INTEGRATING MACHINE LEARNING IN WEARABLE HEALTH ANALYSIS

BTECH PROJECT

By

Chinmaya Kumar Palo

Aniket Patro

Surya Pratap Sahu

Kumar Kausik Sahu

Uttam Gouda

Under the Guidance

of

Dr. Neelamadhab Padhy, Professor, Dy. Dean Research

(Computational Science)

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

GIET UNIVERSITY, GUNUPUR

May 2024

INTEGRATING MACHINE LEARNING IN WEARABLE HEALTH ANALYSIS

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

BTECH

By

Chinmaya Kumar Palo

Aniket Patro

Surya Pratap Sahu

Kumar Kausik Sahu

Uttam Gouda

Under the Guidance

of

Dr. Neelamadhab Padhy, Professor, Dy. Dean Research

(Computational Science)

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

GIET UNIVERSITY, GUNUPUR

May 2024

GIET UNIVERSITY, GUNUPUR

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the thesis entitled **INTEGRATING MACHINELEARNING FOR PROCESS MONITORING AND CONTROL IN WEARABLHEALTH ANALYSIS** in the partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** and submitted in the **DISCIPLINE OF COMPUETR SCIENCE ENGINEERING, GIET UNIVERSITY ,GUNUPUR**, is an authentic record of our own work carried out during the time period from 2021 to 2025 under the supervision of Dr. Neela Madhab Padhy, Professor at Department of Computer Science Engineering , GIET UNIVERSITY.

The matter presented in this thesis has not been submitted by us for the award of any other degree of this or any other institute.

Signature of the student with date

(Chinmaya Kumar Palo, Aniket Patro, Surya Pratap Sahu, Kumar Kausik Sahu, Uttam Gouda)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Signature of the Supervisor of B. Tech thesis

Dr. Neelamadhaba Padhy

Chinmaya Kumar Palo, Aniket Patro, Surya Patrap Sahu, Kumar Kausik Sahu, Uttam Gouda have successfully given their B.Tech. Oral Examination held on**<Date of B.Tech. Oral Examination>**.

Signature(s) of Supervisor(s) of B.Tech. thesis	Dean Computational Science, CSE
Date:	Date:

ACKNOWLEDGEMENTS

We extend our heartfelt gratitude to Professor Dr. Neela Madhab Padhy of the Department of Computer Science and Engineering for his invaluable support and guidance throughout the duration of this project. Without his mentorship and encouragement, this endeavor would not have reached fruition.

We also express our sincere appreciation to the Department of Computational Science, School of Engineering and Technology, for their unwavering assistance and continuous support during the course of our research.

Furthermore, we would like to acknowledge the contributions of our friends, family members, and all those who provided unconditional support and encouragement during the execution of this project. Their belief in our capabilities has been a constant source of motivation.

This project, which focuses on research in machine learning, has been conducted under the auspices of Dr. Neelamadhab Padhy, and his expertise has been instrumental in shaping our work and guiding us towards meaningful outcomes.

Once again, we extend our deepest gratitude to all those who have contributed to the successful completion of this project. Your support has been invaluable, and we are truly grateful for your unwavering assistance.

ABSTRACT

Context: This research project aims to analys valuable data gathered from wearable devices like fitness trackers and smartwatches. We're specifically interested in personal attributes like age, body mass index (BMI), and everyday activities tracked by these devices.

Objective: The main objective of this article is to know the health conditions of the person using Wearable device like smartwatches.

Material/Method: Initially we have collected the data from apple watch and Fitbit data having 6400 records. We employ advanced computer algorithms (Logistic regression, KNN, SVC, Random Forest) to uncover patterns hidden within your health data. By linking various health factors, we gain insights into how they impact your well-being. Our approach encompasses all aspects of your health to paint a comprehensive picture, unlike fragmented views that focus solely on isolated data points. We have created a user-friendly tool that combines collected health data and offers practical advice. Finally, our proposed model recommends the personal health inside it and analyses properly. Among all the classifiers, SVC outperformed with 98.12% accuracy.

Conclusion: This article aims to enhance our health awareness through machine learning. It offers personalized insights and actionable guidance based on your health data, empowering you to make informed lifestyle choices and prioritize preventive care. Our system evaluates your health status as "healthy," "healthy but with scope for improvement," or "requiring medical attention," providing tailored recommendations for optimal health outcomes.

In our analysis, we assessed the 4 machine learning models; Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Random Forest. Out of these two the models - KNN with 98.12 % accuracy score and SVC with 98.12 % accuracy reached the highest accuracy. Logistic Regression- 97.65%_accuracy Lipreading. On the other hand, the Random Forest model had an accuracy of 85.9% which is way behind the others. Hence, other than Random Forest all the model has shown a good performance where KNN and SVC are the top most accurate model for this training set.

TABLE OF CONTENTS

LIST OF FIGURES

ACRONYMS (if any)

KNN: K-Nearest neighbor

ML: Machine Learning

SVM: Support Vector Machine

LDA: Linear Discernment Analysis

SVC: Support Vector Machine

SMOTE: Synthetic Minority Oversampling Technique

ROC: Receiver Operating Characteristic

RF: Random Forest

LR: Logistic Regression

DT: Decision Tree

CHAPTER 1

INTRODUCTION

Digital health technologies have been transformed into an effective means of monitoring our physical condition and general wellness through wearable tech. It is true that the utilization of devices such as smart wrist-wands among others help us keep track of many things including body activities metrics daily however insignificant they might seem at first glance. This information if put into use can greatly help in discerning an individual's state of healthología and identify areas that require attention.

The main subject under study is a comprehensive examination of the data of wearable devices, focusing on particulars like the age, body mass index (BMI), and daily movements through which people are tracked by such gadgets. This will involve using such advanced computation methodologies as Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVC), and Logistic Regression to establish correlations in health data that can give personalized information about an individual's well-being.

The main goal of this article is to explore how wearable technologies can be used to assess someone's health and provide recommendations and practical advice on how to improve overall well-being. We demonstrate the effectiveness of our machine learning model in forecasting health outcomes by analyzing a dataset which contains 6400 records collected from Fitbit as well as Apple Watch devices.

We have several crucial steps in our approach. At first, we collect comprehensive health records using wearable devices with information covering various physical features as well as common daily routines. The research centers around information acquired through this channel and for the creation of prediction models concerning a number of factors each evaluating one's own well-being of an individual.

Sophisticated machine learning techniques allow us to find predictive patterns and discover complex correlated between various health factors. Logistic regression is a typical statistical technique that provides knowledge on the chances of occurrence of specific health events given a set of input parameters. SVC aims at getting the best hyperplane which separates data points into distinct categories unlike KNN that relies on distances to group individual's various fields within the medical sphere. Lastly, the Random Forest is known for the fact that it incorporates multiple decision trees which leads to an improved level of prediction due its employment of ensemble learning methods.

The two among them that are most deserving of mention are the KNN and SVC classifiers, with an accuracy rate of 98.12%. This degree of accuracy demonstrates that we have employed data from wearables effectively when analyzing health related issues. Recommendations for each individual's conditions can be given using precise predictions on health outcomes; this allows for early intervention as well as making informed choices about the way we live.

One of the main advantages of our approach lies in its complexity. Instead of focusing only on isolated information, it takes into account a number of factors such as age, BMI or daily habits when assessing personal well-being. Synthesizing these multiple sources allows us to paint a complete portrait of a person's health status, as well as provide some useful recommendations on how they can improve.

The purpose of this article is to show how machine learning can improve health awareness and also drive plans for personalized wellness. Using the volume of data generated from wearable devices, we can offer personalized advice for mental well-being and valuable information about your own health. Our state-of-the-art method seeks to empower individuals to proactively promote healthy lifestyle habits to increase their longevity

CHAPTER 2

LITRETURE SURVEY

et al. Majumder [1]: Health care that is inexpensive is needed as long as people live longer, in this post I will go over how wearable sensors and remote health monitoring can be an ideal cost-effective alternative and an example can be seen here. The systematic investigation into the various monitoring systems available, with special emphasis on textile-containing sensors and… It covers communication technologies and future issues. Encouraging discovery with the distant light of moon to improve and enhance the quality of health care for the aging population in society.

et al. Wan [2]: At the top of this list is affordable health care for the elderly and chronically ill. We are talking about an innovative system known as WISE (Wearable IoT-cloud-based Health Monitoring System). Instead, the health-monitoring system utilizes body sensors to pass data directly to the cloud in real-time and entirely OMFG no smartphone use needed. The WISE further enables cost-effective healthcare by utilising the Internet of Things to optimise and simplify services.

et al. Al-Khafajiy [3]: The extreme care that is currently lacking in the face of an increasingly aging population and the increase in chronic diseases as well as innovations for modern health care. Please enjoy it, and share your thoughts in the comments section. Latest The latest is a new Smart Healthcare Monitoring System (SW-SHMS) that has been designed for real-time continuous monitoring of the elderly and others in their own home. The company said in a statement that with the help of a wearable electronic sensor and application, the system can collect and analyse information resulting from physical exercises, allowing for an early intervention and remote support. This framework is intended to help overcome healthcare challenges and enhance patient health and wellness.

et al. Prieto-Avalos [4]: The article investigated the practical use of wearable technology in the daily monitoring of heart health to prevent cardiovascular disease (CVD), the most massive epidemic of cardiovascular diseases worldwide. It will explore CVD, the untapped potential of these wearables, and how they can be used for vitals monitoring to detect CVD at the earliest stages enabling tailored CVD treatment plans. The talk covers a variety of types of wearables and their associated health metrics, ranging from non-commercial to commercial. Likewise, the review discusses the barriers and facilitators when using mHealth (and specifically Android) apps for CVD management. Finally, it discusses the present status of wearable in monitoring the CVD.

Et al. Kakria [5]: It talks about a high-end system for heart monitoring in real time and as a great success. The technology will allow patients to have care provided to them without a healthcare professional present, using wearable sensors and mobile devices. The system, which has been tested on a sample group comprising 40 individuals, ensures obstacle-free communication of patients with physicians in terms of safety, performance and data privacy. This, it argues, tackles resource scarcity and improves the level of care for heart patients as it guarantees they get medical attention immediately.

et al. Pantelopoulos, A. [6]: In this paper, a Prognosis as a new technology model is proposed for wearable health monitoring systems (WHMS) to monitor certain physiological parameters of a patient in real-time in order to ensure high quality patient care. A fuzzy regular language model to predict potential health problems with stochastic Petri nets for analysing interactions among individuals and devices. Suitable for chronic diseases and acute attacks, to provide personalized and powerful medical services.

Et al. Ahmed, Z. U. [7]: This paper presents for the realization of healthcare system which monitors the health condition of patient from remote location using IoT and inform healthcare professionals as well as the family members about the changes in patient health during the period. With the power of the Internet of Things, availability can be greatly increased, and costs can be reduced, resulting in more healthcare services and more lives saved. This research paper covers background, system design, device characteristics, performance evaluation and future work.

et al. Evangeline, C. S. [8]: Human Health Monitoring System (HHMS) is a new way of early disease diagnosis and monitoring of its recovery. Sensors can be used to keep a check on patients life signals and geo-location so that data can be sent to care givers without involvement of middle men — all this can be achieved through IoT and GSM tech. HHMS has been demonstrated in a prototype and independently validated in simulated case studies providing lifelong and affordable monitoring. This not only makes the system more efficient, but also cheaper than those that have been tried, he says. If the wireless facility is attainable; its instant capability offers a quicker response. HHMS intends to decrease the cost of healthcare and improve overall healthcare management through remote healthcare solutions.

et al. Taştan, M. [9]: In this paper, we introduce an innovative smartphone application for people with heart disease. Wearable sensors track heart rate, heart rate variability and body temperature and transmit the data to the app via wireless communications. When a critical value is detected, users are notified and receive real time location updates. It also provides email and Twitter notifications for continuous monitoring and social support. This app allows for early identification of cases of heart problems thanks to IoT technology and results in faster medical intervention, therefore better results. Its emphasis is on propelling heart care forward, personalized and pre-emptive.

et al. Sumathy, B. [10]: Modern IoMT wearable technology keeps monitoring of pulse, breathing, and basic temperature using smart devices worn on the chest, and it is intended for use on patients who are chronically ill or elderly. The sensors send the data to the cloud for analysis, so that a remote consultation could be carried out by experts, in case there is any major change in the flow. Pandemics like COVID-19 make temperature monitoring especially important. Hot dark-haired babe gets fucked and churned out by this sucker with monstrous tool devises to give a distant medic and limit the number of times being viewed in hospital A bill split. Increasingly, studies indicate that IoT-enabled solutions that provide real-time information and data-driven insights have the potential to deliver quicker response time along with improved patient outcomes. This monitors the patients continuously that very important at the time of emergency with the help of wireless data & alerts, which genuinely fastens the access to healthcare.

et al. Vijayan, V. [11]: This article looks at how wearable sensor technologies are making inroads in the healthcare space. These sensors are crucial to supporting telecare, telemedicine and real-time patient monitoring (RPM). Wearable technology has the potential to provide healthcare providers new clinical information that may translate into better patient outcomes and in some cases, cost savings. The discussion will also cover the advantages of sensors, long-term monitoring, and customized therapeutic strategies and provide an overview of next-generation sensor technologies and self-care interventions. We also explore modern technology like combining data analysis with deep learning, taking into consideration the question of security and accuracy. In conclusion, the article illustrates how wearable sensors can change healthcare forever as we know it.

et al. Vijayan, V [12]: Application of Wearable sensors in Healthcare (Review):- This research analyses the increasing application of wearable sensors in healthcare to track human health parameters and behaviour. What the report demonstrated is an uptrend in the adoption of these devices for home healthcare, remote monitoring as well as point of care. This review is focused largely on the monitoring of people with musculoskeletal conditions, spanning current apps in clinical and self-care environments. The content will focus on the current issues and progress in topics like data protection, accuracy of measurement, and deep learning for automatic activity and sleep detection. Thus, the results from the study suggest an overarching narrative of wearable sensor technology being able to provide a way for patients to be remote monitored, for their care to be supported and for their data to be objective.

Et al. Zhou, S. [13]: For a work published in Nature Human Behaviour in 2017, they observed 18 seniors who used their wearable for several months beginning from February. The participants had monitored metrics such as heart health, blood pressure, steps taken in a day, their circadian rhythms (sleeping patterns), and mood. Varying heart rate, reduced blood pressure and steps count were linked to good health, the results showed. However, some participants complained that the device made them feel tired. Wearables had social benefits, such as improving social interactions according to interviews. The study has implications for how wearables might promote social engagement among seniors and serve as a resource for health monitoring proactively, the researchers say.

et al. Asthana, S [14]: Meet health Advisor: the ultimate top recommendation tool for personal health monitoring wearables will suggest the appropriate technology depending on the history and background of the user. Machine Learning is applied to understand and predict possibles diseases and tie them to relevant metrics. This will be followed by Text analysis using suitable wearable devices which will measure the same. This not only lets the users control their health but it also assists developers to understand in the trends in wearables. The personalized recommendations by Health Advisor ensures to offer the best healthcare consultation and remedies in an affordable price.

et al. Arshad, A [15]: This paper presents a range of approaches for tracking mobility of the elderly in the context of healthcare facilities, aimed at reducing costs and improving care quality. Pros and Cons of Various Monitoring Techniques: We survey monitoring techniques including wearable devices, wireless sensors and floor sensors. The mission is to find out where to make health monitoring better. We emphasize the surging demand for authentic monitoring systems which provide consistent healthcare assistance specifically to the geriatric population. We hope to enhance care for the elderly by reviewing what we currently know and suggesting what work needs to be done in the future.

et al. Sharma, A [16]: This will be the first in a series of paper that we write on this trend in healthcare, focusing on the growth of personalized healthcare as a result of the need to manage a growing population of chronic disease patients, and the increase of the aging population. Wearable sensors possessing biosensing capabilities are becoming increasingly prominent, allowing for biomarkers from body fluids to be monitored constantly. It concerns only the non-invasive optical and electrochemical biosensors. These wearables, that features microfluidics and multiplexing technology, is seamlessly integrated into wearable textiles that can conform to the body. This discovery could increase our understanding of disease, resulting in earlier detection and specific treatment response assessments.

et al. González-Valenzuela, S [17]: You will be surprised at how using a few sensors can bring a lot of benefits with minimal cost, as evidenced by research in the field. This solution uses two networks, one of them is an access point which stays on its current/ permanent location and the other one a sensor network, which is bandaged up to a user to measure the vital statistics, remotely. In case of motion leading to a weakening signal, the access point switches the data transmission to a wearable sensor that has a better signal strength. Experimental results prove this method of transmission to be much better in terms of minimizing data loss when worn on hand, and when the subject walks (based on wrist reading).

et al. Esposito, M. [18]: This study examines the importance of providing easily accessible health care for an older population suffering from chronic diseases. It proposes a flexible and adaptable mobile design for the development of personalized health monitoring applications that can be used on wearable devices and smartphones. The architecture consists of four layers and can be easily adapted to meet different needs, using an ontology-based data model for communication (verified by ALMA). A case study of a cardiac arrhythmia monitoring application demonstrates the effectiveness of the system in real-world scenarios.

et al. Jiang, W [19]: We aim to determine the impact of COVID-19 on respiratory, mental and cardiovascular health in this study. PROCESSING CONCLUSIONS Conclusion assess the current monitoring methods and some limitations but on the other hand we proposed a wearable telehealth solution. The template reports include temperature (T), heart rate (HR), oxygen saturation (SpO2), and respiratory rate (RR). They will also predict lung function as well with the AI and sensor fusion technology. It could be used as unequivocal patient monitoring system for those who suffer from COVID-19 or chronic disease. Our initial prototypes showed some very promising results and even outperformed some commercial systems. Ultimately, our aim is to enhance patient care and health.

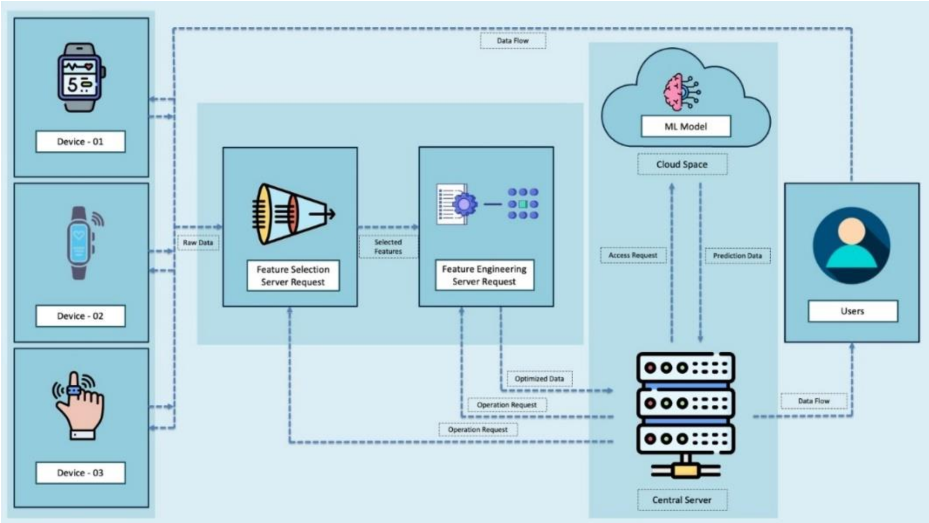
et al. Li, X. [20]: In this research, we explore how wearable biosensors can be used to diagnose disease and manage health by monitoring changes in the body during everyday tasks. Our study involved following 43 participants and tracking their daily fluctuations as well as significant changes during travel and other activities. These variations showed how external factors can affect health, correlating with feelings of fatigue and overall health. Thanks to data obtained from biosensors and medical tests, we were able to detect Lyme disease, inflammation and insulin sensitivity early. This study highlights the potential of wearable biosensors for personalized health monitoring and disease diagnosis.

et al Teixeira, E. [21]: While wearable devices (WDs) are commonly used to track daily activity data, there are concerns about their accuracy, particularly in older individuals. Studies show that WDs designed for younger users may not provide accurate health measurements for this age group. It is imperative that research continues as WD technology advances to ensure their effectiveness. This article examines data on the effectiveness of WD in health and activity monitoring in older individuals to help users and researchers select the most appropriate monitoring device for this demographic.

CHAPTER 3

PROPOSED METHODODLOGY

Proposed Methodology



(Fig: 3.1 Proposed Methodology)

This figure is the data flow design related to processing of wearable data and its system to predict it to wearable users. This starts with ingesting data from a multitude of wearable devices. But before entering the modelling, these raw data is passed into two processes, first is Feature Selection and Feature Engineering. This server process is fired by server request due to an operation request.

After cleaning the data, the data is broadcasted to the Central server. Here we send a request to execute for a machine learning model on cloud at the same location where we sent that request. These predictions are then sent back to Central server, and then to that user. These predictions will be available to the user on any screen that is capable of showing this information.

Basically, this diagram is a way to visualize a simplified end to end flow of data collection through prediction, using Azure and 5G server/cloud resources to manage the processing and delivery of the information to the customers and end-users of this solution.

Phase 1: Different Sources of the Data

Wearable healthcare devices collect diverse data for machine learning models. Sensors in these devices capture physiological data like heart rate, movement, sleep patterns, and even emotional arousal. Users contribute by self-reporting medical history, lifestyle habits, and symptoms. Some devices gather environmental data like temperature and air quality. Additionally, research studies and clinical trials offer valuable datasets under controlled conditions. Ensuring data privacy and security through anonymization is crucial, and ethical considerations should be addressed. This multifaceted data fuels the development of machine learning models that can offer personalized healthcare insights and predictions.

Phase 2: Data preprocessing techniques (feature selection, feature engineering etc)

A prepossessing of raw data to show how we did to obtain the raw data from the wearable devices and enrich the quality of the model and bring some new information. New attributes were created using feature engineering techniques such as BMI (derived height and weight) and steps per second (derived step count and time intervals). These derived features give a more complete picture of the user's wellness and movement. Feature selection techniques were also employed to select the most important features for training the model and also hence reducing noise and increasing efficiency. This guarantees that the most significant information is at the forefront, allowing for precise predictions, and customized health care recommendations!

Phase 3: Cloud Space

Built on top of cloud storage, our Wearable Healthcare Analysis Project has all the data in order, and ready for us to easily manage and access it. All data, raw sensor readings, pre-processed attributes and the trained machine learning model is stored and secured together in the cloud. This makes it easy to access and analyse on demand, without having the drag of local storage. Secondly, by this being a cloud infrastructure, the result is real-time update and user input and model refinements are made available in the next second. This uni-directional, dynamic synchronization not only scales the project to massive scale but also enables real-time health insights — This will be one of my powerful personalized healthcare intervention/plays!

Phase 4:ML models

Logistic Regression: Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome. **Naïve Bayes**: Naive Bayes is a probabilistic classifier based on Bayes' Theorem, assuming feature independence which is a quite unrealistic condition in order to simplify calculations and is extremely efficient.
KNN: K-nearest neighbours (KNN) is a simple and effective algorithm for classifying data points based on the majority vote of its K-nearest neighbours. **SVM**: Support Vector Machine (SVM), An SVM is one of the most powerful Machine Learning Algorithm, because given your data SVM will find the optimal hyperplane that help separate your data points into different classes. **XGBoost**: XGBoost stands for eXtreme Gradient Boosting, which is a popular powerful ensemble learning algorithm that combines multiple decision tree models and claims to be the state of the art machine learning algorithm for tabular/structured data. **Random Forest**: Random Forest is a part of an ensemble method that performs multiple decision trees to have a correct prediction and averages their outcome. **Decision Tree :** A decision tree is a tree-shaped model of decisions and their consequences; it is a tree-like graph that is used to determine a course of action. **LDA** : LDA (Latent Dirichlet Allocation) – is a generative statistical model that allows detecting topics in a collection of documents

1. **Logistic Regression: -**

- Description: Logistic Regression is a linear model used for binary classification. It predicts the probability that a given input belongs to a certain class.
- Equation: The logistic regression model uses the logistic function (sigmoid function) to model the probability:

$$P(Y=1|X) = \frac{1}{1+e^{-\beta^T X}}$$

where $P(Y=1|X)$ is the probability of the positive class given input X, and β represents the model parameters.

2. **Naïve Bayes: -**

- Description: Naïve Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of independence between features.
- Equation: Bayes' theorem is applied:

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)}$$

where $P(C_k|X)$ is the posterior probability of class C_k given feature vector X, $P(X|C_k)$ is the likelihood of observing feature vector X given class C_k, $P(C_k)$ is the prior probability of class C_k, and $P(X)$ is the probability of observing the feature vector X.

3. **KNN: -**

- Description: KNN is a non-parametric method used for classification and regression. It classifies data points based on the majority class among their nearest neighbors.
- Equation: No explicit equation for KNN, it operates by calculating the distance between data points and finding the k-nearest neighbors based on some distance metric, such as Euclidean distance.

4. **SVM: -**

- Description: SVM is a supervised learning model used for classification and regression analysis. It finds the hyperplane that best separates the classes in the feature space.
- Equation: In the case of linear SVM, the decision boundary is represented as $w^T x + b = 0$, where w is the weight vector, x is the input vector, and b is the bias term.

5. **XGBoost: -**

- Description: XGBoost (Extreme Gradient Boosting) is an ensemble learning method based on decision trees, designed to optimize performance and computational speed.
- Equation: The objective function in XGBoost consists of a sum of a loss function and a regularization term:

$$\text{Objective} = \sum_i^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

where L is the loss function, $\hat{y}_i$ is the predicted value for observation i, $\Omega(f_k)$ is the regularization term for tree k.

6. **Random Forest: -**

- Description: Random Forest is an ensemble learning method that constructs a multitude of decision trees at training time and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.
- Equation: No explicit equation, as it's an ensemble method made up of multiple decision trees.

7. **Decision Tree: -**

- Description: Decision Tree is a supervised learning algorithm used for classification and regression. It partitions the data into subsets based on the values of features and makes decisions at each node.
- Equation: No specific equation, decision trees are hierarchical structures where decisions are made at each node based on the value of a feature.

8. **LDA: -**

- Description: LDA is a dimensionality reduction and classification technique. It seeks to find the linear combination of features that best separate two or more classes.
- Equation: The goal is to maximize the between-class scatter while minimizing the within-class scatter. The projection $y = W^T x$ is chosen to maximize the ratio of between-class scatter to within-class scatter, where W is the transformation matrix.

CHAPTER 4

RESULTS AND DISCUSSION / ANALYSIS

SMOTE Technique:

SMOTE, or Synthetic Minority Oversampling Technique, is a statistical technique used in machine learning to balance class distribution in datasets. It's used when the class being analyzed is underrepresented in the data. SMOTE works by creating synthetic minority class samples from existing minority class samples.

Before SMOTE:

Table 4.1

Parameter /Model	LR	Naïve Bayes	KNN	SVM	XGBoost	RF	DT	LDA
Accuracy	93.77	49.92	97.6	95.37	93.46	95.05	75.59	89.15
Precision	94.06	24.92	97.63	96.16	94.94	95.95	62.3	90.07
Recall	93.77	49.92	97.6	95.37	93.46	95.05	75.59	89.15
F1 Score	93.82	33.24	97.61	9538	93.62	95.06	67.08	89.38
AUC-ROC	99.23	50	99.29	99.29	98.002	99.29	83.19	95.87
Sensitivity	87	0	94	83	82	82	0	83
Specificity	95	0	98	94	91	93	69	91
Cohen's Kappa	90.08	0	96.16	92.61	89.63	92.1	60.39	83.14

Accuracy: -

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome.

We got the accuracy for different models applied in our machine learning article is followed by**:** Logistic Regression = 93.77, Naive Bayes = 46.62, KNN = 97.6, SVM = 95.37, XGBoost = 93.46, Random Forest = 95.05, Decision Tree = 75.59, LDA = 89.15

Precision: -

Precision is a metric that measures how often a machine learning model correctly predicts the positive class.

We got the precision for different models applied in our machine learning article is followed by: Logistic Regression = 94.06, Naive Bayes = 24.92, KNN =

97.63, SVM = 96.16 XGBoost = 94.94, Random Forest = 95.95, Decision Tree = 62.3, LDA = 90.07.

Recall: -

Recall is a metric that measures how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset.

We got the precision for different models applied in our machine learning article is followed by: Logistic Regression = 93.77, Naive Bayes = 46.62, KNN = 97.6, SVM = 95.37 XGBoost = 93.46, Random Forest = 95.05, Decision Tree = 75.59, LDA = 89.15.

F1 Score: -

The F1 score or F-measure is described as the harmonic mean of the precision and recall of a classification model. The two metrics contribute equally to the score, ensuring that the F1 metric correctly indicates the reliability of a model.

We got the F1 Score for different models applied in our machine learning article is followed by: Logistic Regression = 93.82, Naive Bayes = 33.24, KNN = 97.61, SVM = 95.38 XGBoost = 93.62, Random Forest = 95.06, Decision Tree = 67.08, LDA = 89.38.

AUC-ROC Score: -

An ROC curve, or receiver operating characteristic curve, is like a graph that shows how well a classification model performs. It helps us see how the model makes decisions at different levels of certainty.

We got the AUC-ROC Score for different models applied in our machine learning article is followed by: Logistic Regression = 99.23, Naive Bayes = 50, KNN = 99.29, SVM = 99.29, XGBoost = 90.002, Random Forest = 99.92, Decision Tree = 83.19, LDA = 95.87

Sensitivity: -

Sensitivity measures how well a machine learning model can detect positive instances. In other words, it measures how likely you will get a positive result when you test for something.

We got the Sensitivity for different models applied in our machine learning article is followed by: Logistic Regression = 87, Naive Bayes = 0, KNN = 94, SVM = 83, XGBoost = 82, Random Forest = 82, Decision Tree = 0, LDA = 83.

Specificity: -

Specificity talks about the number of negative records correctly predicted. So, with Specificity, we can measure how well our model predicts the class that we want to declare as the *negative* class.

We got the Specificity for different models applied in our machine learning article is followed by: Logistic Regression = 95, Naive Bayes = 0, KNN = 98, SVM = 94, XGBoost = 91, Random Forest = 93, Decision Tree =69, LDA = 91.

Cohen's Kappa: -

The Kappa Coefficient, commonly referred to as Cohen's Kappa Score, is a statistic used to assess the effectiveness of machine learning classification models.

We got the Cohen's Kappa for different models applied in our machine learning article is followed by: Logistic Regression = 90.08, Naive Bayes = 0, KNN = 96.16, SVM = 92.61, XGBoost = 89.63, Random Forest = 92.1, Decision Tree = 60.39, LDA = 83.14

After SMOTE:

Table 4.2

Parameters/Models	LR	Naïve Bayes	KNN	SVM	XGBoost	RF	DT	LDA
Accuracy	98.12	46	98.59	99.06	94.36	99.06	83.09	91.07
Precision	98.12	21.16	98.73	99.06	94.84	99.12	73.96	91.77
Recall	98.12	46	98.59	99.06	94.36	99.06	83.09	91.07
F1 Score	98.12	28.99	98.61	99.06	94.38	99.07	77.45	91.23
AUC-ROC Score	99.92	50	50	50	99.68	99.69	85.71	97.23
Sensitivity	99.9	98	100	100	99	98	100	92
Specificity	100	100	99	98	99.56	100	98.23	99.3
Cohen's Kappa	96.87	0	97.67	98.43	90.75	98.44	70.53	85.27

Accuracy: -

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome. We got the accuracy for different models applied in our machine learning article is followed by: Logistic Regression = 98.12, Naive Bayes = 46, KNN = 98.59, SVM = 99.06, XgBoost = 94.36, Random Forest = 99.06, Decision Tree = 83.09, LDA = 91.07.

Precision: -

Precision is a metric that measures how often a machine learning model correctly predicts the positive class.

We got the precision for different models applied in our machine learning article is followed by: Logistic Regression = 98.12, Naive Bayes = 21.16, KNN = 98.73, SVM = 99.06, XGBoost = 94.84, Random Forest = 99.12, Decision Tree = 73.96, LDA = 91.77.

Recall: -

Recall is a metric that measures how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset.

We got the precision for different models applied in our machine learning article is followed by: Logistic Regression = 98.12, Naive Bayes = 46, KNN = 98.59, SVM = 99.06 XGBoost = 94.36, Random Forest = 99.06, Decision Tree = 83.09, LDA = 91.07.

F1 Score: -

The F1 score or F-measure is described as the harmonic mean of the precision and recall of a classification model. The two metrics contribute equally to the score, ensuring that the F1 metric correctly indicates the reliability of a model.

We got the F1 Score for different models applied in our machine learning article is followed by: Logistic Regression = 98.12, Naive Bayes = 28.99, KNN = 98.61, SVM = 99.06 XGBoost = 94.38, Random Forest = 99.07, Decision Tree = 77.45, LDA = 91.23.

AUC-ROC Score: -

An ROC curve, or receiver operating characteristic curve, is like a graph that shows how well a classification model performs. It helps us see how the model makes decisions at different levels of certainty.

We got the AUC-ROC Score for different models applied in our machine learning article is followed by**:** Logistic Regression = 98.92, Naive Bayes = 50, KNN = 50, SVM = 50 XGBoost = 99.68, Random Forest = 99.66, Decision Tree = 85.71, LDA = 97.23.

Sensitivity: -

Sensitivity measures how well a machine learning model can detect positive instances. In other words, it measures how likely you will get a positive result when you test for something.

We got the Sensitivity for different models applied in our machine learning article is followed by: Logistic Regression = 99, Naive Bayes = 98, KNN = 100, SVM = 100, XGBoost = 99, Random Forest = 98, Decision Tree = 100, LDA = 92.

Specificity: -

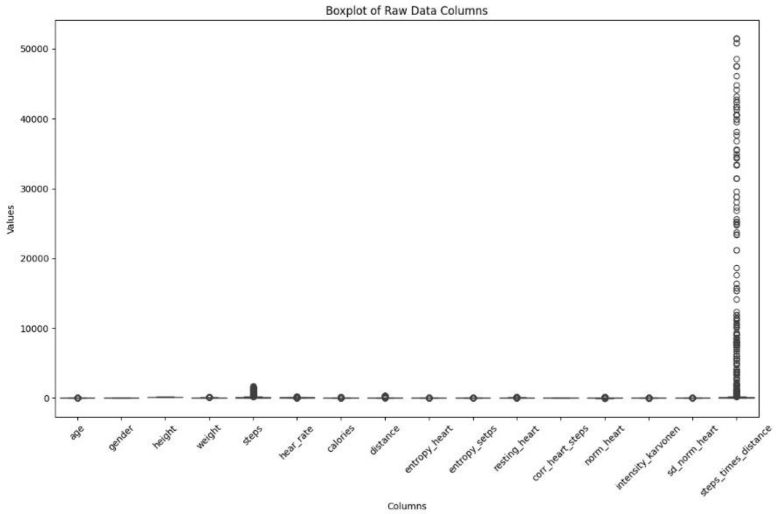
Specificity talks about the number of negative records correctly predicted. So, with Specificity, we can measure how well our model predicts the class that we want to declare as the *negative* class.

We got the Specificity for different models applied in our machine learning article is followed by: Logistic Regression 100, Naive Bayes = 100, KNN = 99, SVM = 98, XGBoost =99.56, Random Forest = 100, Decision Tree = 98.23LDA = 99.3.

Cohen's Kappa: -

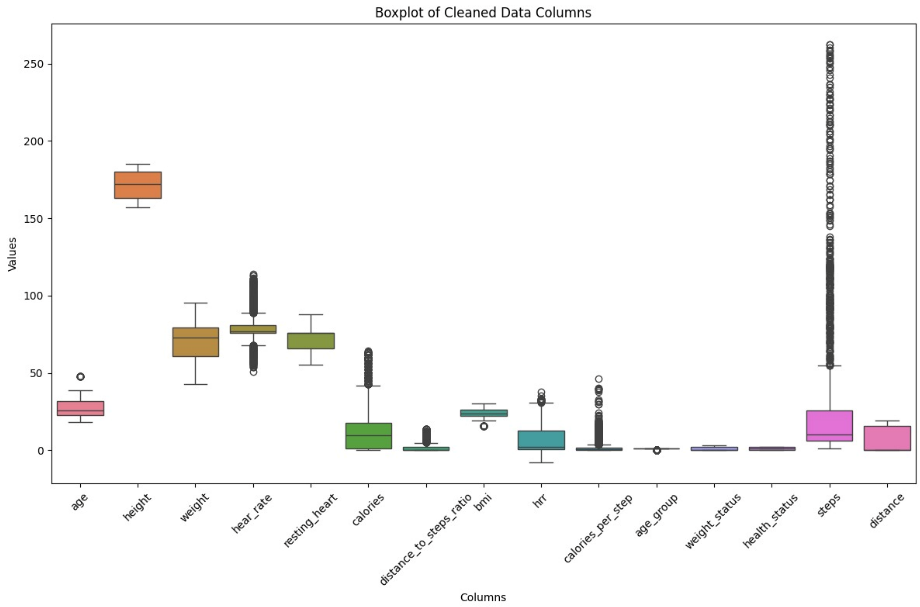
The Kappa Coefficient, commonly referred to as Cohen's Kappa Score, is a statistic used to assess the effectiveness of machine learning classification models.

We got the Cohen's Kappa for different models applied in our machine learning article is followed by: Logistic Regression = 98.87, Naive Bayes = 0, KNN = 97.67, SVM = 98.43, XGBoost = 90.75, Random Forest = 94.44, Decision Tree = 70.53, LDA = 85.27.



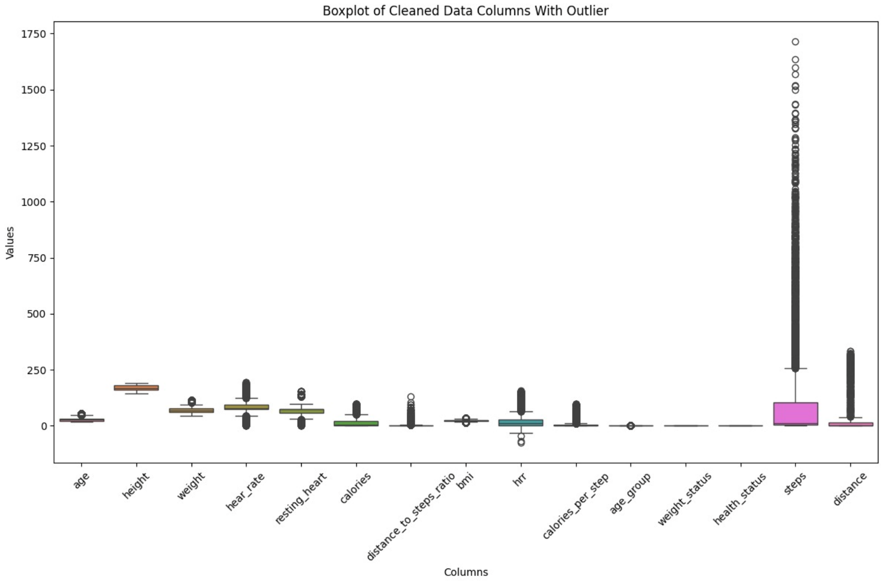
(Fig:4.1 Box Plot of the Raw Data)

- **Steps, Calories, Distance:** These columns have the highest values which represent a wide range of steps taken, calories burned, and distances travelled. Outliers in these lines imply that some of these individuals are much more active than others.
- **Heart Rate:** In the Heart Rate column, there is much less deviation and more values fall into smaller intervals which tells us that those values are likely to be in the normal range.
- **Gender, Height, Weight:** These columns have very small ranges, which is expected as gender is categorical (likely binary), and height and weight values typically don't vary drastically within a dataset.
- **Other Columns:** The remaining columns (entropy, resting heart rate, etc.) have tightly clustered values near zero, suggesting they might be normalized or transformed data. The meaning of these variables would depend on the specific data preprocessing steps taken.



(Fig:4.2 Box Plot Before SMOTE)

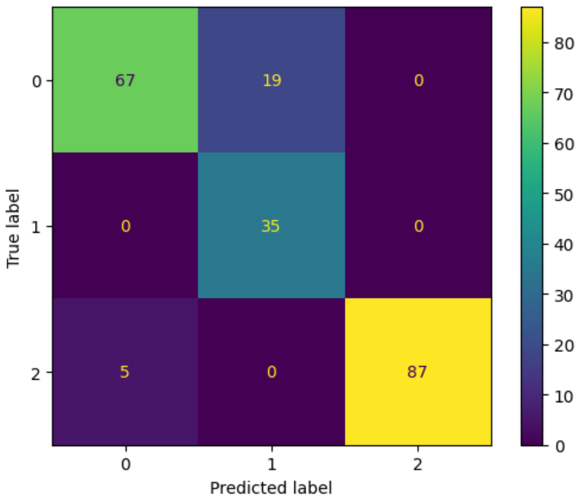
- **Steps and Distance:** These columns have the highest values, representing a variety of numbers regarding steps and distances. This implies that some people might perform a great deal more (outliers).
- **Heart Rate and Resting Heart Rate:** As indicated by the smaller range (almost no outliers), most of the heart rate values tend to fall in a similar and "normal" range.
- **Age, Height, and Weight:** These columns contain numbers in small ranges, which is consistent with these types of columns as the values of age, height, and weight typically do no vary greatly in a dataset.
- **Some of Other Columns:** The rest of the columns (bmi, hrr, calories_per_step, etc.) have ranges and distributions cases which vary. While others like health status are very tightly clustered and calories are much more distributed in the model space The exact interpretations of these variables will vary depending on the preprocessing step as well as the meaning within wearable healthcare data.



(Fig: 4.3 Box Plot After SMOTE)

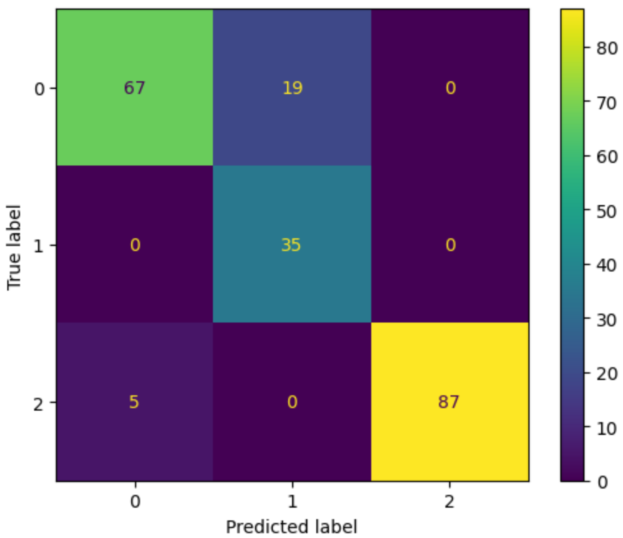
- **Steps and Distance:** These have both the largest medians as well as highest variances, illustrating a wide-ranging number of steps and distance covered. The many outliers in this figure hint that some people are a lot more active than others.
- **Heart Rate and Resting Heart Rate:** These columns have less range with less outliers, which indicates that for most values of heart rate, there are closer to normal values of it.
- **Age, Height, and Weight:** Since these are measurements that do not typically have drastic variation within a dataset, Age, Height, and Weight all have very small ranges.
- **Calories:** This column has a wide range of values due to differences in calorie expenditure between individuals. The fact there are points a long way from the group of points implies some people are burning many calories more than others.
- **Other Columns:** Then we have the rest of the columns such as BMI(HR) HRR etc showing different distribution and ranges. Others clump a little more closely together, showing less variability.

Figure:



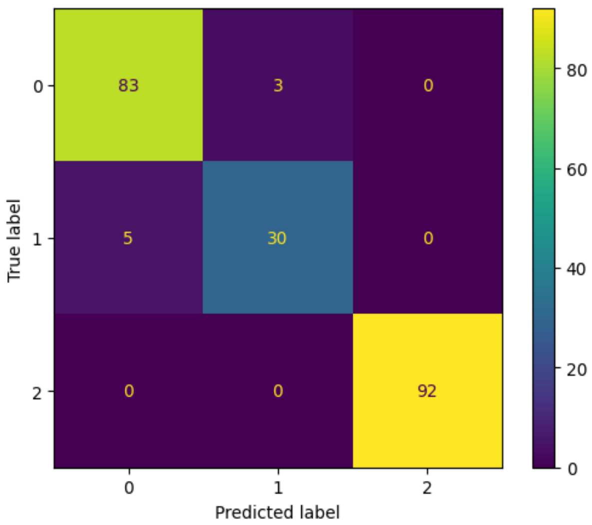
(Fig:4.4 Confusion Matrix – Decision Tree)

The above figure is the confusion matrix of Decision Tree which is used in our project and this has fluctuation false positive and false negative values which may affect our model's accuracy.



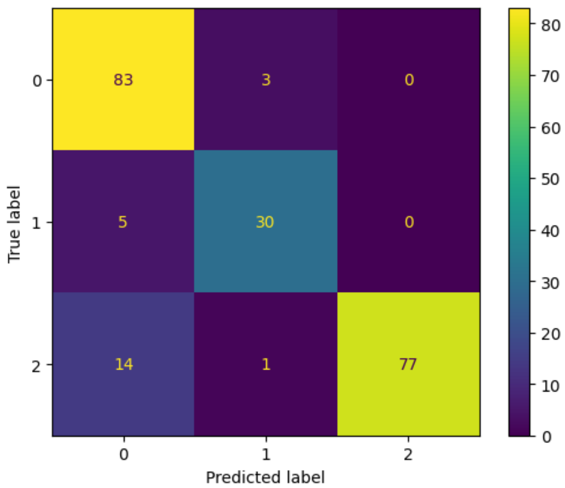
(Fig:4.5 Confusion Matrix – Ensemble Method)

The above figure is the confusion matrix of Ensemble Method (Random Forest and Gradient Boosting) which is used in our project and this has high fluctuation false positive and false negative values which may affect our model's accuracy.



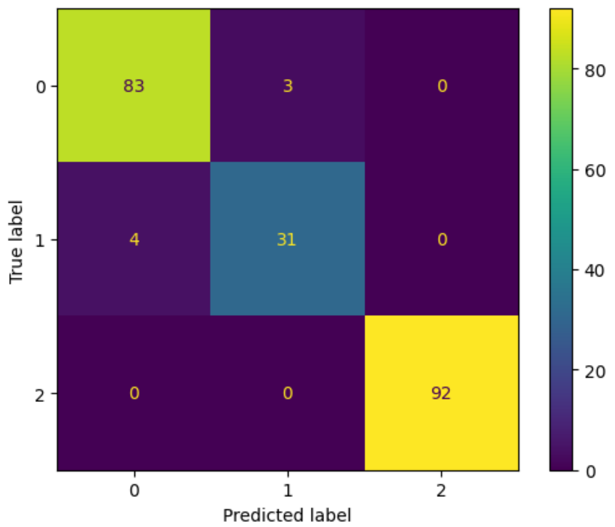
(Fig: 4.6 Confusion Matrix – K Nearest neighbor)

The above figure is the confusion matrix of K Nearest Neighbor which has a stable relation between their false positive and false negative values which can give a very stable result as compared to others.



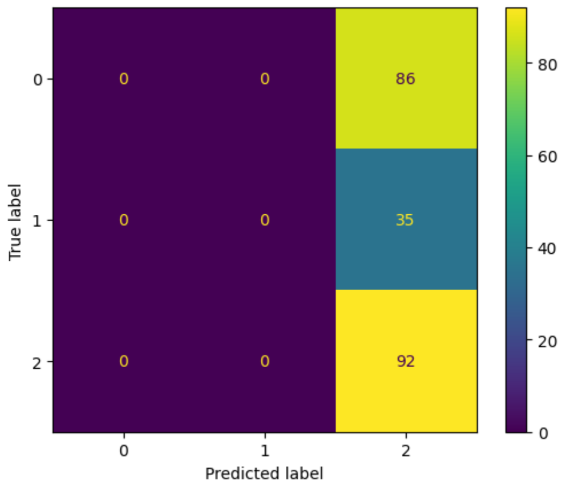
(Fig: 4.7 Confusion Matrix – LDA)

The above figure is the confusion matrix of LDA (Linear Discriminant Analysis) which has less fluctuation false positive and false negative values which can give a very stable result as compared to others.



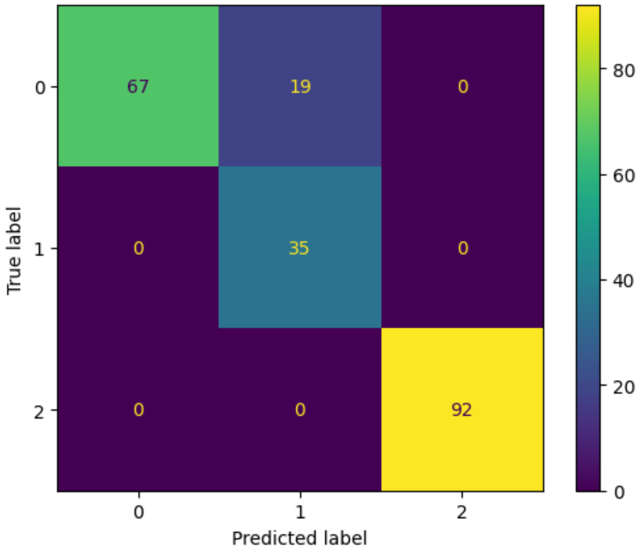
(Fig: 4.8 Confusion Matrix – Logistic Regression)

The above figure is the confusion matrix of Logistic Regression which has a very high relation between their false positive and false negative values which can give a very stable result as compared to others.



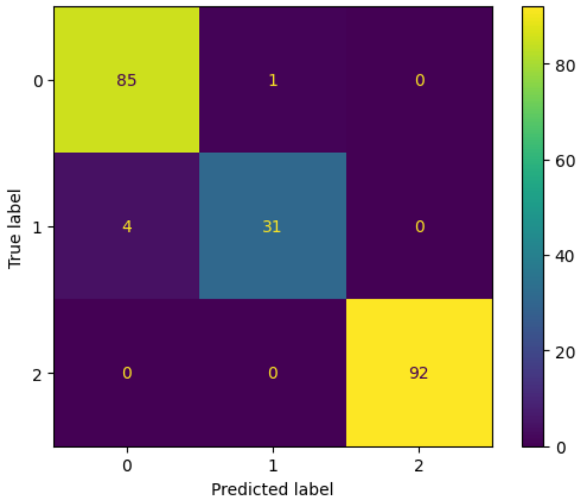
(Fig: 4.9 Confusion Matrix – Naïve Bayes)

This above figure represents the confusion matrix for Naïve Bayes which has very huge difference between the false positive and false negative and can say that no correlation between them which may affect very heavily on our model so, we should avoid using this as a component in our ML model.



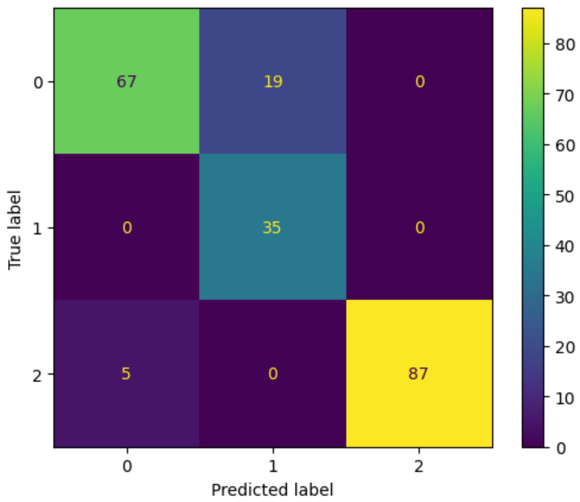
(Fig:4.10 Confusion Matrix – Random Forest)

The above figure is the Confusion Matrix of Random Forest which is used in our project and this has a high fluctuation between false positive and false negative values which may affect our model's accuracy. So, we are avoiding the use of it in out Machine Learning Model.



(Fig:4.11 Confusion Matrix – Support Vector Classifier)

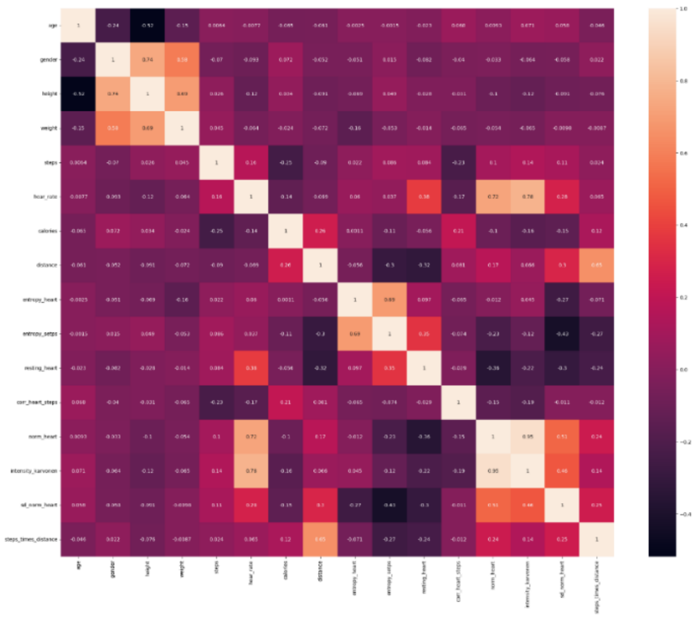
The above figure is the confusion matrix of Support Vector Classifier which has a very high relation between their false positive and false negative values which can give a very stable result as compared to others. So, we got the recommendation of using it strongly for its better accuracy.



(Fig: 4.12 Confusion Matrix - XGBoost)

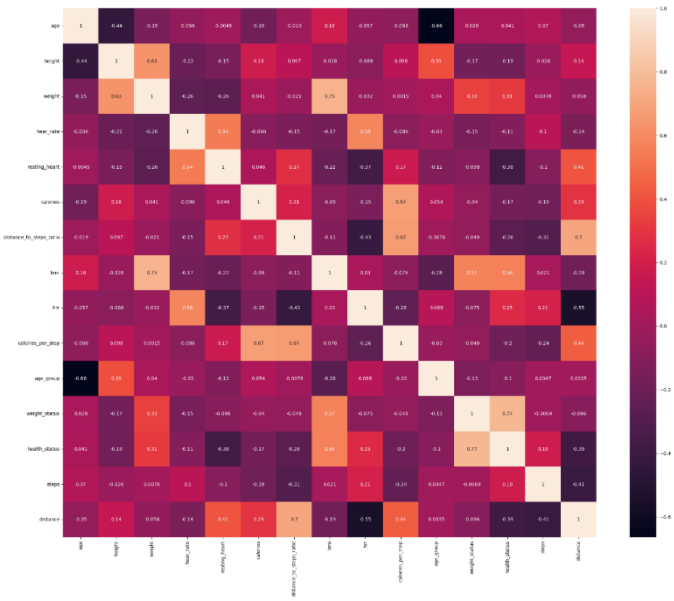
The above figure is the confusion matrix of XgBoost which is used in our project and this has fluctuation false positive and false negative values which may affect our model's accuracy.

Heatmap



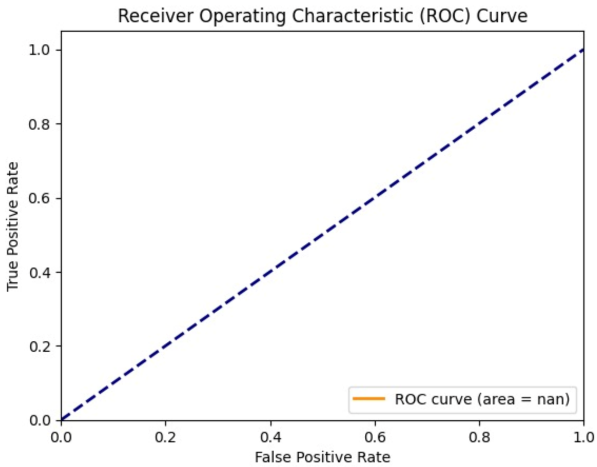
(Fig:4.13 Before Data Cleaning)

This above figure is the Heatmap of correlation between our attributes before cleaning the data which on using to train the model would not give a good result in our model.



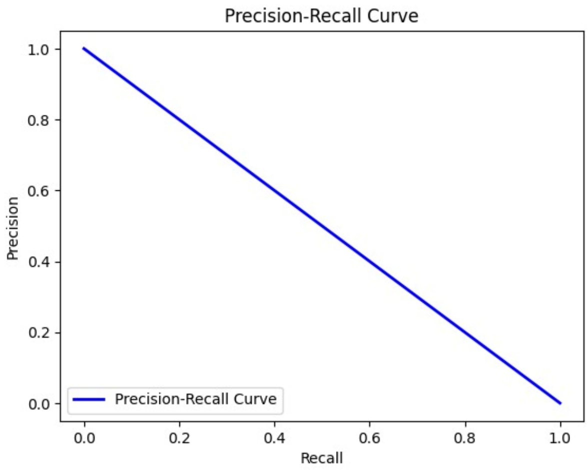
(Fig: 4.14 After Data Cleaning)

This above figure is the Heatmap of correlation between our attributes before cleaning the data which on using to train the model would give a good result in our model as this is showing all the attributes have very good correlation.



(Fig: 4.15 ROC Curve)

The ROC curve in the image shown here means that the model you evaluate is a bad model in comparison to even random chance, this can be because: The model could be a random classifier, there may be an error in calculation, or There is a too imbalanced class. More investigation is definitely needed here to actually find out what is causing this and handle it so that really it can be used for model evaluation.



(Fig: 4.16 Precision-Recall Curve)

If the classes are imbalanced, a precision-recall curve might give a more informative picture than the ROC curve. The precision-recall curve is a plot of precision (TP/P) versus recall (TPR) at different thresholds. This is to measure how well a model performs when the positive class is not only as small as 5% of the entire target set.

CHAPTER: 5

CONCLUSION & FUTURE SCOPE

Future Scope:

Implementation in a New Smartwatch:

A new wristwatch – one with a built-in machine learning model – could provide clients with seamless integrated health monitoring. It may have features such as advanced sensors to enable better data collection and more computing power to run machine learning algorithms on the fly right from their wrist in future smartwatches. The model is built right into the wristwatch and requires no additional devices or connectivity to provide users with real-time health information as well as personalized suggestions.

Software Implementation in Application:

A specialized app could be developed that uses machine learning to increase the number of people who can use it beyond individuals who only own a smartwatch. Any person could access personal medical recommendations using any gadget if they synchronize the device data stored in the fitness tracker with the mentioned software compatible with many smartphones and tablets. The app could also include other features such as tracking progress on health goals, reminding users to take prescriptions or exercise, connecting them with doctors for additional guidance and encouragement. By getting a template on how conventional routines can easily incorporate wellness checks and informed decisions about their health through an intuitive app, people can live safer lives.

Conclusion:

In conclusion, our research on the use of wearable devices in healthcare has revealed a possible path to personal well-being. This paper shows how advanced computational methods, such as random forest analysis or nearest neighbour analysis, together with support vector machines (SVM) or logistic regression. and their effectiveness is discussed using examples based on this paper's data collected through devices such as smart bracelets, or smart watches.

We looked at a large dataset containing 6,400 records from Fitbits and Apple Watches to show how good our machine learning system is at predicting health problems with

98.12 percent accuracy. This accuracy underscores the relevance of wearable technology in analysing patient health information; moreover, it illustrates that if we use timely measures to guide our decisions after making correct predictions, we will be able to achieve this range of accuracy.

Key to the strengths of our approach is that we take into account a range of factors such as age, body mass index (BMI) and daily habits to holistically assess someone's state of well-being or personal health. By combining this breadth of data together, we can represent the full spectrum of health that an individual falls on and suggest ways that are specific to their needs to improve.

The most important thing we aim at is to employ machine learning in order to popularize bespoke wellness schemes and enhance health literacy. In addition, we aim at changing personal behaviour through offering personal advice on mental health and important insights regarding individual health, thereby leading to healthy life practices' proactivity that will eventually result in longer life and higher life quality. Undergoing further research and development around the world is the future and we hope to see where wearables continue to be an integral part of human holistic health.

REFERENCES

[1] Majumder, S., Mondal, T., & Deen, M. J. (2017). Wearable sensors for remote health monitoring. Sensors, 17(1), 130.

[2] Wan, J., AAH Al-awlaqi, M., Li, M., O'Grady, M., Gu, X., Wang, J., & Cao, N. (2018). Wearable IoT enabled real-time health monitoring system. *EURASIP Journal on Wireless Communications and Networking*, *2018*(1), 1-10.

[3] Al-Khafajiy, M., Baker, T., Chalmers, C., Asim, M., Kolivand, H., Fahim, M., & Waraich, A. (2019). Remote health monitoring of elderly through wearable sensors. *Multimedia Tools and Applications*, *78*(17), 24681-24706.

[4] Prieto-Avalos, G., Cruz-Ramos, N. A., Alor-Hernández, G., Sánchez-Cervantes, J. L., Rodríguez-Mazahua, L., & Guarneros-Nolasco, L. R. (2022). Wearable devices for physical monitoring of heart: a review. *Biosensors*, *12*(5), 292.

[5] Kakria, P., Tripathi, N. K., & Kitipawang, P. (2015). A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *International journal of telemedicine and applications*, *2015*, 8-8.

[6] Pantelopoulos, A., & Bourbakis, N. G. (2010). Prognosis—a wearable health-monitoring system for people at risk: Methodology and modeling. *IEEE Transactions on Information Technology in Biomedicine*, *14*(3), 613-621.

[7] Ahmed, Z. U., Mortuza, M. G., Uddin, M. J., Kabir, M. H., Mahiuddin, M., & Hoque, M. J. (2018, December). Internet of Things based patient health monitoring system using wearable biomedical device. In *2018 international conference on innovation in engineering and technology (ICIET)* (pp. 1-5). IEEE.

[8] Evangeline, C. S., & Lenin, A. (2019). Human health monitoring using wearable sensor. *Sensor Review*, *39*(3), 364-376.

[9] Taştan, M. (2018). IoT based wearable smart health monitoring system. *Celal Bayar University Journal of Science*, *14*(3), 343-350.

[10] Sumathy, B., Kavimullai, S., Shushmithaa, S., & Anusha, S. S. (2021). Wearable non-invasive health monitoring device for elderly using IoT. In *IOP Conference*

Series: Materials Science and Engineering (Vol. 1012, No. 1, p. 012011). IOP Publishing.

[11] Vijayan, V., Connolly, J. P., Condell, J., McKelvey, N., & Gardiner, P. (2021). Review of wearable devices and data collection considerations for connected health. *Sensors*, *21*(16), 5589.

[12] Vijayan, V., Connolly, J. P., Condell, J., McKelvey, N., & Gardiner, P. (2021). Review of wearable devices and data collection considerations for connected health. *Sensors*, *21*(16), 5589.

[13] Zhou, S., Ogihara, A., Nishimura, S., & Jin, Q. (2018). Analyzing the changes of health condition and social capital of elderly people using wearable devices. *Health information science and systems*, *6*, 1-10.

[14] Asthana, S., Strong, R., & Megahed, A. HealthAdvisor: Recommendation System for Wearable Technologies enabling Proactive Health Monitoring. arXiv 2016. *arXiv preprint arXiv:1612.00800*.

[15] Arshad, A., Khan, S., Alam, A. Z., Ahmad, F. I., & Tasnim, R. (2014). A study on health monitoring system: recent advancements. *IIUM Engineering Journal*, *15*(2).

[16] Sharma, A., Badea, M., Tiwari, S., & Marty, J. L. (2021). Wearable biosensors: an alternative and practical approach in healthcare and disease monitoring. *Molecules*, *26*(3), 748.

[17] González-Valenzuela, S., Chen, M., & Leung, V. C. (2011). Mobility support for health monitoring at home using wearable sensors. *IEEE Transactions on* Information *Technology in Biomedicine*, *15*(4), 539-549.

[18] Esposito, M., Minutolo, A., Megna, R., Forastiere, M., Magliulo, M., & De Pietro, G. (2018). A smart mobile, self-configuring, context-aware architecture for personal health monitoring. *Engineering Applications of Artificial Intelligence*, *67*, 136-156.

[19] Jiang, W., Majumder, S., Kumar, S., Subramaniam, S., Li, X., Khedri, R., ... & Deen, M. J. (2021). A wearable tele-health system towards monitoring COVID-19 and chronic diseases. *IEEE Reviews in Biomedical Engineering*, *15*, 61-84.

[20] Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S. M., ... & Snyder, M. P. (2017). Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS biology*, *15*(1), e2001402.

[21] Teixeira, E., Fonseca, H., Diniz-Sousa, F., Veras, L., Boppre, G., Oliveira, J., ... & Marques-Aleixo, I. (2021). Wearable devices for physical activity and healthcare monitoring in elderly people: A critical review. *Geriatrics*, *6*(2),

Direct_from_him.docx

ORIGINALITY REPORT

12%	11%	6%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.evidentlyai.com Internet Source	2%
2	Submitted to University of Bradford Student Paper	1%
3	www.analyticsvidhya.com Internet Source	1%
4	dlibrary.univ-boumerdes.dz:8080 Internet Source	1%
5	isset.shiksha.com Internet Source	1%
6	arxiv.org Internet Source	1%
7	Submitted to University of Technology, Sydney Student Paper	1%
8	Submitted to Universiti Teknologi Malaysia Student Paper	<1%
9	www.ijraset.com Internet Source	<1%

10	bookdown.org Internet Source	<1 %
11	pypi.org Internet Source	<1 %
12	uvadoc.uva.es Internet Source	<1 %
13	Submitted to Domain Academy Student Paper	<1 %
14	fastercapital.com Internet Source	<1 %
15	www.jsr.org Internet Source	<1 %
16	www.mdpi.com Internet Source	<1 %
17	Submitted to Coventry University Student Paper	<1 %
18	www.techscience.com Internet Source	<1 %
19	Submitted to University of Hertfordshire Student Paper	<1 %
20	ijisrt.com Internet Source	<1 %
21	"Vision, Sensing and Analytics: Integrative Approaches", Springer Science and Business	<1 %

Media LLC, 2021
Publication

22	0-www-mdpi-com.brum.beds.ac.uk Internet Source	<1 %
23	Submitted to Central Queensland University Student Paper	<1 %
24	vnexplorer.net Internet Source	<1 %
25	Sabyasachi Pramanik, Alex Khang. "chapter 14 Cardiovascular Diseases", IGI Global, 2024 Publication	<1 %
26	link.springer.com Internet Source	<1 %
27	www.iapress.org Internet Source	<1 %
28	doctorpenguin.com Internet Source	<1 %
29	jwcn-eurasipjournals.springeropen.com Internet Source	<1 %
30	libweb.kpfu.ru Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		