derived from the dataset, fostering a user-friendly interface for in-depth exploration of key metrics related to physical activity.

3.2 Data Understanding

The method employed to investigate the dataset in order to build a robust machine learning model, integral to attaining the project objectives, is described in this section. In the initial step, the requisite libraries essential for project manipulation were imported. The utilization of the Google Colab Jupyter notebook is chosen for project management due to its adaptable nature and open-source attributes (Kimm, Paik and Kimm, 2021).

In the following code snippet, the necessary libraries are imported. This step installs some of the packages that were used in the analysis. i.e. the libraries not loaded by default in Jupyter notebook. Furthermore, the necessary libraries, notably Pandas and NumPy for data preprocessing and cleaning, as well as Seaborn and Matplotlib for data visualization, are imported. Additionally, machine learning libraries and tools for model explainability are included in this set of useful libraries.

Import neccessary libraries

```
!pip install catboost
!pip install shap
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import chi2_contingency
from sklearn.preprocessing import LabelEncoder
import shap
%matplotlib inline
```

```
## Remove warnings
import warnings
warnings.filterwarnings('ignore')
warnings.filterwarnings('ignore', category=UserWarning, module='sklearn')
# Importing necessary libraries from scikit-learn
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import train_test_split

# Importing boosting libraries
from lightgbm import LGBMClassifier
from xgboost import XGBClassifier
from catboost import CatBoostClassifier

# Importing libraries for Ensemble
from sklearn.ensemble import StackingClassifier

## Evaluation metrics

from sklearn.metrics import accuracy_score, f1_score, recall_score, precision_score, matthews_corrcoef,confusion_matri

```

The following data investigation steps involved checking the data dimension, datatypes, duplicate values, and missing values in our dataset.

3.2.1 Dimension of the data

This is the act of checking the total rows (observations) and total columns (features) of the dataset (Giri and Lengyel, 2023). This aided in understanding the complexity of the dataset, providing guidance for subsequent processes involving data visualization and machine learning. The following code snippet shows that the dataset comprises 6264 observations with 20 columns.

```
print(f'The total shape of the data set {df.shape}')
print(f'The total rows of the data set {df.shape[0]}')
print(f'The total columns of the data set {df.shape[1]}')
```

```
The total shape of the data set (6264, 20)
The total rows of the data set 6264
The total columns of the data set 20
```

3.2.2 Data types of the features

The process involved checking the data types of each feature in the dataset. This guided the subsequent data cleaning phase, ensuring correction of any features with incorrect data types to mitigate bias during machine learning model training. The following code snippet shows that the dataset comprises 14 numerical features, 4 integers and 2 categorical features.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6264 entries, 0 to 6263
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Unnamed: 0            6264 non-null   int64
 1   X1                    6264 non-null   int64
 2   age                   6264 non-null   int64
 3   gender                6264 non-null   int64
 4   height                6264 non-null   float64
 5   weight                6264 non-null   float64
 6   steps                 6264 non-null   float64
 7   hear_rate             6264 non-null   float64
 8   calories              6264 non-null   float64
 9   distance              6264 non-null   float64
 10  entropy_heart         6264 non-null   float64
 11  entropy_setps         6264 non-null   float64
 12  resting_heart         6264 non-null   float64
 13  corr_heart_steps      6264 non-null   float64
 14  norm_heart            6264 non-null   float64
 15  intensity_karvonen    6264 non-null   float64
 16  sd_norm_heart         6264 non-null   float64
 17  steps_times_distance  6264 non-null   float64
 18  device                6264 non-null   object
 19  activity              6264 non-null   object
dtypes: float64(14), int64(4), object(2)
memory usage: 978.9+ KB
```

3.2.3 Duplicate Rows

The scrutiny for duplicate rows within the dataset is considered a prudent data investigation procedure (Qizi, Kadambayevich and Atabaevich, 2023). Duplicate values here are defined as instances where the entire set of feature values for a particular record is repeated. The significance of this verification lies in the necessity to eliminate such duplicates, as their presence contributes to dataset dimensionality and may introduce redundancy into the analytical process. Furthermore, detection and subsequent removal of these duplicates serve to streamline the dataset and enhance the precision of the analysis. The following code snippet shows that there are no duplicate values in the dataset.

```
df[df.duplicated()]
```

Unnamed: 0	X1	age	gender	height	weight	steps	hear_rate	calor

3.2.4 Missing Values

This step was done to check if there are omitted records in dataset. The following code snippet shows that there are no missing values in the dataset. This concludes the data investigation steps.

```
Unnamed: 0             0
X1                     0
age                    0
gender                 0
height                 0
weight                 0
steps                  0
hear_rate              0
calories               0
distance               0
entropy_heart          0
entropy_setps          0
resting_heart          0
corr_heart_steps       0
norm_heart             0
intensity_karvonen     0
sd_norm_heart          0
steps_times_distance   0
device                 0
activity               0
dtype: int64
```

3.3 Data Cleaning

Data cleaning is a process of handling any incorrect, duplicate, missing, or erroneous values in the data set (Chai, 2020). Due to the nature of the dataset, during investigation it was observed that some columns will not be useful in further analysis of the data. This data cleaning step involved dropping the columns. This is shown in the following code snippet.

```
[14]  to_drop = ['Unnamed: 0', 'X1']
      df = df.drop(to_drop, axis=1)
```

3.4 Data Visualization

Data visualization constituted an integral component of the initial steps in the analysis, aiming to extract insights from the data. Various charts, including bar charts, histograms, and scatter plots, were employed to reveal patterns within the dataset. The visualization encompassed univariate, bivariate, and multivariate analyses. The chi-square statistical test was applied to discern patterns within categorical features, while correlation analysis was utilized to understand associations between numerical variables.

3.4.1 Univariate Analyses

The bar chart in Figure 3.1 below shows that lying is the most frequent activity captured. Moreover, the classes are slightly balance.

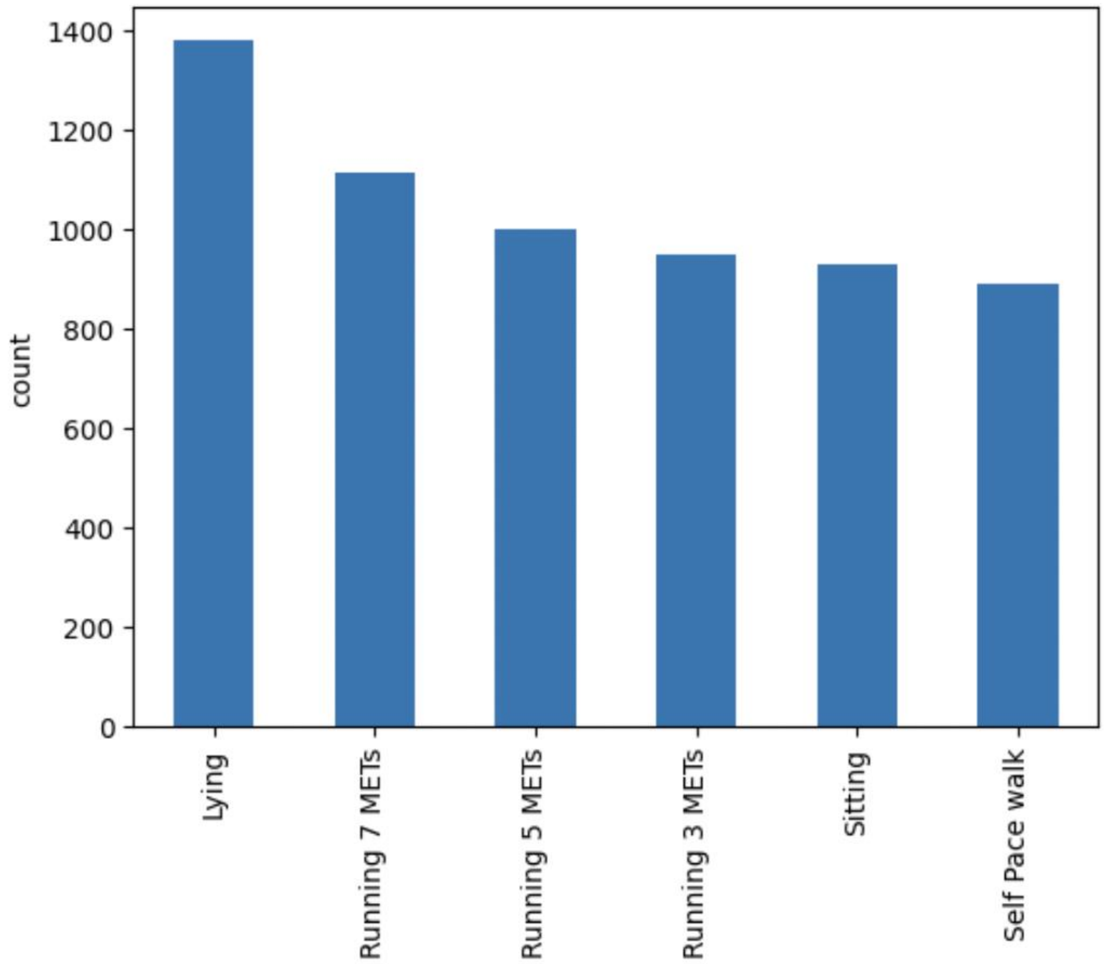

Figure 3.1: Activity occurrence

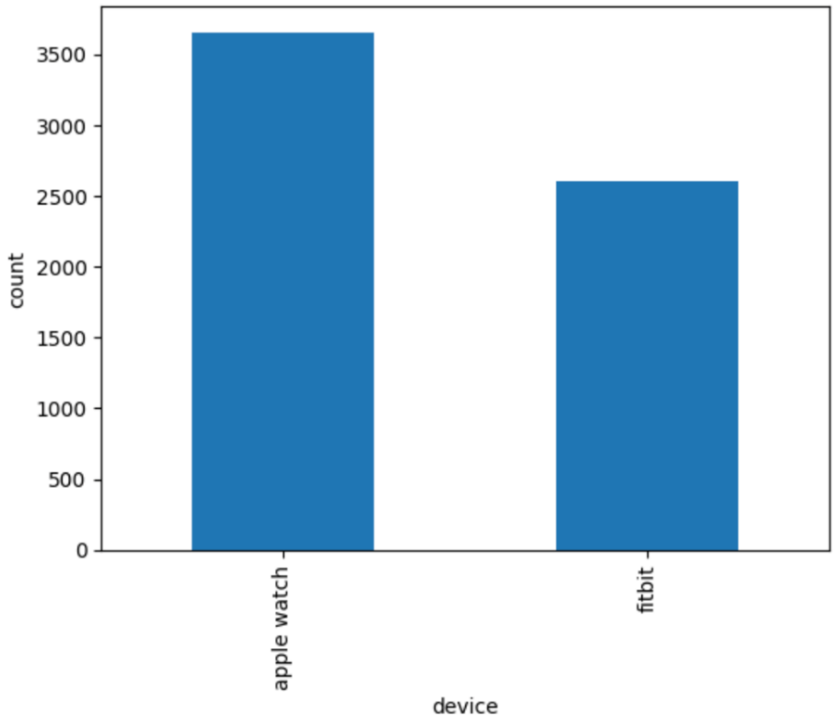

Figure 3.2: Device occurrence

The bar chart in Figure 3.2 above shows that Apple watch users were more in the dataset compared to Fitbit users.

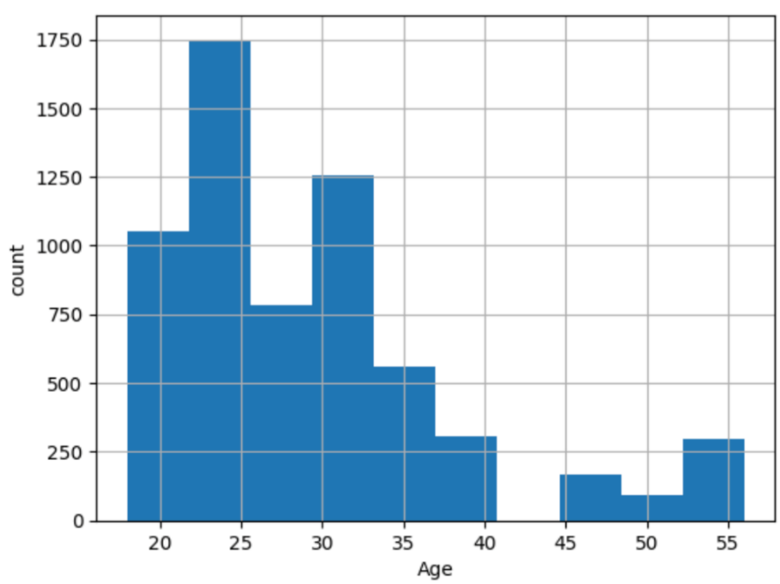

Figure 3.3: Age distribution of users

Figure 3.3 above shows that the majority of users fall between 20 - 25 years old, the maximum age is 58 years, and, there is an age gap from 40-45 years range where no users were found. The absence of users within the age range of 40 to 45, as illustrated in Figure 3.3, implies a notable gap or lack of representation for individuals falling within this specific age bracket in the dataset. This gap could potentially impact the robustness of the analysis, as insights derived from the dataset may not adequately capture the characteristics or behaviors of users within the 40 to 45 years age range.

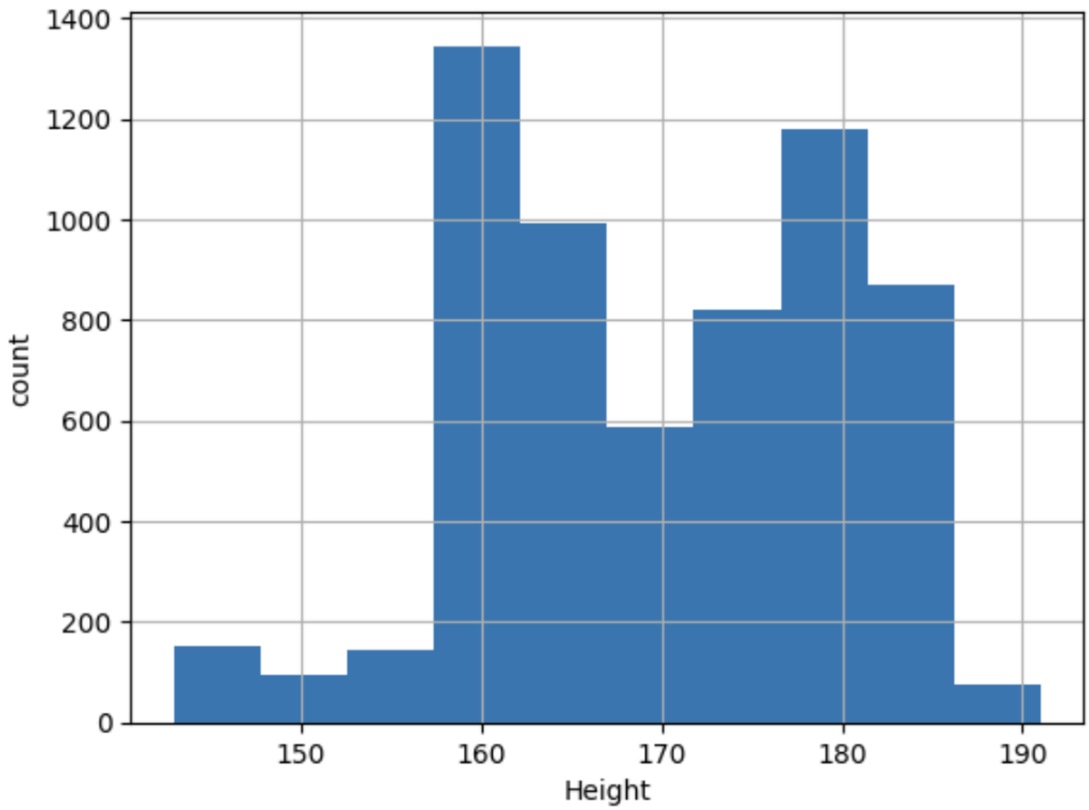

Figure 3.4: Height distribution of users

Figure 3.4 above shows most height of users falls between 160-165cm.

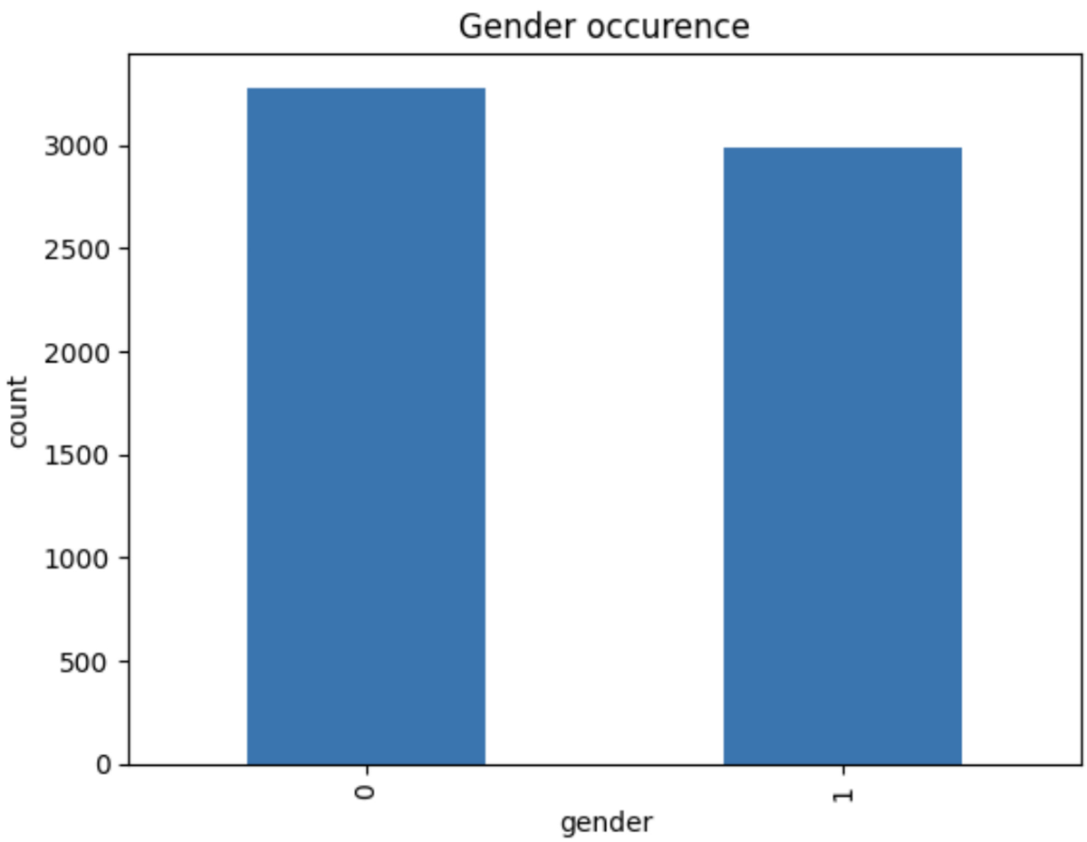

Figure 3.5: Gender distribution of users (0=females; 1=males)

Figure 3.5 above shows that the dataset comprises users with more females using the fitness devices than male.

3.4.2 Bivariate Analysis

The Chi-square test is a statistical method employed to assess whether there is a significant association between two categorical variables. This evaluates the likelihood that any observed association in the data could have occurred by chance. The test is premised on the null hypothesis, which posits that the two variables are independent of each other. If the observed data significantly deviate from what would be expected under this null hypothesis, the test would have concluded that the variables have a dependent relationship. Figure 3.6 below shows that lying activities is the most frequent in both devices (Apple, Fitbit) with 13% and 9.5% respectively. Furthermore, chi square test shows that the correlations between variables for the two devices are not statistically significant.

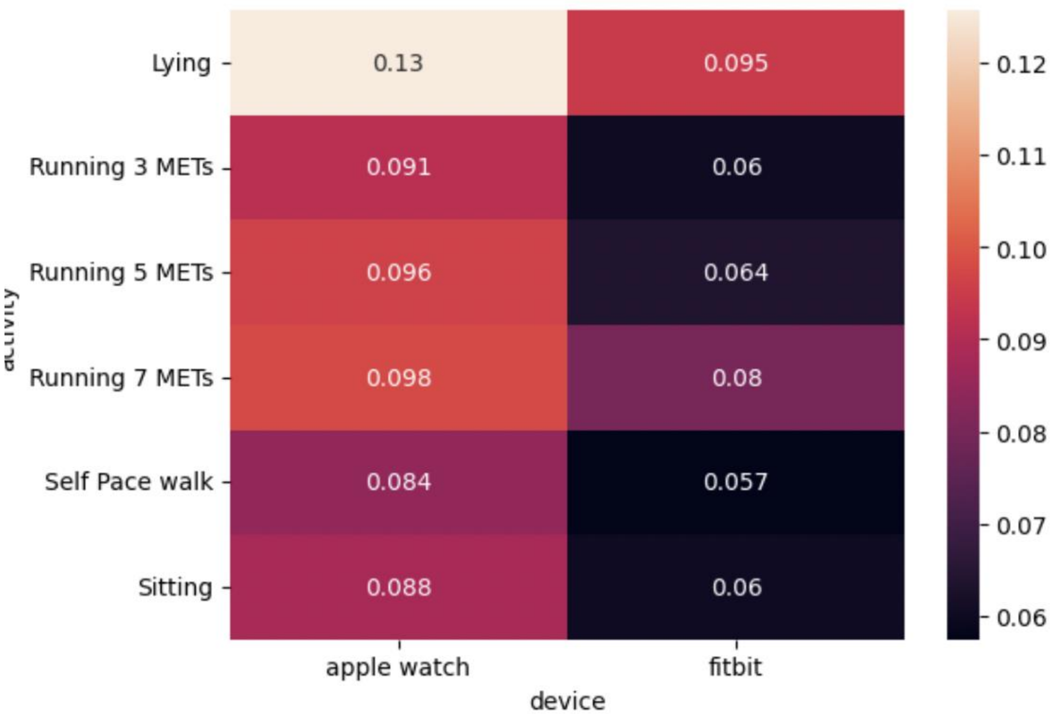

Figure 3.6: Correlation between variables

In terms of gender, Figure 3.7 shows that females dominated the category of activity in the dataset.

Further chi square test for this category shows that the correlations between variables for the two devices were not statistically significant.

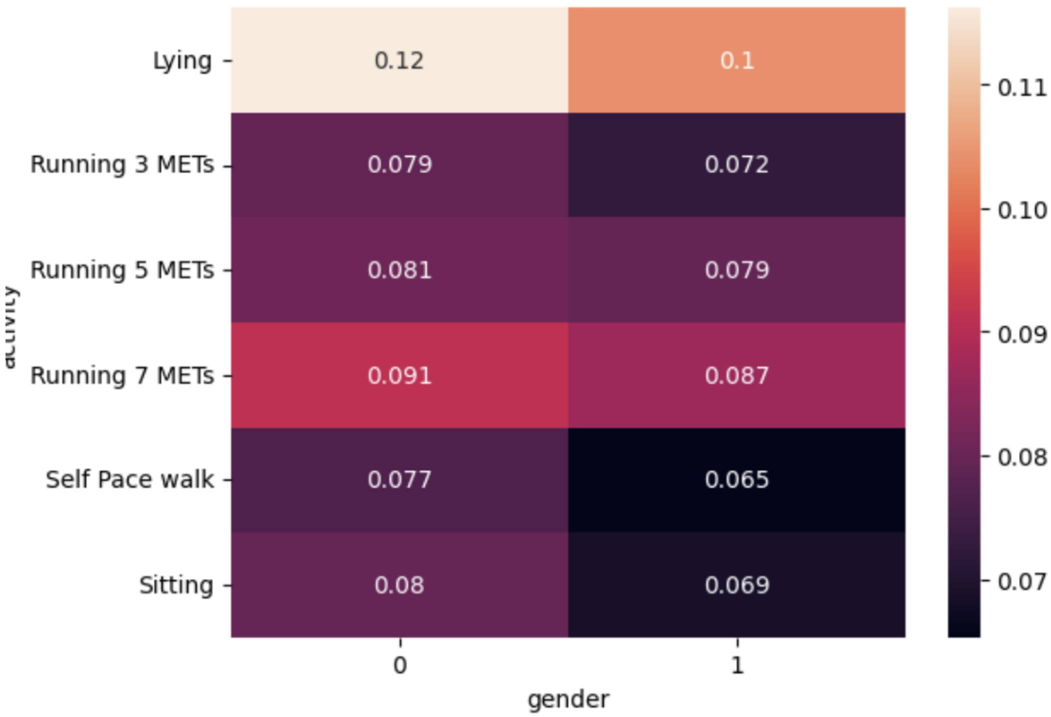

Figure 3.7: Bivariate gender analysis

3.4.4 Multi Variate Analysis

In the multivariate analysis in Figure 3.8 below, it can be deduced that the higher the normal heart rate the higher the intensity of the activity, having a 95% correlation. Furthermore, there is strong correlation between height and weight with 69%. More still, gender is correlated with height and weight with 74% and 58% respectively. The correlation graph also shows that age is mostly negative correlated with some features.

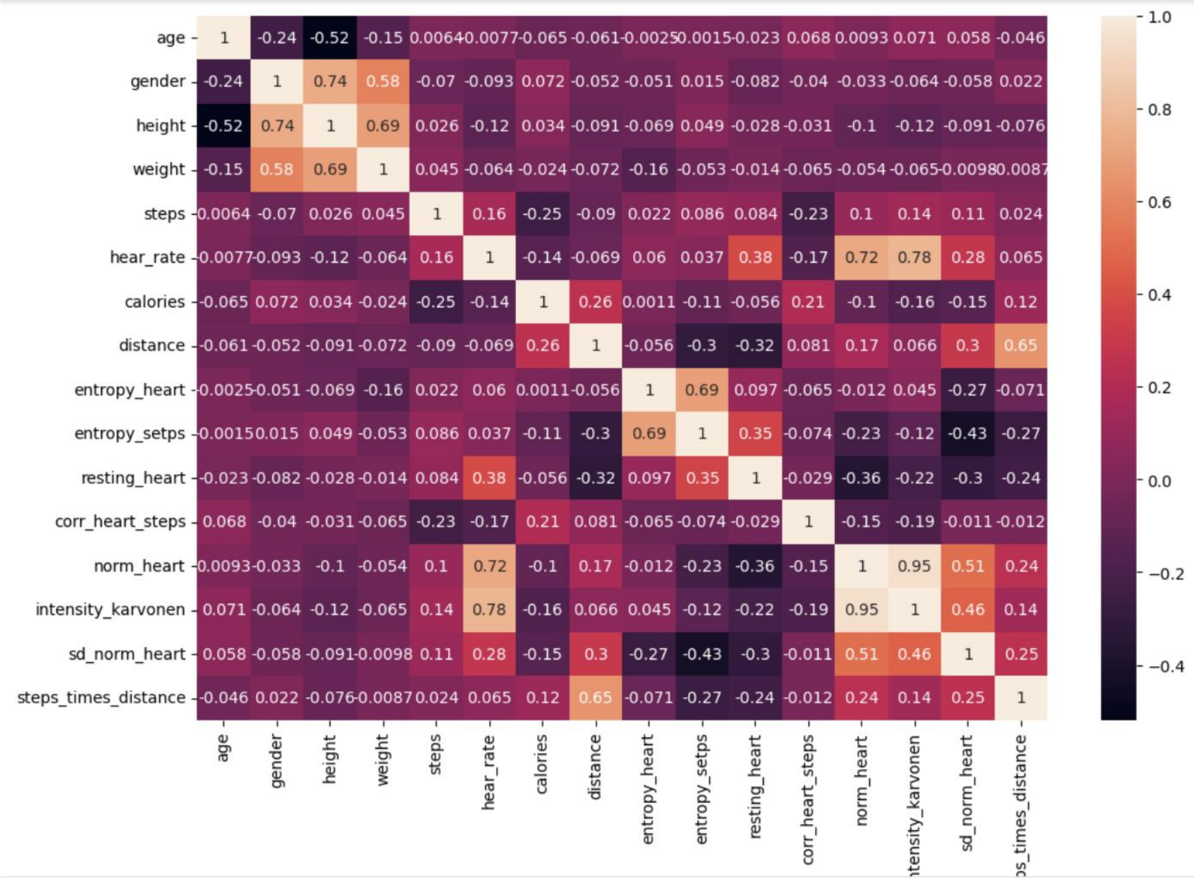

Figure 3.8: Correlations among variables

3.5 Data Splitting

To facilitate the achievement of the project objectives, the dataset was split by device into Apple Watch and Fitbit data as shown by the following code snippet.

```
ap_df = df[df['device'] == 'apple watch'].reset_index(drop=True)
ft_df = df[df['device'] == 'fitbit'].reset_index(drop=True)
```

3.6 Feature Engineering

The feature engineering process involves the science and art of generating new features from existing ones in the dataset, thereby enhancing the accuracy of the model (Kanjilal and Uysal, 2021). The creation of these new features is informed by domain-specific knowledge and extensive research. The process is shown in the following code snippet.

```
# BMI (Body Mass Index)
df['bmi'] = df['weight'] / ((df['height'] / 100) ** 2)

# Calories burn per Step
df['calories_per_step'] = np.where(df['steps'] > 0, df['calories'] / df['steps'], 0)

#  Cardiovascular Fitness Index (using resting heart rate and maximum estimated heart rate
df['cv_fitness_index'] = df['resting_heart'] / df['hear_rate'].max()

# Active vs Resting Heart Rate Differential
df['heart_rate_differential'] = df['hear_rate'] - df['resting_heart']

# Step Intensity (combining 'steps' with 'hear_rate')
df['step_intensity'] = df['steps'] * df['hear_rate']

```

In this regard, Calories Burned per Step quantifies the average caloric expenditure per step taken. Derived by dividing the total calories burned by the total number of steps, this feature provides a detailed perspective on energy efficiency during physical activity. Its significance lies in its utility for personalized health and fitness assessments, particularly in evaluating the caloric cost associated with various activities. The Cardiovascular Fitness Index represents the ratio of the resting heart rate to the maximum estimated heart rate, offering insights into an individual's cardiovascular fitness. A lower ratio is generally indicative of superior fitness, suggesting a lower heart rate for a given level of exertion. This index serves as a potential proxy for assessing cardiovascular efficiency and overall heart health. The Active vs Resting Heart Rate Differential captures the variance between the active heart rate and the resting heart rate. This metric provides insights into the heart's responsiveness to physical activity, offering a measure of the adaptability and resilience of the cardiovascular system. Step Intensity, a composite measure derived from the multiplication of the number of steps and the heart rate, reflects the intensity of physical activity. Higher values indicate more vigorous exercise. This compound measure holds potential significance in studies exploring the interplay between the volume and intensity of physical activity.

3.6 Dashboard design and Build

As part of the project's implementation, the process of constructing the dashboard using Power BI involved several steps. Initially, the relevant dataset, encompassing the assessed commercial wearables data, was imported into the Power BI environment. Subsequently, data cleaning procedures were executed to ensure the integrity and consistency of the dataset. Following data preparation, the construction of the dashboard commenced with the identification and selection of key performance indicators (KPIs) and metrics to be visualized. Visualization components such as charts, graphs, and tables were strategically chosen to effectively represent the essential aspects of the dataset, aligning with the project's objectives. The Power BI tools were then employed to create and customize the chosen visualizations, ensuring coherence and clarity in conveying insights. Various features within Power BI, such as filters, slicers, and drill-through functionalities, were utilized to enhance interactivity and facilitate a more dynamic exploration of the data. Additionally, considerations were given to the aesthetic aspects of the dashboard, including colour schemes, font choices, and layout design, to optimize user experience and comprehension. The dashboard was refined iteratively, incorporating feedback and adjustments to improve its effectiveness in conveying meaningful insights derived from the assessed commercial wearables data. Finally, the completed Power BI dashboard was shared at the following link: https://app.powerbi.com/view?r=eyJrIjoiOGRmYjBhNjUtMTJmNi00OTU3LWExZTQtZmUyNzcyNzAyODk2IiwidCI6Ijg4ZTlhN2RjLTU2MzMtNGM2Ni1iNjZjLTkyZGY1Y2E3NDhmYyJ9 allowing stakeholders and end-users to interact with the visualizations, explore patterns, and draw insights relevant to the project's objectives.

The dashboard provides a comprehensive visualization of various metrics related to physical activity, segmented by age grades, and distinguished between Apple Watch and Fitbit devices. The visualizations include averages of calories, average distance, entropy steps, average heart rate, resting heart rate, normalized heart rate, steps, and steps times distance across different age groups. The upper section of the dashboard displays bar charts depicting the average values of calories and average distance, as well as entropy steps and entropy heart, categorized by age grades. This visual representation enables a comparative analysis of these metrics across different age groups. The subsequent section showcases the average heart rate, resting heart rate, and normalized heart rate, providing insights into the cardiovascular aspects of the users. The bar charts are segmented by age grades, allowing for a nuanced examination of heart rate trends within each group. The third section exhibits the average values of steps and steps times distance, differentiated by age grades. This comparison provides a perspective on the physical activity levels across various age groups. The dashboard also features a breakdown of individual metrics, allowing users to explore specific details such as average calories, average distance, average heart rate, and average steps. Users can filter the data by device (Apple Watch or Fitbit), gender (female or male), and specific activities. Furthermore, a demographic analysis is presented with a bar chart illustrating the count of users within each age group. This adds an additional layer of understanding regarding the distribution of users across different age categories.

CHAPTER FOUR

RESULTS

This chapter unveils the results of the activity classification process and generated insights from the dashboard. The preceding chapter described aspects of data investigation, data cleansing, data visualization, and feature engineering, aimed at extracting insights from the data to enhance the robustness of the results of this chapter. The following sections will expound upon the comparative performances of the generalized machine learning models, achieved through the establishment of training and test data, utilizing stratified one-holdout cross-validation techniques. Furthermore, the focus of this chapter is also directed towards the use of the interpretable machine learning, specifically Shapley values, to describe the features in the context of demystifying the inherent complexities of the various machine learning models. This chapter also presents the results of the dashboard.

4.1 Cross Validation

Cross-validation is a technique employed to partition the dataset into training and test sets, serving as a means to assess the model's performance on previously unseen data (de Rooij and Weeda, 2020). In this project, the approach of stratified one-holdout cross-validation (LandJr and Schaffer, 2019) is adopted to ensure the preservation of class distribution across both the training and test sets throughout the training and evaluation phases of the machine learning models. The dataset was consequently partitioned into 80% training data and 20% test data, with a designated random seed of 2023. This random seed is chosen to maintain consistency in the dataset splitting process, ensuring uniformity in the allocation of data for training and testing purposes. The process is shown in the following code snippet.

```
[32] lb = LabelEncoder()
     ap_df['activity'] = lb.fit_transform(ap_df['activity'])
```

```
[33] X = ap_df.drop(['activity', 'device'], axis=1)
     y = ap_df['activity']
```

```
[34] # Separating the 80% data for training data and 20% for testing data
     X_train, X_test, y_train, y_test = train_test_split(X,y,test_size = 0.2, random_state =2023, stratify=y)
```

```
print('Shape of the X_train {}'.format(X_train.shape))
print('Shape of the y_train {}'.format(y_train.shape))
print('Shape of the X_test {}'.format(X_test.shape))
print('Shape of the y_test {}'.format(y_test.shape))
```

```
Shape of the X_train (2924, 21)
Shape of the y_train (2924,)
Shape of the X_test (732, 21)
Shape of the y_test (732,)
```

The code snippet above shows the cross validation for Apple data.

```
[40] lb = LabelEncoder()
     ft_df['activity'] = lb.fit_transform(ft_df['activity'])
```

```
[41] X = ft_df.drop(['activity', 'device'], axis=1)
     y = ft_df['activity']
```

```
[42] # Separating the 80% data for training data and 20% for testing data
     X_train, X_test, y_train, y_test = train_test_split(X,y,test_size = 0.2, random_state =2023, stratify=y)
```

The code snippet above shows the cross validation for Fitbit data.

The mapping of the activity image is shown in the code below. This aims to enhance the clarity and understanding of how different activities are categorized or classified within the context of the project.

```
df['activity'].unique()
```

```
array(['Lying', 'Sitting', 'Self Pace walk', 'Running 3 METs',
       'Running 5 METs', 'Running 7 METs'], dtype=object)
```

```
pd.DataFrame(lb.fit_transform(df['activity']))[0].unique()
```

```
array([0, 5, 4, 1, 2, 3])
```

4.2 Machine Learning

The algorithms designated for predicting the likelihood of a transaction being fraudulent or non-fraudulent encompass the following models:

1. Logistic Regression
2. Naive Bayes
3. Decision Tree
4. Random Forest
5. LightGBM
6. XGBoost
7. CatBoost
8. Artificial Neural Network (ANN)

The evaluation metrics employed to assess the performance of the models are outlined as follows (Chicco and Jurman, 2020; Chicco and Jurman, 2023; Foody, 2023):

- **Accuracy:** This metric quantifies the ratio of correctly predicted observations to the total number of observations, providing an indication of the overall correctness of model predictions. Accuracy=Number of Correct Predictions/Total Number of Predictions
- **F1 Score:** The F1 Score, being the harmonic mean of Precision and Recall, serves to balance the trade-off between Precision and Recall, offering a comprehensive measure of the model's performance. F1=2×(Precision×Recall)/Precision+Recall
- **Recall:** Also known as sensitivity or true positive rate, Recall signifies the ratio of correctly predicted positive observations to all actual positives, providing insights into the model's ability to identify positive instances.
  Recall=True Positives/True Positives + False Negatives
- **Precision:** Precision represents the ratio of correctly predicted positive observations to the total predicted positives, elucidating the accuracy of positive predictions made by the model. Precision=True Positives/True Positives + False Positives
- **MCC (Matthews Correlation Coefficient):** The MCC serves as a correlation coefficient between observed and predicted binary classifications, considering both true and false positives and negatives, providing a comprehensive measure of model performance.
  MCC=(TP×TN−FP×FN)/sqrt(TP+FP)×(TP+FN)×(TN+FP)×(TN+FN)

	Model	Accuracy	F1_Score	Recall	Precision	MCC
4	LightGBM	0.826503	0.826508	0.826503	0.827547	0.791121
6	CatBoost	0.818306	0.817790	0.818306	0.818465	0.781347
3	Random_Forest	0.815574	0.814558	0.815574	0.814573	0.777835
5	XGBoost	0.811475	0.810601	0.811475	0.810732	0.773101
2	Decision_Tree	0.699454	0.697525	0.699454	0.698200	0.638233
1	Naive_Bayes	0.321038	0.305302	0.321038	0.333138	0.191429
0	Logistic_Regression	0.295082	0.204629	0.295082	0.233500	0.158485
7	ANN	0.239071	0.173466	0.239071	0.442787	0.130318

Figure 4.1: Classification results comparisons among models based on Apple data

From Figure 4.1 above, the best model in classifying the activity based on Apple data is the LightGBM model.

	Model	Accuracy	F1_Score	Recall	Precision	MCC
4	LightGBM	0.904215	0.904253	0.904215	0.904547	0.884294
6	CatBoost	0.904215	0.904074	0.904215	0.904592	0.884515
3	Random_Forest	0.894636	0.894622	0.894636	0.895258	0.872798
5	XGBoost	0.892720	0.892553	0.892720	0.892926	0.870571
2	Decision_Tree	0.860153	0.859681	0.860153	0.859962	0.831326
7	ANN	0.574713	0.565625	0.574713	0.619200	0.494641
0	Logistic_Regression	0.377395	0.316354	0.377395	0.338010	0.248916
1	Naive_Bayes	0.300766	0.274553	0.300766	0.376751	0.185812

Figure 4.2: Classification results comparisons among models based on Fitbit data

From Figure 4.2, the best model in classifying the activity on Fitbit data is also LightGBM model.

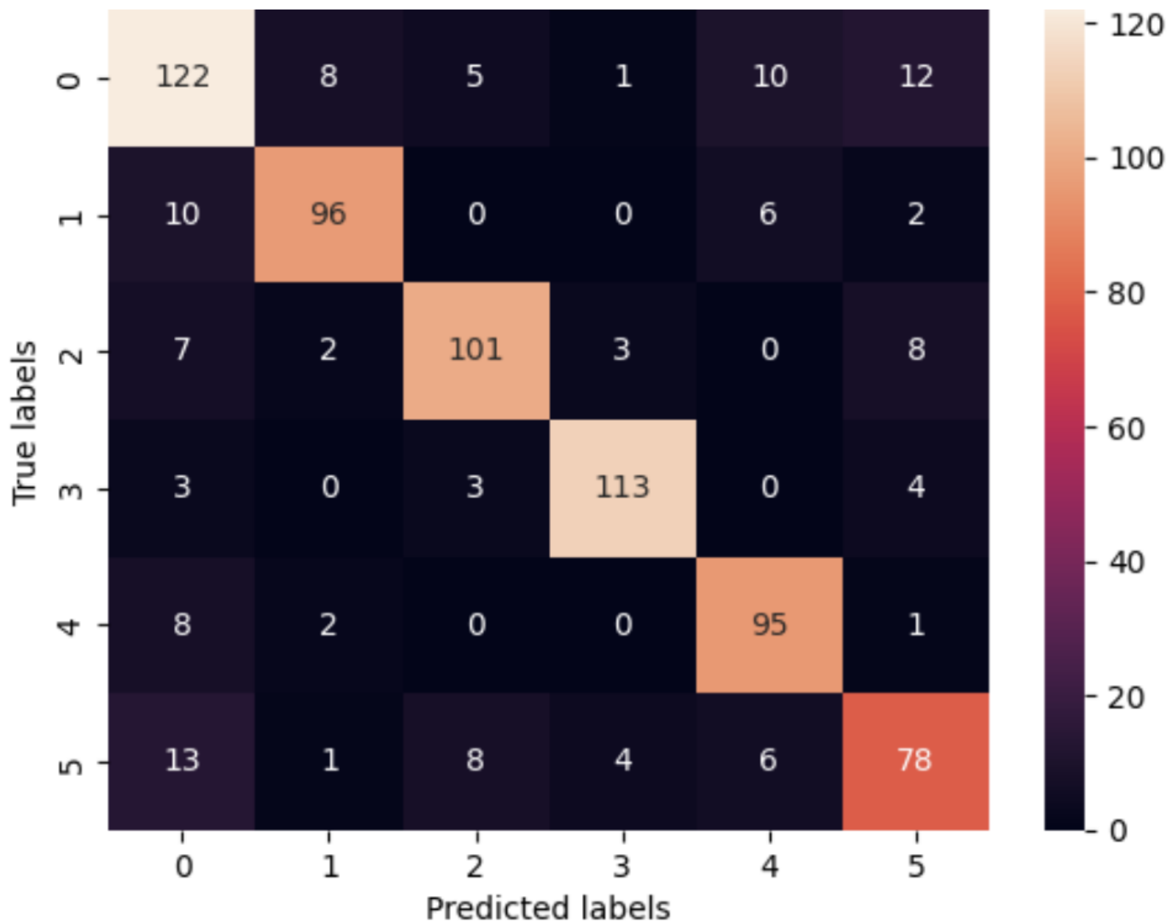

Figure 4.3: Confusion matrix for Apple result

From the confusion matrix in Figure 4.3, it can be seen that the Self pace walk (3) and Running 7 METs (4) have less misclassified classes than others.

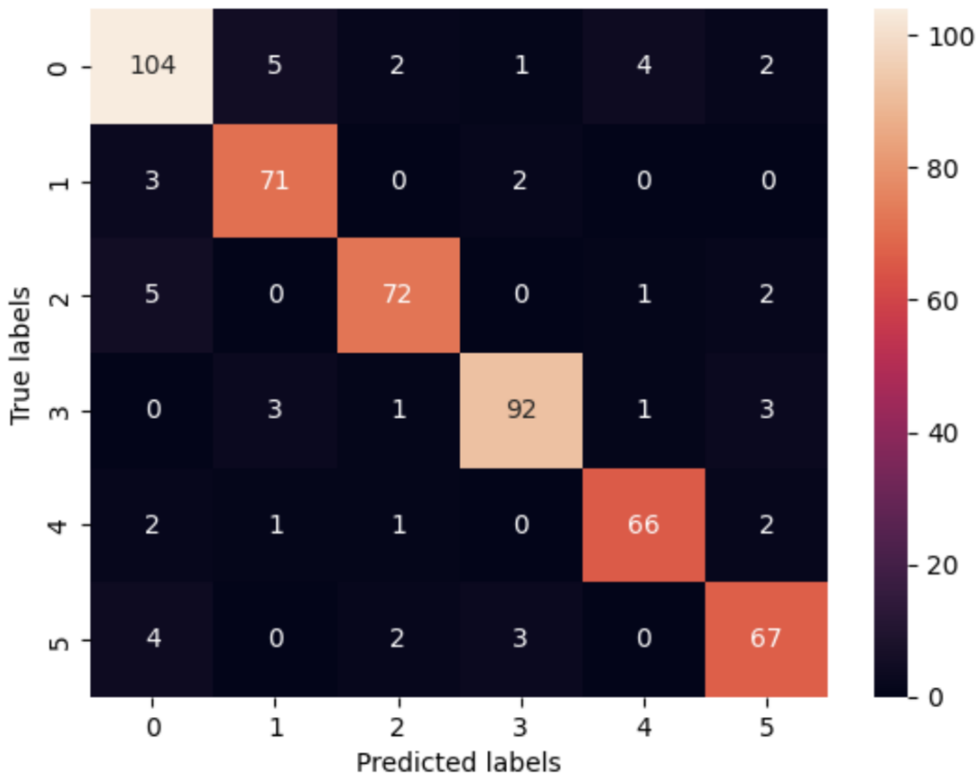

Figure 4.4: Confusion matrix for Fitbit result

From Figure 4.4, the Running 3 METs (1) Sitting (5), and Running 5 METs (2) have less misclassified classes than others.

4.3 Interpretable Machine Learning

Interpretable machine learning serves as a methodology that enables an understanding of the underlying mechanisms of black-box models, facilitating informed decision-making (Mi, Li and Zhou, 2020). The adoption of interpretable models is instrumental in fostering trust in the model's predictions, diagnosing errors, ensuring ethical compliance, and facilitating model debugging. Given that our primary model operates as a black-box model, Shapley values will be employed for interpretation purposes. SHAP (Shapley Additive Explanation) presents a game-theoretic approach designed to explain the output of any machine learning model, as detailed in the documentation (https://shap.readthedocs.io/en/latest/). Shapely Additive Explanations play a crucial role in discerning the contribution of each feature to the model's output. This understanding enhances the trustworthiness of the model and provides insights into the impact of individual features on the assigned class. The ensuing analysis will focus on elucidating the top features' contributions to determining the classification results. Furthermore, Shapley values will aid in anomaly detection, specifically in identifying features that predominantly consist of outliers during the training and evaluation phases of the models. This multifaceted approach contributes to a comprehensive interpretation of the model's decision-making processes.

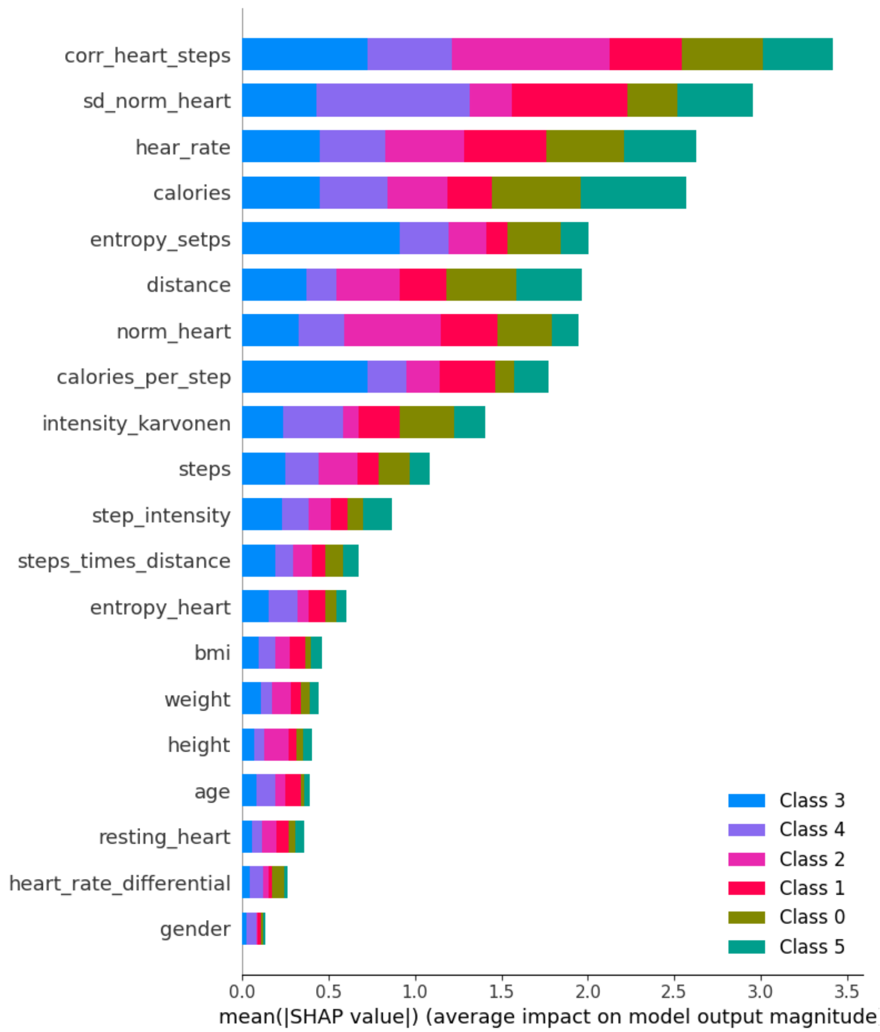

Figure 4.5: Shapely plot summary for Apple classification result

In Figure 4.5, it is evident that the predominant features crucial for the classification of activities in Apple data are identified. Notably, the feature "corr heart steps" exhibits the most substantial influence, particularly on Class 2 (Running 5 METs). Additionally, it is observed that the feature "sd norm heart" holds significant influence over Class 5 (Sitting), and this pattern persists across other classes as well.

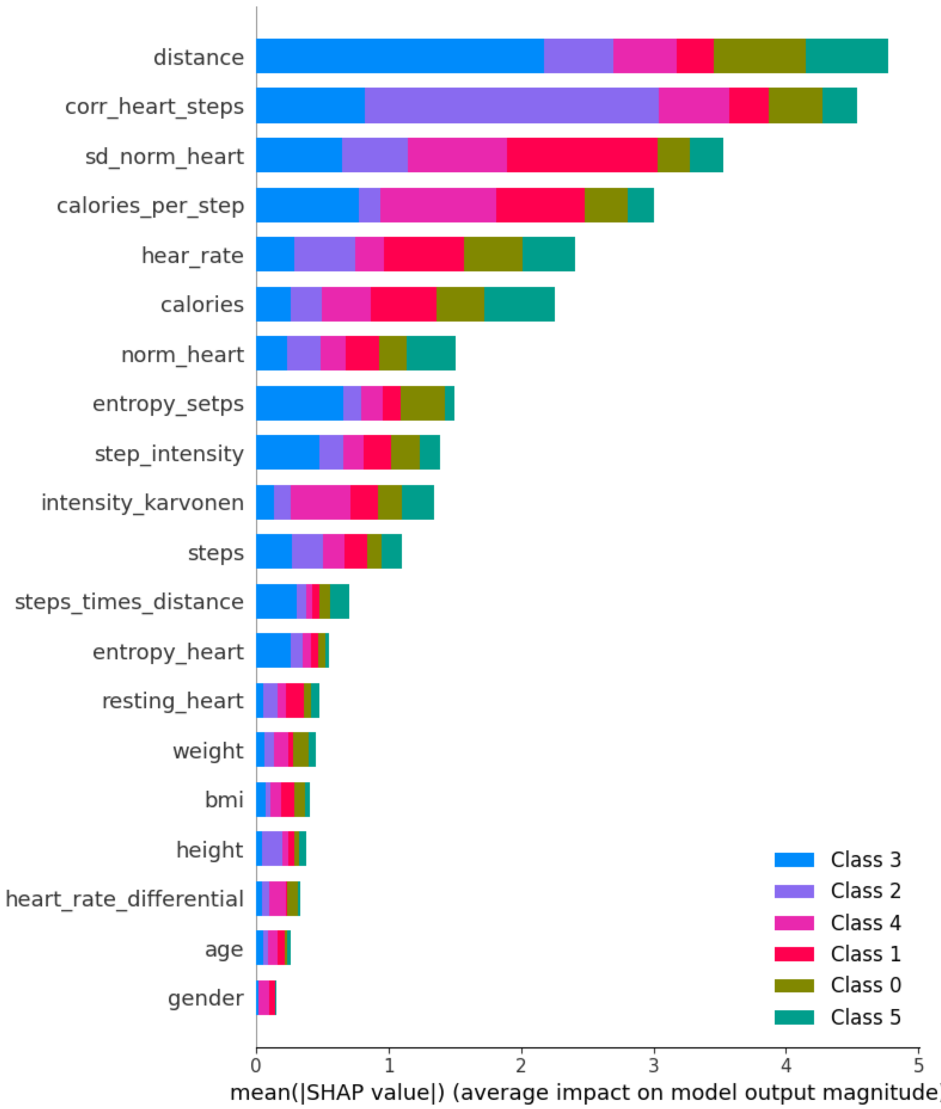

Figure 4.6: Shapely plot summary for Fitbit classification result

Similarly, from Figure 4.6, the feature "distance" is observed to exert the most significant influence, particularly on Class 3 (Running 7 METs). Furthermore, it is noted that the feature "corr heart steps" holds the most influence over Class 2 (Running 7 METs), and this trend extends to other classes as well.

4.4 Stack Model for Apple and FitBit

Stacking, also recognized as "stacked generalization," represents an ensemble machine learning technique employed to amalgamate multiple classification or regression models (Brownlee, 2021; Çodur, 2023). The fundamental concept underlying stacking is to enhance predictive performance by harnessing the collective strengths of diverse base models.

	Model	Accuracy	F1_Score	Recall	Precision	MCC
0	Stacked_model	0.180328	0.13548	0.180328	0.186035	0.036453

The code snippet above shows that stacking models for the apple data classification performed so poorly.

	Model	Accuracy	F1_Score	Recall	Precision	MCC
0	Stacked_model	0.574713	0.565625	0.574713	0.6192	0.494641

Similarly, from the code above, the stacking models for Fitbit data classification is slightly better that with Apple data. However, the conclusion still remains that single LightGBM model is far better than stacking in classification performance.

4.4 Dashboard Insights

The dashboard offers a comprehensive comparative analysis between users of Apple Watch and Fitbit, unraveling insights across several dimensions. An overview of the dashboard is shown in Figure 4.7 below. It is important to note that the dashboard is interactive. Therefore, the following insights are generated by interacting with the dashboard. Notably, the data illustrates a considerable preference for Apple Watch, with 3,656 individuals (58.37%) opting for this device, as opposed to 2,608 (41.63%) who chose Fitbit. This suggests a significant dominance of Apple Watch in the user base, implying, albeit weakly, a higher market share. Upon closer examination of gender distribution, both devices showcase comparable patterns. Female users slightly outnumber males, with Apple Watch having 1,925 females (52.65%) and 1,731 males (47.34%). Similarly, Fitbit reflects 1,925 females (52.65%) and 1,731 males (47.34%). These findings indicate a consistent gender distribution across both devices.

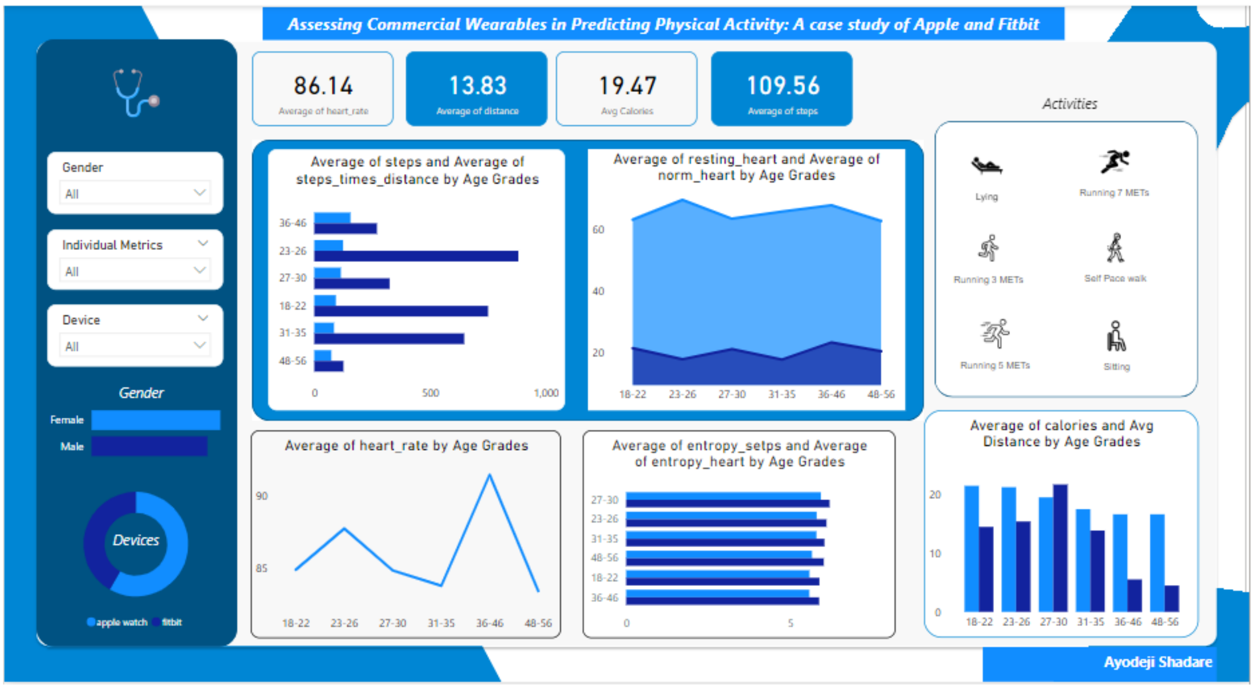

Figure 4.7: An overview of the dashboard

The absence of users in the age range of 40 to 45 years stands out as a noteworthy observation in the demographic analysis, prompting a closer examination of potential contributing factors. This peculiar gap sparks intriguing questions, leading to the exploration of various considerations that could elucidate the reasons behind this phenomenon. The absence of users aged 40-45 may suggest a deliberate focus on different age groups or possibly a gap in marketing efforts tailored to attract individuals within this particular age bracket. Furthermore, differences in preferences for wearable technology across age groups become a relevant avenue for exploration. Influenced by factors such as lifestyle, technological affinity, and health consciousness, understanding why individuals in the 40-45 age range refrain from utilizing these wearables could provide valuable insights. The lack of users in the specified age gap might also be indicative of broader health and fitness trends. It raises the possibility that individuals in the 40-45 age range may not be as inclined to use commercial wearables for tracking physical activities. A detailed exploration into their lifestyle patterns, fitness habits, and health priorities could shed light on whether wearables align with their preferences and needs. Patterns of technological adoption within the 40-45 age range could also offer valuable insights. Factors such as familiarity with wearable technology, comfort with digital devices, and attitudes toward health-tracking tools may play a pivotal role in their decision-making process. Investigating the technological behaviours and preferences of individuals within this age gap is crucial for understanding their reluctance to embrace commercial wearables.

The average physiological metrics, such as heart rate and distance covered, reveal intriguing patterns. The collective average heart rate for both devices is 86.14 bpm, suggesting a commonality in physiological responses. However, a more detailed exploration of gender-specific averages unveils subtle distinctions. For instance, male Apple Watch users exhibit an average heart rate of 82.81 bpm, whereas male Fitbit users record an average of 86.06 bpm. The average distance covered during activities stands at 13.83 meters for both devices, indicating parity in this aspect of user engagement. The following detailed exploration of specific activities sheds light on diverse user engagement patterns:

a. Lying:

- Apple Watch users: Average heart rate of 80.31 bpm, distance covered 0.09 km, and calories burnt 6.15.
- Fitbit users: Average heart rate of 78.16 bpm, distance covered 27.24 km, and calories burnt 10.28.

b. Running 3 METs:

- Apple Watch users: Average heart rate of 86.3 bpm, distance covered 0.06 km, and calories burnt 7.57.

- Fitbit users: Average heart rate of 83.55 bpm, distance covered 31.95 km, and calories burnt 59.17.

c. Running 5 METs:

- Apple Watch users: Average heart rate of 98.07 bpm, distance covered 0.08 km, and calories burnt 5.72.
- Fitbit users: Average heart rate of 80.58 bpm, distance covered 33.89 km, and calories burnt 67.35.

d. Running 7 METs:

- Apple Watch users: Average heart rate of 113.31 bpm, distance covered 0.10 km, and calories burnt 2.48.
- Fitbit users: Average heart rate of 78.73 bpm, distance covered 37.23 km, and calories burnt 33.74.

e. Running Self-Paced Walking:

- Apple Watch users: Average heart rate of 85.66 bpm, distance covered 0.06 km, and calories burnt 7.62.
- Fitbit users: Average heart rate of 78.05 bpm, distance covered 37.33 km, and calories burnt 57.92.

f. Sitting:

- Apple Watch users: Average heart rate of 85.66 bpm, distance covered 0.06 km, and calories burnt 7.62.
- Fitbit users: Average heart rate of 76.23 bpm, distance covered 33.13 km, and calories burnt 20.50.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

5.1 Summary of Findings

This chapter presents the outcomes of the activity classification process and insights derived from the dashboard, aligning with the objectives outlined in the study. The focus on comparing advanced machine learning models, implementing feature engineering techniques, exploring ensemble learning methods, ensuring interpretability, and constructing a comprehensive dashboard using Microsoft Power BI is systematically addressed. Cross-validation, an essential technique for assessing model performance, was employed through stratified one-holdout cross-validation. The dataset was partitioned into 80% training data and 20% test data, maintaining class distribution across both sets. This process aimed to ensure consistency in the dataset splitting, crucial for training and evaluating machine learning models.

Eight models, including Logistic Regression, Naive Bayes, Decision Tree, Random Forest, LightGBM, XGBoost, CatBoost, and Artificial Neural Network (ANN), were utilized to predict the likelihood of transaction fraudulence. Evaluation metrics such as Accuracy, F1 Score, Recall, Precision, and MCC were employed. Notably, LightGBM emerged as the best-performing model for both Apple and Fitbit data. Interpretable machine learning using Shapley values was introduced to elucidate the black-box model. Shapley Additive Explanations provided insights into feature contributions for both Apple and Fitbit classifications. Top features influencing classification outcomes were identified, enhancing the trustworthiness and transparency of the models. The stacking of models, aimed at combining diverse models for superior predictive performance, was explored. However, results indicated that a single LightGBM model outperformed the stacking approach for both Apple and Fitbit classifications.

The dashboard presented a comparative analysis of Apple Watch and Fitbit users. Apple Watch dominated with 58.37% users, while Fitbit accounted for 41.63%. Gender distribution revealed a slight female majority for both devices. The average heart rate for both devices was 86.14 bpm. Notably, a demographic gap was observed, with no users in the 40-45 age range, prompting further exploration. Detailed physiological metrics unveiled insightful patterns. Specific activities demonstrated varied user engagement, with notable distinctions between Apple Watch and Fitbit users. For instance, in Running 5 METs, Apple Watch users exhibited an average heart rate of 98.07 bpm, while Fitbit users recorded 80.58 bpm. The absence of users in the 40-45 age range raised questions about market targeting, user preferences, health trends, and technological adoption, requiring further investigation.

5.2 Discussions

This study's findings can be discussed and juxtaposed in the light of several studies that have explored similar topics. For instance, Ridgers et al. (2021) conducted a cluster-randomized controlled trial to examine the effect of commercial wearables and digital behavior change resources on the physical activity of adolescents attending schools in socio-economically disadvantaged areas. The study used Fitbit Flex devices and online digital resources delivered via Facebook to promote physical activity among the participants. The study found no significant differences between the intervention and control groups in terms of accelerometer-derived and self-reported physical activity levels immediately post-intervention and at 6-months follow-up. The study suggested that commercial wearables alone may not be sufficient to increase physical activity levels among youth, especially in socio-economically disadvantaged settings.

Another relevant study is by Bunn et al. (2018), which conducted a systematic review of the current state of commercial wearable technology in physical activity monitoring. The study evaluated several devices, including Fitbit Flex, for their accuracy and validity in measuring biometric information such as steps, heart rate, and caloric expenditure. The study found that most devices were reasonably accurate in measuring steps, but less so in measuring heart rate and caloric expenditure. The study also noted that device accuracy varied depending on the type and intensity of physical activity, as well as individual characteristics such as age, gender, and body mass index. The study recommended that users and researchers should be aware of the limitations and sources of error in commercial wearable devices and use them with caution.

A third pertinent study is by Melton et al. (2016), which investigated the feasibility and effectiveness of using commercial wearable devices and smartphone applications to increase physical activity and fitness among college students. The study used a randomized controlled trial design to compare the effects of Fitbit Charge 2 devices and MyFitnessPal applications on the participants' physical activity levels, fitness, and body composition. The study found that both interventions resulted in significant increases in physical activity and fitness, as well as significant decreases in body fat percentage, compared to the control group. The study also found that the Fitbit group had higher adherence and satisfaction rates than the MyFitnessPal group. The study concluded that commercial wearable devices and smartphone applications can be feasible and effective tools for promoting physical activity and fitness among college students.

In comparing the findings of this study with existing academic research, several key insights and methodological approaches can be discussed. This study's emphasis on employing advanced machine learning models aligns with the broader trend in predictive analytics research. Various studies, such as those by Wang and Jiang (2023) and Kit et al. (2023), have demonstrated the efficacy of LightGBM and the importance of selecting appropriate algorithms for specific physical activity tasks. These studies provide a broader context and perspective for the current study, as they address similar research questions and use similar methods and devices. However, they also highlight some of the differences and gaps that the current study attempts to fill. For example, the current study focuses on comparing two specific devices, Apple Watch and Fitbit, rather than a range of devices. Furthermore, the study by Ridgers et al. (2021), focusing on adolescents in socio-economically disadvantaged areas, highlights the potential limitations of relying solely on commercial wearables. The lack of significant differences in physical activity levels raises questions about the effectiveness of these devices, especially in specific demographic contexts. It emphasizes the importance of considering socio-economic factors and supplementary interventions for a holistic approach to promoting physical activity among youth.

In conclusion, the identification of LightGBM as the best-performing model with Fitbit data, showcasing higher accuracy than Apple watch data, is a noteworthy outcome in the context of this study. However, it is essential to approach these findings with a critical perspective, acknowledging the influence of contextual variations and contrasting results from other studies. Firstly, the generalizability of model performance across different studies can be challenging due to variations in data characteristics, participant demographics, and study designs. Models that excel in one context may not necessarily perform optimally in another. Therefore, while LightGBM may exhibit superior accuracy in this study, it is crucial to recognize that the effectiveness of machine learning models can be highly context-dependent. Moreover, the choice of evaluation metrics can impact the perceived performance of a model. Different studies may prioritize various metrics such as accuracy, precision, recall, or F1 score based on the specific objectives and characteristics of their datasets. Therefore, a model deemed effective in one study based on a specific metric might not demonstrate the same level of effectiveness when considering alternative evaluation criteria. Comparisons with other studies, especially those using different wearable devices or diverse populations, should be approached cautiously. Factors such as device-specific features, user behaviors, and health conditions can significantly influence the performance of predictive models. Consequently, discrepancies in findings may stem from variations in these factors rather than inherent deficiencies in the models themselves. The dynamic nature of technology and advancements in machine learning also contribute to the evolving landscape of predictive analytics. New algorithms, techniques, and models emerge, leading to variations in recommendations across different studies. As technology continues to progress, the superiority of certain models may change, and adopting a flexible approach to accommodate emerging trends is crucial.

5.3 Implication of Findings/Recommendation

Based on these findings and insights, the following recommendations are made to enhance the user experience, effectiveness, and market positioning of commercial wearables, specifically with regards to Apple Watch and Fitbit.

- Implement personalized health insights and recommendations based on individual user data. These wearables can provide tailored advice for improving health and achieving fitness goals by leveraging AI, taking into account gender-specific variations.

- Fine-tune algorithms and sensors to better capture and analyze data for specific activities. This could involve refining heart rate monitoring during different exercises and improving distance accuracy to provide users with more precise feedback on their performance.

- Recognizing the gender-specific differences in heart rates and activity preferences, consider incorporating features that cater to these variations. This could include tailored workout suggestions, recovery strategies, and health insights designed specifically for male and female users.

- Promote compatibility and interoperability between wearables and other health-related devices or applications. This could foster a more holistic approach to health management by allowing users to integrate data seamlessly from multiple sources.
- Collaborate with research institutions and health professionals to further validate the accuracy and reliability of the wearables' health-related metrics. This collaboration can contribute to the credibility of the devices and their potential integration into healthcare practices.

5.4 Limitation/Future Research

The following are specific limitations and corresponding areas of future research directions based on the findings of this research:

1. **Demographic Representation:** The absence of users in the 40-45 age range in the dataset poses a limitation in understanding the preferences and behaviours of this demographic. The findings may not be generalized to this age group, impacting the overall representativeness of the study. Future research can be targeted on the 40-45 age group, exploring their preferences, technological adoption patterns, and health trends, could provide insights into the reasons behind their absence in the wearable user base.
2. **Wearable Device Dependency:** The study focused on Apple Watch and Fitbit data, limiting the generalizability of findings to users of other wearable devices. Future research should consider a more diverse set of wearables for a comprehensive analysis.
3. **Model Generalization:** While the selected machine learning models demonstrated efficacy, their generalization to other datasets and contexts may be limited. Assessing model performance across various datasets and scenarios is essential for broader applicability. Future research can consider this.
4. **Interpretability Challenges:** Despite efforts to enhance interpretability using Shapley values, the interpretability of complex machine learning models remains a challenge.

Future research should explore additional techniques to improve model interpretability further.

5. **Data Imbalance:** The class distribution in the dataset was not be perfectly balanced, potentially impacting the models' ability to generalize well to minority classes. Addressing class imbalance and exploring techniques to handle it could improve model performance in future research.
6. **Real-Time Monitoring:** This study did not explore the potential of real-time monitoring using wearable devices. Future research could explore the effectiveness of implementing models that are capable of providing instantaneous feedback and personalized recommendations based on users' current activities.

Reference List

Amit, S., Talasila, V., Nagaveni, C.R., Subbaraya, C.K., Mahesh, D.S. and Premnath, P.M. (2020). *Design and Development of Wearable Antenna integrated with RF Transceiver for Health Monitoring System*. [online] IEEE Xplore. doi:10.1109/RTEICT49044.2020.9315674.

Arvidsson, D., Fridolfsson, J. and Börjesson, M. (2019). Measurement of Physical Activity in Clinical Practice Using Accelerometers. *Journal of Internal Medicine*, 286(2). doi:https://doi.org/10.1111/joim.12908.

Bansal, R., Singh, J. and Kaur, R. (2020). *Machine Learning and Its applications: a Review*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/338685514_Machine_learning_and_its_applications_A_Review.

Bayoumy, K., Gaber, M., Elshafeey, A., Mhaimeed, O., Dineen, E.H., Marvel, F.A., Martin, S.S., Muse, E.D., Turakhia, M.P., Tarakji, K.G. and Elshazly, M.B. (2021). Smart Wearable Devices in Cardiovascular care: Where We Are and How to Move Forward. *Nature Reviews Cardiology*, [online] 18, pp.1–19. doi:10.1038/s41569-021-00522-7.

Bianchim, M.S., McNarry, M.A., Barker, A.R., Williams, C.A., Denford, S., Thia, L.P., Evans, R. and Mackintosh, K.A. (2023). A Machine Learning Approach for Physical Activity Recognition in Cystic Fibrosis. *Measurement in Physical Education and Exercise Science*, pp.1–10. doi:https://doi.org/10.1080/1091367x.2023.2271444.

Blond, K., Brinkløv, C.F., Ried-Larsen, M., Crippa, A. and Grøntved, A. (2019). Association of High Amounts of Physical Activity with Mortality risk: A Systematic Review and meta-analysis. *British Journal of Sports Medicine*, [online] p.bjsports-2018-100393. doi:10.1136/bjsports-2018-100393.

Brickwood, K.-J., Watson, G., O'Brien, J. and Williams, A.D. (2019). Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR mHealth and uHealth*, 7(4), p.e11819. doi:10.2196/11819.

Brownlee, J. (2021). *Ensemble Learning Algorithms with Python*. Machine Learning Mastery.

Bunn, J., Navalta, J., Fountaine, C. and Reece, J. (2018). Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *International Journal of Exercise Science*, [online] 11(7), pp.503–515. Available at: https://digitalcommons.wku.edu/ijes/vol11/iss7/1/ [Accessed 8 Jan. 2024].

Case, M.A., Burwick, H.A., Volpp, K.G. and Patel, M.S. (2015). Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data. *JAMA*, 313(6), p.625. doi:10.1001/jama.2014.17841.

Chai, C.P. (2020). The Importance of Data Cleaning: Three Visualization Examples. *CHANCE*, 33(1), pp.4–9. doi:https://doi.org/10.1080/09332480.2020.1726112.

Chicco, D. and Jurman, G. (2020). The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*, 21(1). doi:https://doi.org/10.1186/s12864-019-6413-7.

Chicco, D. and Jurman, G. (2023). The Matthews Correlation coefficient (MCC) Should Replace the ROC AUC as the Standard Metric for Assessing Binary Classification. *BioData Mining*, 16(1). doi:https://doi.org/10.1186/s13040-023-00322-4.

Çodur, M.K. (2023). Ensemble Machine Learning Approaches for Prediction of Türkiye’s Energy Demand. *Energies*, 17(1), pp.74–74. doi:https://doi.org/10.3390/en17010074.

Conn, V.S., Hafdahl, A.R. and Mehr, D.R. (2011). Interventions to Increase Physical Activity among Healthy Adults: Meta-Analysis of Outcomes. *American Journal of Public Health*, 101(4), pp.751–758. doi:10.2105/ajph.2010.194381.

Cosoli, G. and Scalise, L. (2019). Accuracy and Metrological Characteristics of Wearable Devices: A Systematic Review. *Lecture Notes in Electrical Engineering*, pp.377–387. doi:10.1007/978-3-030-04324-7_47.

Cosoli, G., Spinsante, S. and Scalise, L. (2020). Wrist-worn and Chest-strap Wearable Devices: Systematic Review on Accuracy and Metrological Characteristics. *Measurement*, p.107789. doi:10.1016/j.measurement.2020.107789.

de Rooij, M. and Weeda, W. (2020). Cross-Validation: A Method Every Psychologist Should Know. *Advances in Methods and Practices in Psychological Science*, 3(2), pp.248–263. doi:https://doi.org/10.1177/2515245919898466.

Evenson, K.R., Goto, M.M. and Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1). doi:10.1186/s12966-015-0314-1.

Foody, G.M. (2023). Challenges in the Real World Use of Classification Accuracy metrics: from Recall and Precision to the Matthews Correlation Coefficient. *PLOS ONE*, 18(10), pp.e0291908–e0291908. doi:https://doi.org/10.1371/journal.pone.0291908.

Fuller, D., Anaraki, J.R., Simango, B., Rayner, M., Dorani, F., Bozorgi, A., Luan, H. and A Basset, F. (2021). Predicting lying, sitting, Walking and Running Using Apple Watch and Fitbit Data. *BMJ Open Sport & Exercise Medicine*, [online] 7(1), p.e001004. doi:10.1136/bmjsem-2020-001004.

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M.A., Simango, B., Buote, R., Heerden, D.V., Luan, H., Cullen, K., Slade, L. and Taylor, N.G.A. (2020). Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR mHealth and uHealth*, [online] 8(9), p.e18694. doi:10.2196/18694.

Germini, F., Noronha, N., Borg Debono, V., Abraham Philip, B., Pete, D., Navarro, T., Keepanasseril, A., Parpia, S., de Wit, K. and Iorio, A. (2022). Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature. *Journal of Medical Internet Research*, 24(1), p.e30791. doi:10.2196/30791.

Giri, J. and Lengyel, A. (2023). Explainable Machine Learning Approach to Yield and Quality Improvements Using Deep Topological Data Analytics. *ASME 2023 International Technical*

Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems. doi:https://doi.org/10.1115/ipack2023-111525.

Hardekopf, P. and Fountaine, C. (2017). Validity of The DiagnostixTM 2100 And Fitbit Charge HRTM In Assessing Heart Rate. *Medicine & Science in Sports & Exercise*, 49, p.762. doi:10.1249/01.mss.0000519028.25537.34.

Harrison, D., Marshall, P., Bianchi-Berthouze, N. and Bird, J. (2015). *Activity tracking: barriers, Workarounds and Customisation*. [online] ACM Digital Library. doi:10.1145/2750858.2805832.

Jackson, C.L., Ward, J.B., Johnson, D.A., Sims, M., Wilson, J. and Redline, S. (2019). Concordance between Self-reported and Actigraphy-assessed Sleep Duration among African-American Adults: Findings from the Jackson Heart Sleep Study. *Sleep*, 43(3). doi:10.1093/sleep/zsz246.

Kanady, J.C., Ruoff, L., Straus, L.D., Varbel, J., Metzler, T., Richards, A., Inslicht, S.S., O'Donovan, A., Hlavin, J. and Neylan, T.C. (2020). Validation of Sleep Measurement in a Multisensor Consumer Grade Wearable Device in Healthy Young Adults. *Journal of Clinical Sleep Medicine*, 16(6), pp.917–924. doi:10.5664/jcsm.8362.

Kanjilal, R. and Uysal, I. (2021). The Future of Human Activity Recognition: Deep Learning or Feature Engineering? *Neural Processing Letters*. doi:https://doi.org/10.1007/s11063-020-10400-x.

Kimm, H., Paik, I. and Kimm, H. (2021). *Performance Comparision of TPU, GPU, CPU on Google Colaboratory over Distributed Deep Learning*. [online] IEEE Xplore. doi:https://doi.org/10.1109/MCSoC51149.2021.00053.

Kit, M., Weldon, A., Yong, J., Lam, C.-P. and Sampaio, J. (2023). Using Machine Learning Algorithms to Pool Data from Meta-Analysis for the Prediction of Countermovement Jump Improvement. *International Journal of Environmental Research and Public Health*, 20(10), pp.5881–5881. doi:https://doi.org/10.3390/ijerph20105881.

Krey, M. (2020). *Wearable Technology in Health Care – Acceptance and Technical Requirements for Medical Information Systems*. [online] IEEE Xplore. doi:10.1109/ICIM49319.2020.244711.

Krittanawong, C. (2018). The Rise of Artificial Intelligence and the Uncertain Future for Physicians. *European journal of internal medicine*, [online] 48, pp.e13–e14. doi:10.1016/j.ejim.2017.06.017.

LandJr, W.H. and Schaffer, J.D. (2019). The Generalized Regression Neural Network Oracle. *Springer eBooks*, pp.77–105. doi:https://doi.org/10.1007/978-3-030-18496-4_3.

Lewis, Z. H., Lyons, E. J., Jarvis, J. M. and Baillargeon, J. (2015). Using an electronic activity monitor system as an intervention modality: A systematic review. *BMC Public Health*, 15(1). doi:10.1186/s12889-015-1947-3.

Lewis, Z.H., Pritting, L., Picazo, A.-L. and JeanMarie-Tucker, M. (2020). The Utility of Wearable Fitness Trackers and Implications for Increased engagement: an exploratory, Mixed Methods Observational Study. *Digital Health*, 6, p.205520761990005. doi:10.1177/2055207619900059.

Liu, L., Hou, Y., He, J., Lungu, J. and Dong, R. (2020). An Energy-Efficient Fall Detection Method Based on FD-DNN for Elderly People. *Sensors*, 20(15), p.4192. doi:10.3390/s20154192.

Manne, R. and Kantheti, S.C. (2021). Application of Artificial Intelligence in Healthcare: Chances and Challenges. *Current Journal of Applied Science and Technology*, pp.78–89. doi:10.9734/cjast/2021/v40i631320.

McGrow, K. (2019). Artificial Intelligence. *Nursing*, 49(9), pp.46–49. doi:10.1097/01.nurse.0000577716.57052.8d.

Melton, B.F., Buman, M.P., Vogel, R.L., Harris, B.S. and Bigham, L.E. (2016). Wearable Devices to Improve Physical Activity and Sleep. *Journal of Black Studies*, 47(6), pp.610–625. doi:https://doi.org/10.1177/0021934716653349.

Mi, J.-X., Li, A.-D. and Zhou, L.-F. (2020). Review Study of Interpretation Methods for Future Interpretable Machine Learning. *IEEE Access*, 8, pp.191969–191985. doi:https://doi.org/10.1109/access.2020.3032756.

Mondal, S., Zehra, N., Choudhury, A. and Iyer, P.K. (2020). Wearable Sensing Devices for Point of Care Diagnostics. *ACS Applied Bio Materials*. doi:10.1021/acsabm.0c00798.

Nelson, M.B., Kaminsky, L.A., Dickin, D.C. and Montoye, A.H.K. (2016). Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. *Medicine & Science in Sports & Exercise*, 48(8), pp.1619–1628. doi:10.1249/mss.0000000000000933.

Nweke, H.F., Teh, Y.W., Al-garadi, M.A. and Alo, U.R. (2018). Deep Learning Algorithms for Human Activity Recognition Using Mobile and Wearable Sensor networks: State of the Art and Research Challenges. *Expert Systems with Applications*, [online] 105, pp.233–261. doi:https://doi.org/10.1016/j.eswa.2018.03.056.

Panesar, A. and Panesar, H. (2020). Artificial Intelligence and Machine Learning in Global Healthcare. *Handbook of Global Health*, pp.1–39. doi:10.1007/978-3-030-05325-3_75-1.

Qizi, A.A.A., Kadambayevich, K.O. and Atabaevich, K.D. (2023). Data Collection Approached in Echocardiography for Cardiovascular Disease Classfication. *Journal of Science, Research and Teaching*, [online] 2(12), pp.50–54. Available at: https://jsrt.innovascience.uz/index.php/jsrt/article/view/353 [Accessed 8 Jan. 2024].

Rapp, A. and Cena, F. (2016). Personal Informatics for Everyday life: How Users without Prior self-tracking Experience Engage with Personal Data. *International Journal of Human-Computer Studies*, [online] 94, pp.1–17. doi:10.1016/j.ijhcs.2016.05.006.

Raykar, S.S. and Shet, V.N. (2020). Cognitive Analysis of Data Mining Tools Application in Health Care Services. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. doi:10.1109/ic-etite47903.2020.442.

Ray, P., Nagarajan, G. and Minu, R.I. (2022). *Applied Edge AI: concepts, platforms, and Industry Use Cases*. Boca Raton: Auerbach.

Ridgers, N.D., Timperio, A., Ball, K., Lai, S.K., Brown, H., Macfarlane, S. and Salmon, J. (2021). Effect of Commercial Wearables and Digital Behaviour Change Resources on the Physical Activity of Adolescents Attending Schools in socio-economically Disadvantaged areas: the RAW-PA cluster-randomised Controlled Trial. *International Journal of Behavioral Nutrition and Physical Activity*, 18(1). doi:https://doi.org/10.1186/s12966-021-01110-1.

Saleh, Z. (2019). *Artificial Intelligence Definition, Ethics and Standards*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards.

Simon, H. A. (1983). Why Should Machines Learn? In: Machine learning. Morgan Kaufmann, Los Altos, pp 25–37.

Sunny, J.S., Patro, C.P.K., Karnani, K., Pingle, S.C., Lin, F., Anekoji, M., Jones, L.D., Kesari, S. and Ashili, S. (2022). Anomaly Detection Framework for Wearables Data: A Perspective Review on Data Concepts, Data Analysis Algorithms and Prospects. *Sensors*, 22(3), p.756. doi:10.3390/s22030756.

Sypes, E.E., Newton, G. and Lewis, Z.H. (2019). Investigating the Use of an Electronic Activity Monitor System as a Component of Physical Activity and Weight-Loss Interventions in Nonclinical Populations: A Systematic Review. *Journal of Physical Activity and Health*, 16(4), pp.294–302. doi:10.1123/jpah.2017-0660.

Than, A., Zan, P. and Chen, P. (2020). Transdermal Theranostics. *View*, 1(2). doi:10.1002/viw2.21.

Tokucoglu, F. (2018). Monitoring Physical Activity with Wearable Technologies. *Archives of Neuropsychiatry*. doi:https://doi.org/10.29399/npa.23333.

Vatandsoost, M. and Litkouhi, S. (2019). The Future of Healthcare Facilities: How Technology and Medical Advances May Shape Hospitals of the Future. *Hospital Practices and Research*, [online] 4(1), pp.1–11. doi:10.15171/hpr.2019.01.

Wang, Q. and Jiang, Y. (2023). Leisure Time Prediction and Influencing Factors Analysis Based on LightGBM and SHAP. *Mathematics*, 11(10), pp.2371–2371. doi:https://doi.org/10.3390/math11102371.

Wang, S. (2020). A Comprehensive Review of Wearable Applications and Material Construction. *Open Journal of Applied Sciences*, 10(06), pp.364–408. doi:10.4236/ojapps.2020.106027.

Wang, Y., Cang, S. and Yu, H. (2019). A Survey on Wearable Sensor Modality Centred Human Activity Recognition in Health Care. *Expert Systems with Applications*, [online] 137, pp.167–190. doi:https://doi.org/10.1016/j.eswa.2019.04.057.

World Health Organization (2015). WHO | Global Recommendations on Physical Activity for Health. *Who.int*. [online] doi:/entity/dietphysicalactivity/factsheet_recommendations/en/index.html.

Xiong, J., Lan, L., Lian, Z. and De dear, R. (2020). Associations of bedroom temperature and ventilation with sleep quality. *Science and Technology for the Built Environment*, pp.1–11. doi:10.1080/23744731.2020.1756664.

Xuan, Q., Wu, J., Shen, J., Ji, X., Lyu, Y. and Zhang, Y. (2020). Assessing cognitive load in adolescent and adult students using photoplethysmogram morphometrics. *Cognitive Neurodynamics*, 14(5), pp.709–721. doi:10.1007/s11571-020-09617-2.

Zhang, Y., Sun, W. and Chen, J. (2022). Application of Embedded Smart Wearable Device Monitoring in Joint Cartilage Injury and Rehabilitation Training. *Journal of Healthcare Engineering*, 2022, pp.1–11. doi:10.1155/2022/4420870.