PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
MOHAMED BOUDIAF UNIVERSITY - M'SILA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
COMPUTER SCIENCE DEPARTMENT

THESIS

Submitted in partial fulfillment of the requirements for the degree of *Master*

Specialty: Computer Sciences

Option: Artificial intelligence

By: MIHOUBI Hadil

SAOUDI Racha

CALORIE BURN PREDICTION

Publicly defended on:13/06/2023 in front of the jury composed of:

Dr. DEBBI Hichem	MBU of M'sila	President
Dr. HEMMAK Allaoua	MBU of M'sila	Supervisor
Dr. CHALABI Nourelhouda	MBU of M'sila	Examiner

Academic Year: 2022 /2023

ACKNOWLEDGMENT

Dedication

To our dear parents

"أهدي تخرجي هذا إلى روح والدي الذي لم يشاهدني وأنا أتوج على هذه المنصة، فكم كنت أتمنى أن تكون بجانبي في هذه اللحظة الجميلة من حياتي، لكن وعد يا أبي سأرفع رأسك عالياً بكل عزيمة وإصرار، رحمك الله وأنار قبرك يا أغلى سند وأغلى أب رحل عن الدنيا."

سعودي رشا

Acknowledgment

above all, thankful and heartfelt gratitude to god, we thank to express our respect for Mr. HEMMAK Allaoua guidance.

we would like to thank our parents and siblings who have supported us during this academic journey and our entire lives.

and our friends for their invaluable assistance, continuing encouragement. We thank jury members for the honor they have given us by agreeing to evaluate and judge this humble act.

TABLE OF CONTENTS

LIST OF FIGURES

GENERAL INTRODUCTION

General Introduction

Today, technology has become one of the most important areas of human life. It plays a major role to determine the future especially artificial intelligence.

Artificial Intelligence (AI) is one of the most important technologies of our time. AI is a field of computer science that focuses on the creation of intelligence machines that can perform tasks that would typically require human intelligence, such as visual perception, speech recognition. AI systems are designed to learn and improve overtime, using algorithms that enable them to recognize patterns in data, make predictions, and take actions based on their analysis, these systems rely on techniques such as machine learning, deep learning, and natural language processing to simulate human intelligence and perform complex tasks.

Machine Learning (ML) is a subset of AI that involves training algorithms to recognize patterns in data and make predictions based on that data, it is a way of enabling machines to learn and improve over time without being explicitly programmed to do so ML algorithms are used in a wide variety of applications such as image recognition, speech recognition natural language processing, and predictive modeling.

The main idea in this project is all about apply our machine learning models in order to find out how many calories each individual will burn by gathering the right dataset. Pre-processing of the records needs to be done before the statistics feeding operation after that data analysis is completed. Then by using some visualization techniques the data arranged as graphs and plots, we use XGBoost regressor as machine learning model.

Hence, this document is divided into five chapters

- In the first chapter we described briefly the field of the burned calorie and its factors.
- In the second chapter we provide a definition of artificial intelligence and machine learning that’s also its types and algorithms.
- The third chapter consists to present the state of the art and related work to the same subject.
- The fourth chapter we talk about the methodology and design of our project.
- In the last chapter we describe the design and the implementation of our application then we dress some comments, discussions and deductions.

CHAPTER I

CALORIE BURN SYSTEM

Introduction

The human body needs calorie to survive, without energy he would die, people absorb this energy from food and drink.
If people consumed only the number of calorie needed every day, they would probably have healthy lives. Calorie consumption that is too low or too high will eventually lead to health problems. And that's why the human body needs to burn calorie, and this what we are going to explain in this chapter.

1. Understanding calories

The amount of energy in an item of food or drink is measured in calories. When we eat and drink more calorie than we use up. Our bodies store the excess as body fat, if this continues, over time we may put on weight. [1]

A calorie is a measure of energy expenditure and stored energy. The calories referred to in diet (calories eaten) and exercise (calories burned) are kilocalories. [2]A calorie is the amount of energy that is needed to raise 1 gram (g) of water by 1°C. this measurement can be applied to lots of different energy releasing mechanisms outside of the human body, for the human body calorie are a measure of how much energy the body needs to function [3]

As a guide, an average man needs around 2.500kcal (10.500kj) a day to maintain a healthy body weight. For average woman, that figure is around 2.000kcal (8.400kj) a day. [1]

2. Calorie and energy balance

Our bodies need energy to keep us alive and our organs functioning normally. When we eat and drink, we put energy into our bodies, our bodies use up that energy through everyday movement, which includes everything from breathing to running. to maintain a stable weight, the energy we put into our bodies must be the same as the energy we use through normal bodily functions and physical activity.

An important part of a healthy diet is balancing the energy we put into our bodies with the energy we use. For example, the more physical activity we do, the more energy we use. if we consume too much energy on 1 day, we just try to take in less energy on the following days. [1]

Energy is another word for "calories". Our energy balance is the balance of calorie consumed through eating and drinking compared to calorie burned through physical activity. What we eat and drink is ENERGY IN. What we burn through physical activity is ENERGY OUT.

We burn a certain number of calorie just by breathing air and digesting food. We also burn a certain number of calorie (ENERGY OUT) through our daily routine.

An important part of maintaining energy balance is the amount of ENERGY OUT (physical activity) that we do. People who are more physically active burn more calorie than those who are not as physically active.

The same amount of ENERGY IN (calories consumed) and ENERGY OUT
(calories burned) over time = weight stays the same
More IN than OUT over time = weight gain
More OUT than IN over time = weight loss [4]
This figure explains how to get energy balance by the calories consumed and calories burned (as shown in figure1.1)

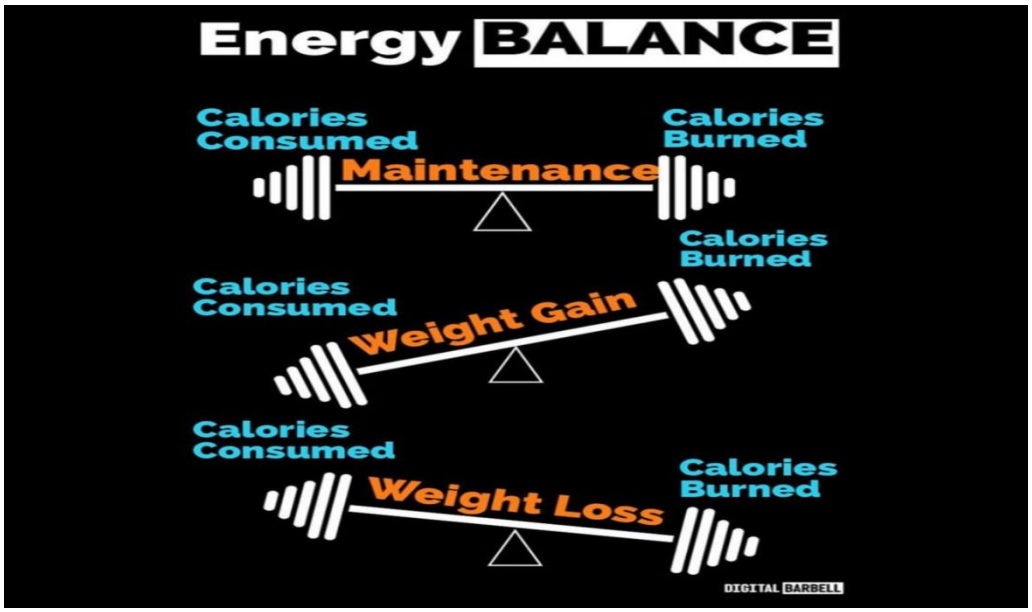

Figure 1.1 Energy balance *[5]*

3. Factors affecting calorie burn

Many factors how many calorie a person burns each day, some of the factors that influence daily calorie burn are not is a person's control while others can be changed these factors include: [3]

- Age: the older person is the fewer calorie burned per day.
- Sex: men burn more calorie than women.
- Amount of daily activity: those who move more, burn more calories.
- Body composition: those with more muscle burn more calorie than those who have less muscle.
- Body size: larger people burn more calorie than smaller people, even at rest.
- The mogenesis: this the amount of energy the body uses to break down food.
- Pregnancy: pregnant woman burns more calorie than non-pregnant women.
- Breast-feeding: women who are breast-feeding also burn extra calories.

4. Basal Metabolic Rate

It is essential for the body to use energy continuously for maintenance and to function correctly. Although the body burns some of its calorie through physical activity, basal metabolism accounts for around 80% of energy expenditure.

Basal metabolism refers to all of the processes that keep the body alive, functioning, and healthy, these include:

• Breathing.
• Blood circulation.
• Cellular growth and repair.
• Brain and nerve function.
• Temperature control.

These metabolic processes require a constant supply of energy, so the body is always burning calories, even when a person is resting or sleeping.
BMR is the number of calorie that the body burns every 24 hours due to basal metabolism.
Essentially, it is how many calories a person would use in day if they just rested and did no physical activity.

Many different factors affect BMR including age, body, sex, and genetics, for this reason, everyone’s BMR is different. [3]

4.1 Calculating daily calorie burn

Being able to work out how many calories are burned each day is essential to any person looking to maintain, lose or gain weight. Knowing what factors contribute to calorie burning can help a person alter their diet or exercise program to accommodate the goal. An accepted method to calculate how many calories a person burns each day is the harris-bendict formula.

Originally developed in the early 20th century, it was revamped in 1984 and again in 1990 to help improve its accuracy. The Harris-benedict formula is a relatively simple process in which a person multiplies their basal metabolic (BMR) by their average daily activity level.

BMR is the number of calorie a person burns by simply existing. BMR varies based on age, sex, size, and genetics, to calculate BMR, a person uses inches for height, pounds for weight, and years for age in the following formulas:

• For men: 66+(6.2*weight) +(12.7*height) -(6.76*age). (1)

• For women: 655.1+(4.35*weight) +(4.7*height) -(4.7*age). (2)

The result of the BMR calculation are then used to multiply against the average daily activity of the person. Points are awarded based on how active a person is.

Points for activity levels are as follows:

• 1.2 points for a person who does little to no exercise.
• 1.37 points for a slightly active person who does light exercise 1-3 days a week.
• 1.55 points for a moderately active person who performs moderate exercise 3-5 days a week.
• 1.725 points for a very active person who exercise hard 6-7 days a week.
• 1.9 points for an extra active person who either has a physically demanding job or has a particularly challenging exercise routine.

When the BMR is calculated and the activities points are determined, the two scores are multiplied. The total is the number of calorie burned on an average day. For example, to calculate how many calorie a 37-year-old, foot-tall, and 170-pound man who is moderately active burns, the formula would look like:

(66+(6.2*170) +(12.7*72) -(6.76*37)) *1.55=2.766 calorie/ day. (1)

This figure activity shows that a man of this age, height, weight, and activity level can consume 2.766 calorie and maintain his current weight, he could increase or decrease weight by consuming more or less than this amount over the course of serval days [3]

This figure presents the BMR formulas for men and women (as shown in figure1.2).

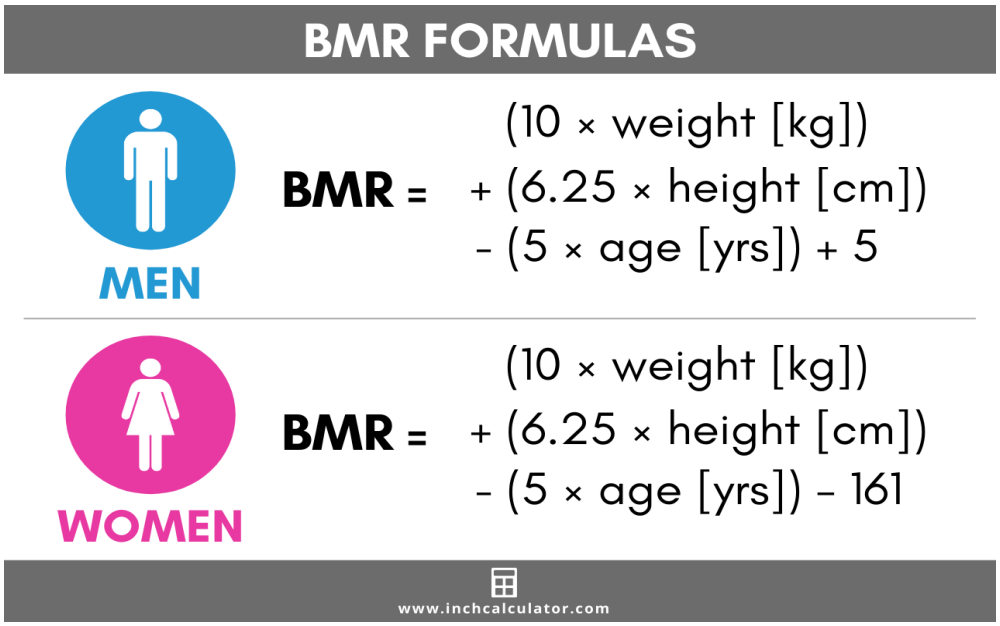

Figure 1.2 BMR formulas *[6]*

5. Burning calories

The amount of calorie that people uses by doing a certain physical activity varies, depending on a range of factors, including size and age.

The more vigorously on activity, the more calorie use. For example, fast walking will burn more calorie than walking at a moderate pace. If the person wants gain weight, it could mean regularly eat and drink more calorie than before.

To lose weight, need to use more energy than consume it, and continue this over a period of time. The best approach is to combine diet changes with increased physical activity. [1]

5.1 How to measure calorie burned

If we looking for a means to acutely measure the number of calorie we just burned from a workout, unfortunately, it’s challenging to get an exact number. "Generally, finding an accurate caloric burn

from a specific workout is incredibly tough," says Miriam Fried, Founder and Head Trainer at MF Strong.

"The caloric metrics you might find on a machine at the gym are just estimations based on average body weight, so it won't be near accurate for most users. Even most studies on popular wearables currently on the market showed none were able to accurately track your caloric expenditure." [7]

That doesn't mean it's not worth tracking, though. "It's hard to measure the exact number of calories burned in a given workout, but many resources are available to provide close estimations," Froerer says. These include:

- Heart Rate Monitor

If you want to go a bit older school, invest in a heart rate monitor. "Your heart rate is one of the best physiological factors to determine how much effort it takes in order to perform a certain activity," Froerer says [8]"Heart rate sensors are able to pick up this biological feedback and can then be used to calculate caloric expenditure." Of course, now most fitness trackers have built-in heart rate monitors—but you can get a separate heart rate monitor to use while exercising if you wish.

- MET Value Charts

If you really want to level up and do a deep dive into calorie counts, look into MET, or metabolic equivalent charts. "These show a ratio of your working metabolic rate in relation to your resting metabolic rate," Froerer explains. [9]."Metabolic rate is the rate of energy expended per unit of time and a good way to estimate how many calories are burned during a specific physical activity. » One MET is the energy you spend sitting at rest, or your resting or basal metabolic rate.

5.2 Calories burned in physical activity

Physical activity burns calorie beyond the basal metabolic rate, the muscles use both readily available and stored energy sources in the body. The exercise calorie burned during cardiovascular activities such as walking, running, swimming, and cycling depend on the intensity of the exercise, the body weight, and the amount of time spending at exercise. Moderate-intensity exercises such as brisk walking burn fewer calorie per minute than more vigorous-intensity exercises such as running. [2]

5.3 The calories burned in typical workout

Calorie counts for exercise and activity will vary from person, age, sex, body type, and size influence how many calories an individual will burn doing a physical activity.

In general, more intense or strenuous activity will burn more calorie than lighter effort exercise. It's difficult to list out how many calorie each and every person would burn in single workout, froerer says the numbers below are the average calorie burned by a 150- pound woman in a 30- or 60- minute workout [10]

- 60 minutes HIIT: 800 calories.
- 60 minutes of running: 600 calories (10:00/ miles pace).
- 60 minutes boxing:560 calories.
- 60 minutes weight training: 450 calories.
- 60 minutes Pilates: 300 calories.
- 60 minutes yoga: 240 calories.
- 60 minutes of barre: 220 calories.
- 30 minutes HIIT: 400 calories.
- 30 minutes of running: 300 calories (10:00/mile pace).
- 30 minutes boxing: 280 calories.
- 30 minutes weight training: 220 calories.
- 30 minutes Pilates: 150 calories.
- 30 minutes yoga: 120 calories.
- 30 minutes of barre: 110 calories.

5.4 Workout that burn the most calories

HIIT workout and aerobic exercise like running, rowing, and cycling and strength training, either "the more focus on total body exercise, the more muscle recruitment, resulting in more calorie burned" froerer says. "HIIT workout are a quick and efficient way to burn calorie and boost metabolism while continuing to burn calorie through the day ". [11]

6. Activity tracker and fitness apps

While not exact, apple watch or Fitbit give a pretty good idea of how workouts are going. "the software behind these devices combines many factors such as age, sex, weight, workout duration, and heart rate to provide estimate, "foerer says." Because each wearable or app will give you a different

estimated calorie burn, I would recommend finding the one you like the most and sticking with the numbers it gives to you". [10]

7. Benefits of burning calories

Health is much more important than wealth, once the person decide to work on his health, it will achieve wonders, burning calorie will not only keep it healthy, but also ensures that the body remain active throughout the day, burning calorie eventually means losing extra weight, and it has various health advantages, following the benefits of burning those extra calories that the body didn't need.

- It ensures the blood vessels stay healthier and lower the risk of cardiovascular diseases.
- It protects the body from the illnesses like diabetes and helps to control the condition if it already have.
- It improves the mood and lowers the stress and feeling of anxiety. [12]

7.1 Ways to burn calories

Below are the ways to burn calorie and take care of your overall health-

• Regular exercise- One of the most effective ways to burn calorie is doing regular exercise. When you exercise diligently, your body starts to burn calorie to fuel your activity. The effect of exercise lasts for a long time, and that's how you keep burning calorie even when you stop exercising. Experts suggest, if you wish to have a prolonged calorie-burning effect, exercise for more extended periods.

• Strength training- As you age, opting for strength training becomes more critical as your metabolism tends to slow down. Doing strength training for a couple of days a week will ensure

the calorie-burning deposited in the thighs, abdomen and other areas.

• Consume caffeinated teas- Consuming hot caffeinated green and black teas ensures stimulation in your body. Stimulants tend to calorie makes you feel energetic as well. It also results in metabolic changes and that lead to burning more calories. Having a zero-calorie cup of tea will also reduce the number of calorie you take in a sweetened beverage and reduce your cravings.

• Eat in proportions- Eating smaller meals help you digest all the nutrients nicely and keeps your digestion healthy. Calories make human digestion active; hence, smaller meals you consume throughout the day help burn calorie faster. Experts suggest eating large meals does no good to our health compared to smaller meals as it encourages you to burn your calorie

even better.

• Don't skip breakfast- Experts suggest that skipping breakfast is one of the unhealthiest practices starting your day. You tend to eat more throughout the day if you don't eat your breakfast and consume more calorie end of the day.

• Fidget as much as possible- If you can't spare yourself time to work out, ensure that you move your body as much as you can. Even restless walking in routine can help you burn calories. Fidgeting after meals would ensure you feel light and active throughout the day. [12]

Conclusion

In this chapter, we have provided a comprehensive overview of calories, what they are, how they are burned, how to calculate them, and the factors that influence them. Additionally, we have highlighted the numerous benefits of burning calorie and have presented various ways to do so. By understanding the science behind calorie and how to burn them, we can make informed decisions about our diet and exercise habits, ultimately leading to a healthier and happier lifestyle.

CHAPTER II

MACHINE LEARNING

Introduction:

When it comes to AI, it's impossible to ignore the importance of machine learning. In this chapter, we will provide a comprehensive overview of intelligence artificial, machine learning, beginning with definitions and concepts that are essential to understanding this field. We will then delve into the various algorithms that are used in machine learning, exploring their different types.

INTELLIGENCE ARTIFICIAL

Artificial intelligence is an increasingly important technology that enable computers to simulate human intelligence-based tasks. With a robust foundation of software. It works by building and training algorithms using programming languages like python, R, and more. Once you have an algorithm, the next step is to in gest training data in the model, analyze it, and find patterns/ abnormalities to forecast future states, there are many learning materials on artificial intelligence and related technologies. However, not all of those materials credibly explain AI concepts and applications. [13]

1. MACHINE LEARNING

2.1. Definition

Machine Learning is a subset of Artificial Intelligence that uses statistical learning algorithms to build systems that have the ability to automatically learn and improve from experiences without being explicitly programmed. Most of us use machine learning in our day-to-day life when we use services like recommendation systems on Netflix, YouTube, Spotify; search engines like google and yahoo; voice assistants like google home and amazon Alexa. In Machine Learning we train the algorithm by providing it with a lot of data and allowing it to learn more about the processed information. [14]

Or we can say Machine learning is to predict the future based on the past. For example, we can predict how user Omar will choose a movie he hasn't seen yet, based on his reviews of films he has seen (the passive) So, predicting the future based on the past (a set of examples) is at the heart of all the algorithms of machine learning. [15]

This figure represents the general scheme of machine learning process (as shown in figure2.1).

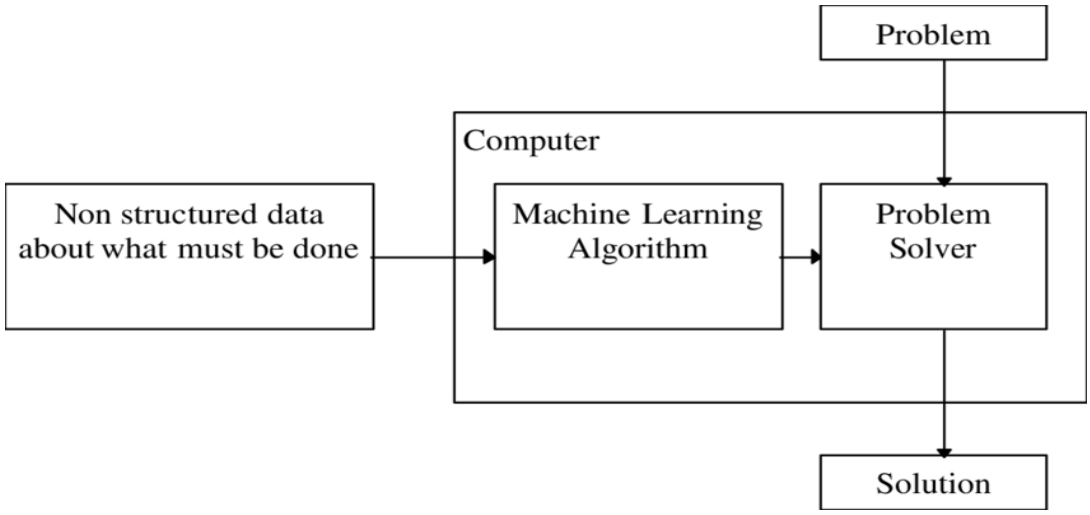

Figure 2.1 : Machine learning general scheme *[16]*

2. Types of machine Learning

There are four types of machine learning (as shown in figure2.2):

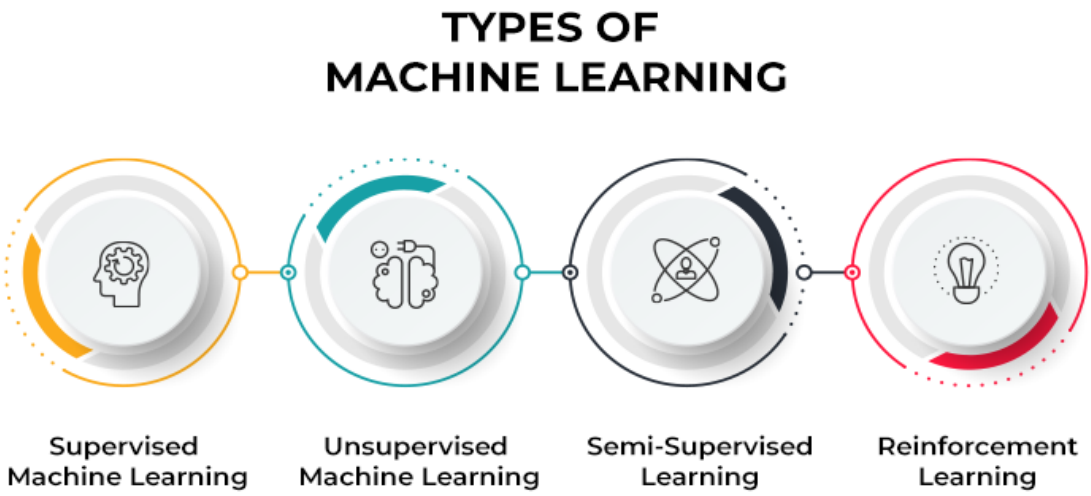

Figure 2.2 : Machine Learning types *[17]*

3.1. Supervised learning

In supervised learning, the machine is taught by example. The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs. While the operator

knows the correct answers to the problem, the algorithm identifies patterns in data, learns from observations and makes predictions. The algorithm makes predictions and is corrected by the operator – and this process continues until the algorithm achieves a high level of accuracy/performance. [18]

This figure explains how supervised learning process work (as shown in figure2.3).

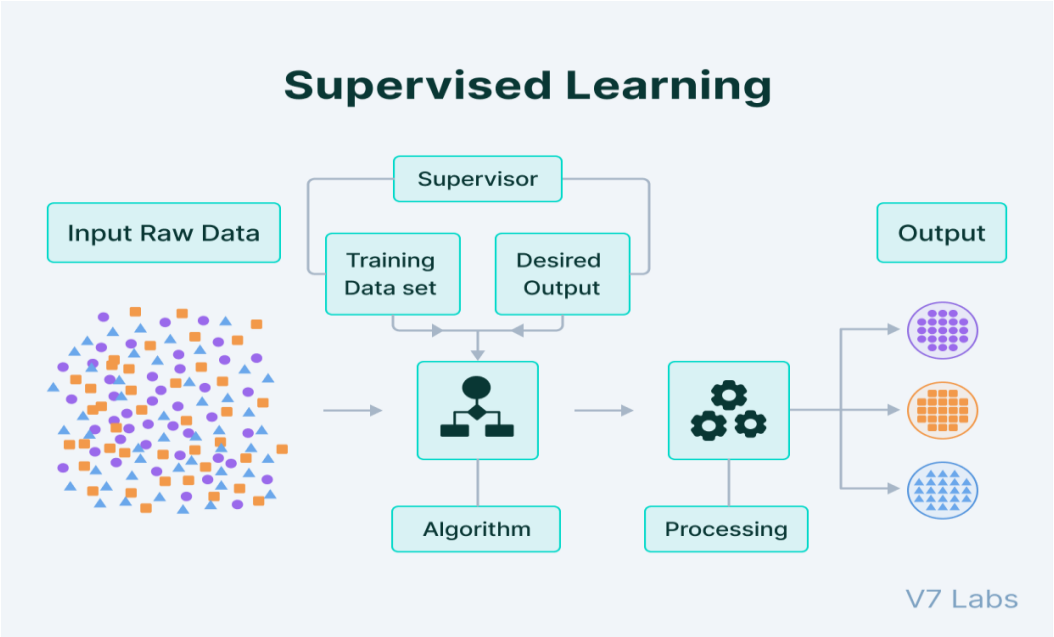

Figure 2.3 : Supervised Learning *[19]*

3.1.1. Regression

Regression is a method for understanding the relationship between independent variables or features and a dependent variable or outcome. Outcomes can then be predicted once the relationship between independent and dependent variables has been estimated. Regression is a field of study in statistics which forms a key part of forecast models in machine learning. It's used as an approach to predict continuous outcomes in predictive modelling, so has utility in forecasting and predicting outcomes from data. Machine learning regression generally involves plotting a line of best fit through the data points. [20]The distance between each point and the line is minimized to achieve the best fit line. In this case, a dataset is modeled using the following equation: $\boldsymbol{Y=M*X+B}$

It is possible to train a regression model with multiple pairs of data, such as x, y. To do this, you need to define a position, as well as the slope of the line, with a minimal distance from all known data

points. This is the line that best approximates the observations in the data, and can help make predictions for new unseen data. [21]

Some of the most common regression techniques in machine learning can be grouped into the following types of regression analysis:

- Simple Linear Regression
- Multiple linear regression
- Logistic regression [20]

3.1.2. Classification

Classification algorithms can explain or predict a class value. It is an essential component for many AI applications, but it is especially useful for ecommerce applications. For example, classification algorithms can help predict if a customer will purchase a product, or not. The two classes in this case are “yes” and “no”. Classification algorithms are not limited to two classes and can be used to classify items into a large number of categories. Logistic regression is considered the simplest and most basic classification algorithm. A logistic regression algorithm can take more than one input, and use the data to estimate the probability of an event occurring. [21]

There are many different types of classification tasks that you may encounter in machine learning and specialized approaches to modeling that may be used for each.

There a different types of classification predictive modeling in machine learning:

- Classification predictive modeling involves assigning a class label to input examples.
- Binary classification refers to predicting one of two classes and multi-class classification involves predicting one of more than two classes.
- Multi-label classification involves predicting one or more classes for each example and imbalanced classification refers to classification tasks where the distribution of examples across the classes is not equal. [22]

3.1.3. Forecasting

Forecasting is the process of making predictions about the future based on the past and present data, and is commonly used to analyses trends.

Advantages

- Supervised learning allows collecting data and produces data output from previous experiences.
- Helps to optimize performance criteria with the help of experience.
- Supervised machine learning helps to solve various types of real-world computation problems.
- It performs classification and regression tasks.
- It allows estimating or mapping the result to a new sample.
- We have complete control over choosing the number of classes we want in the training data. [23]

Disadvantages

- Classifying big data can be challenging.
- Training for supervised learning needs a lot of computation time. So, it requires a lot of time.
- Supervised learning cannot handle all complex tasks in Machine Learning.
- Computation time is vast for supervised learning.
- It requires a labelled data set.
- It requires a training process. [23]

3.2. Unsupervised learning

Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself. [23] . that’s an example how to work the unsupervised learning at the figure bellow (as shown in figure2.4).

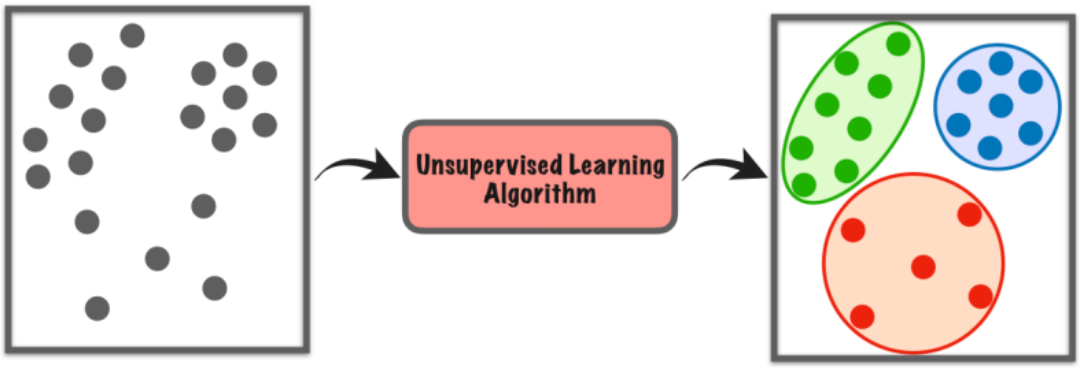

Figure 2.4 : Unsupervised learning *[24]*

Unsupervised learning is classified into two categories of algorithms:

3.2.1. Clustering

A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

3.2.2. Association

An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Advantages of unsupervised learning:

- It does not require a training data to be labelled.
- Dimensionality reduction can be easily accomplished using unsupervised learning.
- Capable of finding previously unknown patterns in data. [23]

Disadvantages of unsupervised learning:

- Difficult to measure accuracy or effectiveness due to lack of predefined answers during training.
- The results often have lesser accuracy.
- The user needs to spend time interpreting and label the classes which follow that classification. [23]

3.3. Semi-supervised learning

Semi-supervised learning is a type of machine learning that falls in between supervised and unsupervised learning. It is a method that uses a small amount of labeled data and a large amount of unlabeled data to train a model. The goal of semi-supervised learning is to learn a function that can accurately predict the output variable based on the input variables, similar to supervised learning. However, unlike supervised learning, the algorithm is trained on a dataset that contains both labeled and unlabeled data. Semi-supervised learning is particularly useful when there is a large amount of unlabeled data available, but it's too expensive or difficult to label all of it5(as shown in figure2.5) . [25]

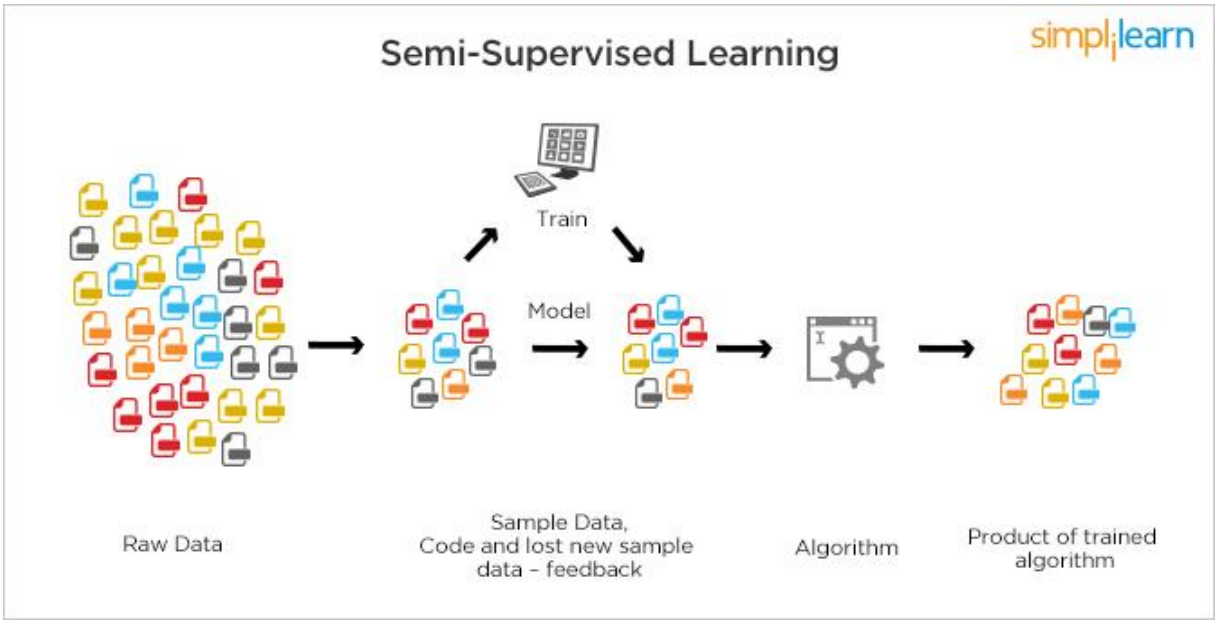

Figure 2.5 : Semi-supervised Learning *[26]*

3.4. Reinforcement learning

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation. Reinforcement learning differs from supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.

Reinforcement learning uses algorithms that learn from outcomes and decide which action to take next. After each action, the algorithm receives feedback that helps it determine whether the choice it made was correct, neutral or incorrect. It is a good technique to use for automated systems that have to make a lot of small decisions without human guidance.

Reinforcement learning is an autonomous, self- teaching system that essentially learns by trial and error. It performs actions with the aim of maximizing rewards, or in other words, it is learning by doing in order to achieve the best outcomes (as shown in figure2.6)

. [27]

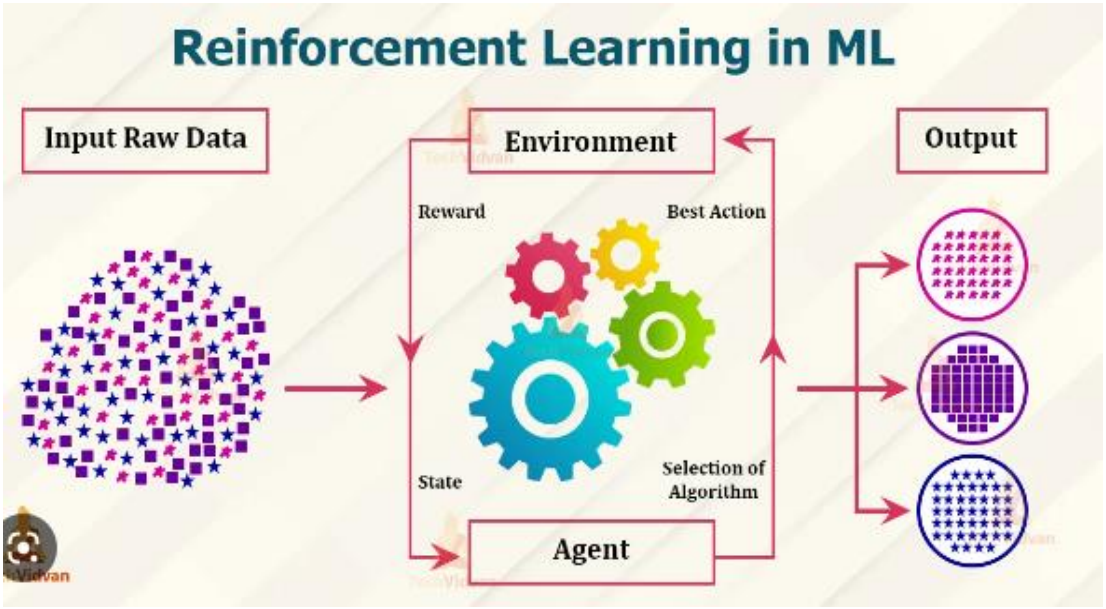

Figure 2.6 : Reinforcement Learning *[28]*

3. Machine Learning algorithms

4.1. Linear regression

Linear Regression is an ML algorithm used for supervised learning. Linear regression performs the task to predict a dependent variable(target) based on the given independent variable(s). So, this regression technique finds out a linear relationship between a dependent variable and the other given independent variables. Hence, the Name of This algorithm is Linear Regression (as shown in figure2.7).

.

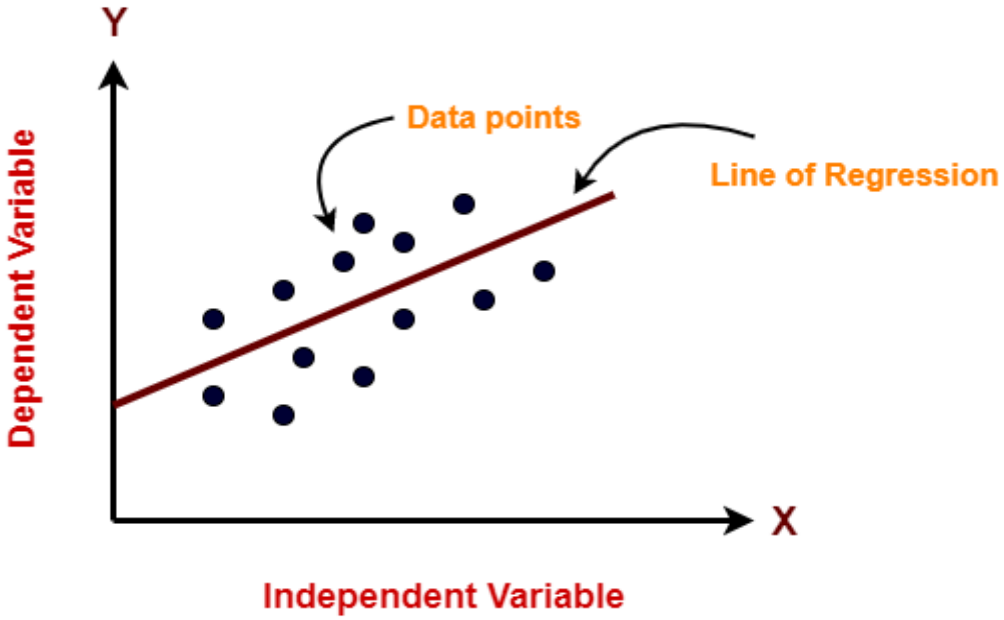

Figure 2.7 : Linear regression example *[29]*

In the figure above, on X-axis is the independent variable and on Y-axis is the output. The regression line is the best fit line for a model. And our main objective in this algorithm is to find this best fit line. [30]

4.2. Support Vector machine (SVM)

Support Vector Machines are one of the most well-known and discussed machine learning techniques. It is essentially a more refined approach. Constructs and performs classification tasks hyper planes are planes in a multidimensional space that are parallel to each other. Cases with various class designations are separated. SVM (Systematic Variable Modeling). [31]**.** We can present it and explain it at this figure (as shown in figure2.8).

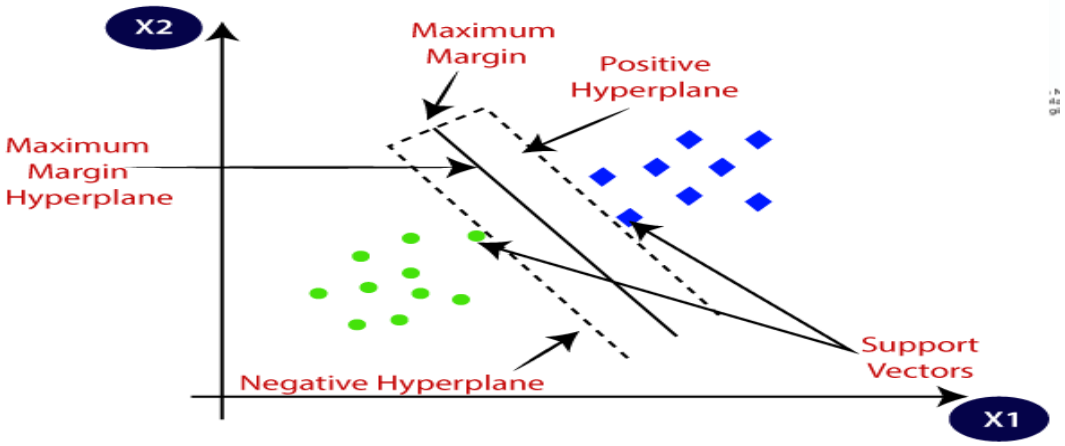

Figure 2.8 : Principle of the svm algorithm *[32]*

4.3. XGBOOST Regression

This belongs to the class of models known as boosting algorithms. This model improves the final decision tree learner by combining the capabilities of several weal learner trees (that is models with higher bias but lower variance) with each other in a sequential manner such that at the end a much more robust decision tree learner is obtained (as shown in figure2.9).

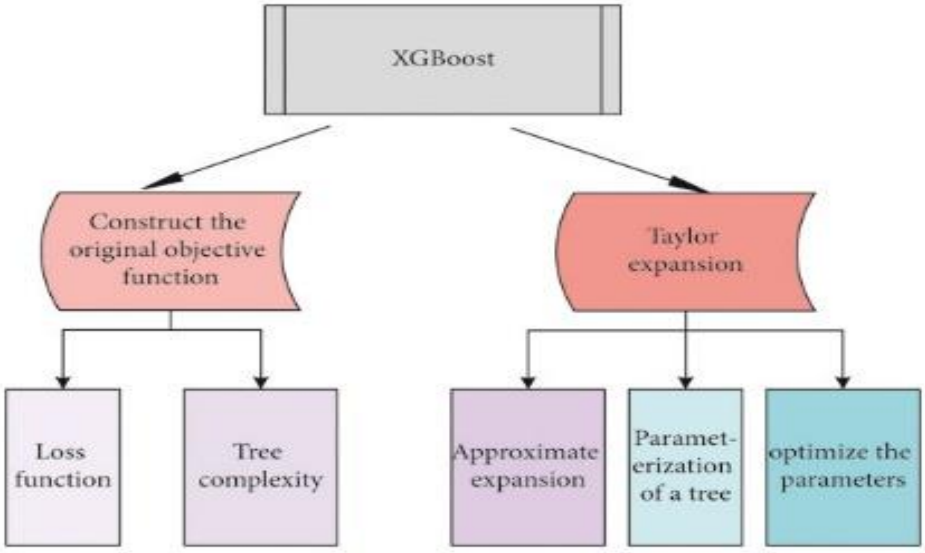

Figure 2.9: Principle of xgboost algorithm *[33]*

4.4. Naïve bayes

A naive Bayes classifier is not a single algorithm, but a family of machine learning algorithms which use probability theory to classify data with an assumption of independence between predictors It is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. (as shown in figure2.10)
. [31]

$$P(A/B)=P(BA))*P(A)\ /P(B). \quad (1)$$

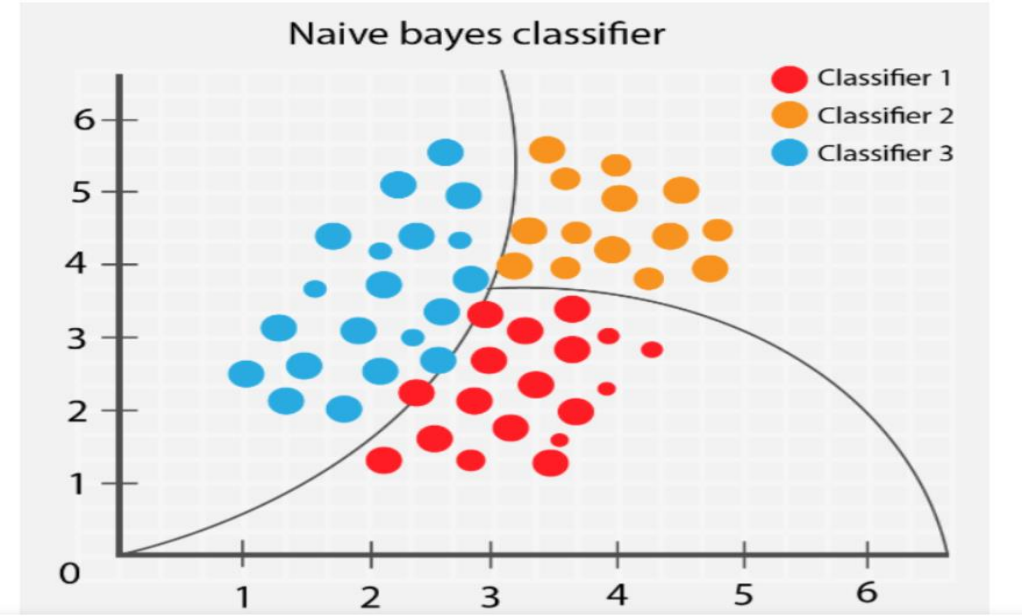

Figure 2.10: Naïve bayes classifier *[34]*

4.5. Random forest

A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification and Regression problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees more it will be robust. Similarly, the greater the number of trees in a Random Forest Algorithm, the higher its accuracy and problem-solving ability. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model. [35]

The principal of random forest algorithm as the following figure (as shown in figure2.11)
:

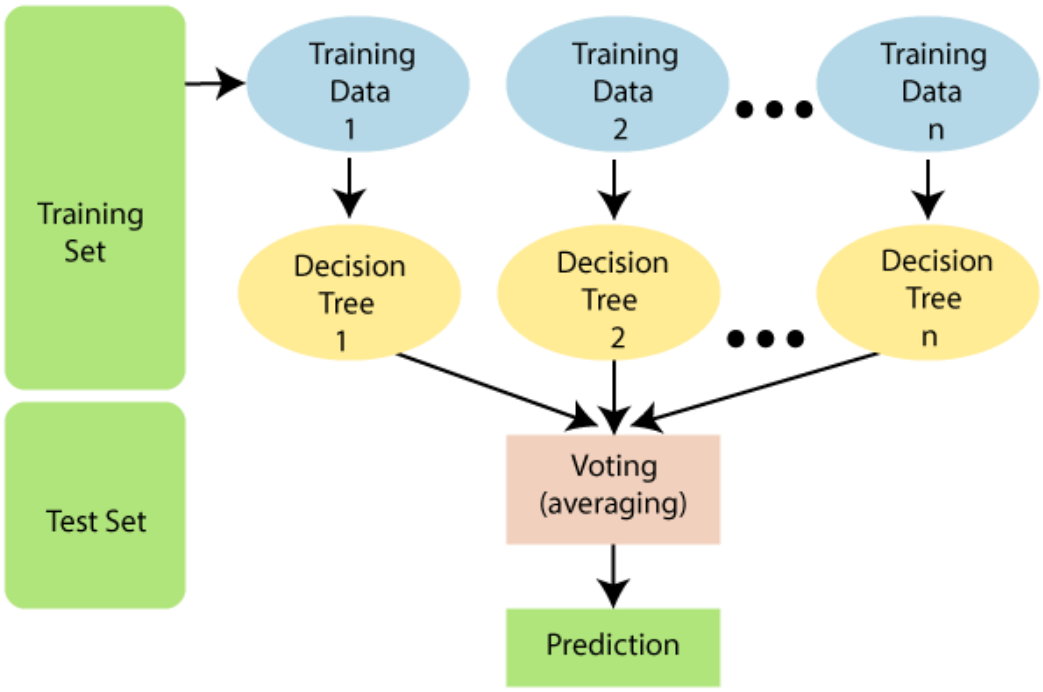

Figure 2.11 : Random Forest principle 1 *[36]*

4.6. Decision tree

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node. [37]

In this figure we describe the principal of decision tree algorithm (as shown in figure2.12).

.

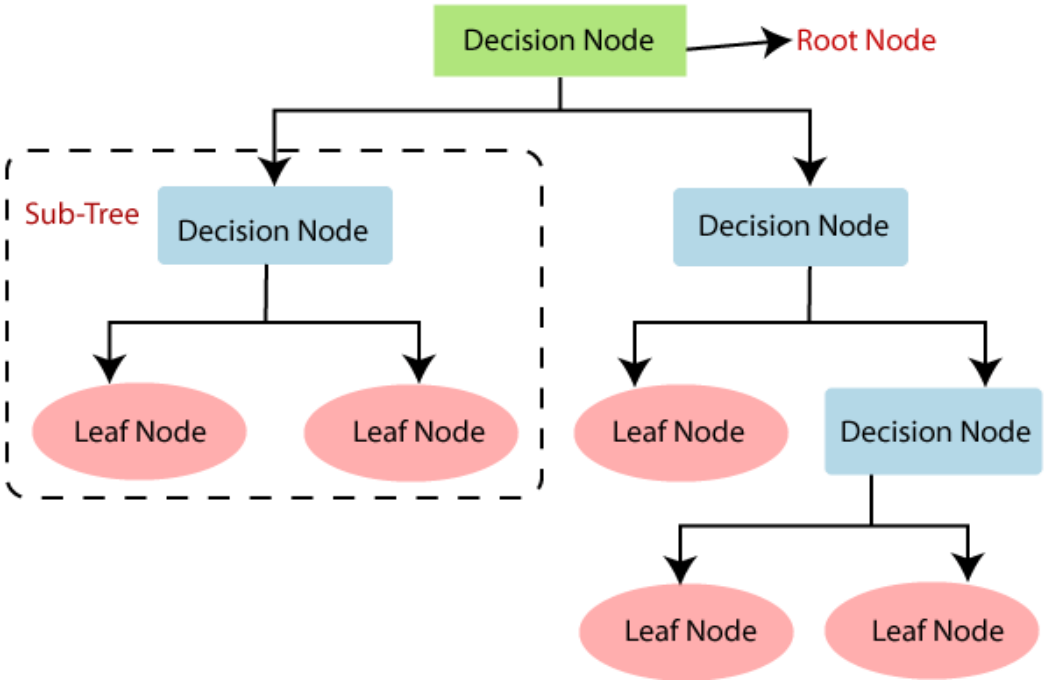

Figure 2.12 : Principle of decision tree *[38]*

Conclusion

Throughout this chapter, we have explored the fascinating world of machine learning, which is a crucial component of artificial intelligence. We have covered the fundamental concepts and techniques of machine learning, as well as the different types of algorithms that are used in this filed. We have also examined the various types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning.

CHAPTER III

STATE OF THE ART

Introduction

In this chapter, we will provide a comprehensive overview of the current research and knowledge in this field, with a specific focus on articles related to calorie burn prediction. Additionally, we will introduce a calorie calculator and provide a guide on how to use it. By summarizing the content as follows.

1. Related work

1) Burned calorie prediction using supervised machine learning regression algorithm. This study is The estimation of calorie burned by individuals is based on a formula and MET charts. This study aims to predict the calorie burned using a regression model as one of the machine learning algorithms to give more accurate results. Model training and testing using K-fold validation were done to determine the best model for the study. The performance and prediction accuracy of regression models were evaluated based on the result of model testing after ten (10) iterations. The average accuracy was computed and the result shows that Random Forest regression is the best model for the study with an accuracy of 95.77%. It is very important to visualize and study the relationships of the variables in the data because it may affect the performance of the algorithm in predicting the value of the target variable. The Random Forest regression model was able to predict the calorie burned with a high accuracy rate. [39]
2) calorie burn prediction using machine learning. By Suvarna Shreyas Ratnakar, Vidya S, at this study, the use of machine learning to help people to get a healthy life. The main part here is people should have adequate knowledge about their calorie intake and burn, keeping a track of their calorie intake is easy. Keeping track of calorie burnt is a difficult part as there are very few devices for that. Calories burned by an individual are based on MET charts and formulas. The main agenda of this study is a prediction of the burnt calorie with the help of an XG boost regression model as the ML (machine learning) algorithm to show accurate results. The model is fed with more than 15,000 data and its mean absolute error is 2.7 which will become better over time by feeding the XG boost regression model with more data. In this study, they had concentrated on the seven primary factors that influence how many calorie the body burns, but there are other factors that also play a role. [40]
3) calorie burn prediction using machine learning. By Gupta, Rachit kumar singh vaibhav, at this paper they talk about Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The object of this research paper is to create a project that can be used predict calorie burnt using Machine Learning with Python. Xgboost Regression model is used in this project.

In this paper, they described xgboost algorithm, linear regression, logistic regression and lasso regression and how they can be used to implement an algorithm for finding the concrete calorie burnt which depend on a number of factors. [41]

2. Calorie calculator

The Calorie Calculator can be used to estimate the number of calorie a person needs to consume each day. This calculator can also provide some simple guidelines for gaining or losing weight. This Calorie Calculator is based on several equations, and the results of the calculator are based on an estimated average. The Harris-Benedict Equation was one of the earliest equations used to calculate basal metabolic rate (BMR), which is the amount of energy expended per day at rest. It was revised in 1984 to be more accurate and was used up until 1990, when the Mifflin-St Jeor Equation was introduced. The Mifflin-St Jeor Equation also calculates BMR, and has been shown to be more accurate than the revised Harris-Benedict Equation. The Katch-McArdle Formula is slightly different in that it calculates resting daily energy expenditure (RDEE), which takes lean body mass into account, something that neither the Mifflin-St Jeor nor the Harris-Benedict Equation do. Of these equations, the Mifflin-St Jeor Equation is considered the most accurate equation for calculating BMR with the exception that the Katch-McArdle Formula can be more accurate for people who are leaner and know their body fat percentage. The three equations used by the calculator are listed below:

Mifflin-St Jeor Equation:

For men:

$$BMR = 10W + 6.25H - 5A + 5 \quad (1)$$

For women:

$$BMR = 10W + 6.25H - 5A – 161 \quad (2)$$

Revised Harris-Benedict Equation:

For men:

$$BMR = 13.397W + 4.799H - 5.677A + 88.362 \quad (3)$$

For women:

$$BMR = 9.247W + 3.098H - 4.330A + 447.593 \quad (4)$$

Katch-McArdle Formula:

$$BMR = 370 + 21.6(1 - F)W \quad (5)$$

where:

W is body weight in kg

H is body height in cm

A is age

F is body fat in percentage

[42]

There are some captures of calorie calculator (as shown in figure3.1) (as shown in figure3.2) :

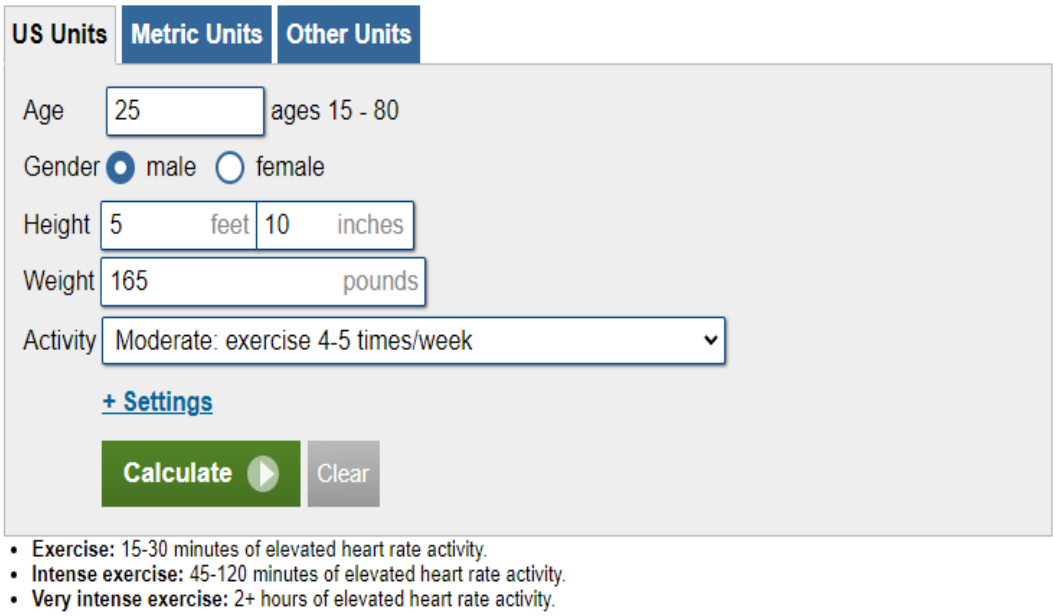

Figure 3.1 : Calorie calculator *[42]*

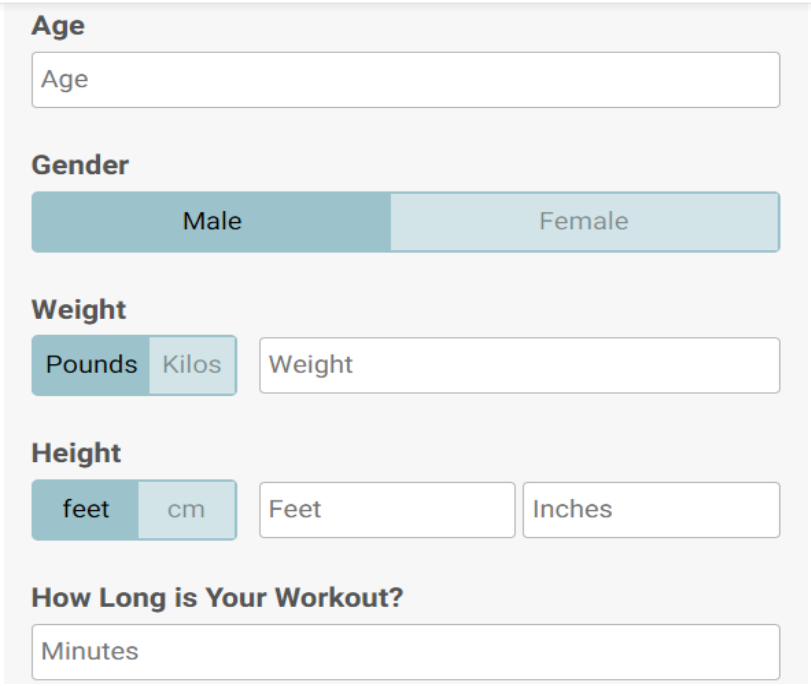

Figure 3.2 : Calorie calculator 2 *[42]*

CHAPTER IV

DESIGN METHODOLOGY

Introduction

In this chapter we will discuss our methodology for building a machine learning model that predicts the number of calorie burned during the workout. We will describe the algorithm we employed and we will also explain how we evaluated the performance of the model.

1. Problematical

The problem we are trying to solve is the inaccuracy of traditional methods of calculating calorie burn during exercise.

These methods are often based on general formulas and do not take into account individual differences such as age, gender, weight. This inaccurate calculation can lead to individuals overestimating or underestimating their calorie burn, which can negatively impact their fitness progress and overall health outcomes.

Our motivation for this work is to help people achieve their fitness goals more effectively and efficiently. By providing a more accurate calculation of calorie burn, our app can help individuals better understand their fitness progress and make more informed decisions about their exercise routines. This can ultimately lead to improved health outcomes and a better quality of life.

The solution we propose is to use machine learning to develop a more accurate and personalized method for predicting calorie burn based on individual characteristics and duration exercise. By using machine learning, we can create a more accurate model that takes into account individual characteristics.

This can help individuals achieve their fitness goals more efficiently and effectively, leading to improved health outcomes and a better quality of life.

2. Methodology steps

Our work consists of the following steps:

- Download the right dataset by kaggle.
- Data preprocessing.
- Training phase: build machine learning models from dataset.
- Testing phase: use features from the dataset.
- Predicting the calorie burn.

the following figure represent the work flow for our ML model (as shown in figure4.1).

.

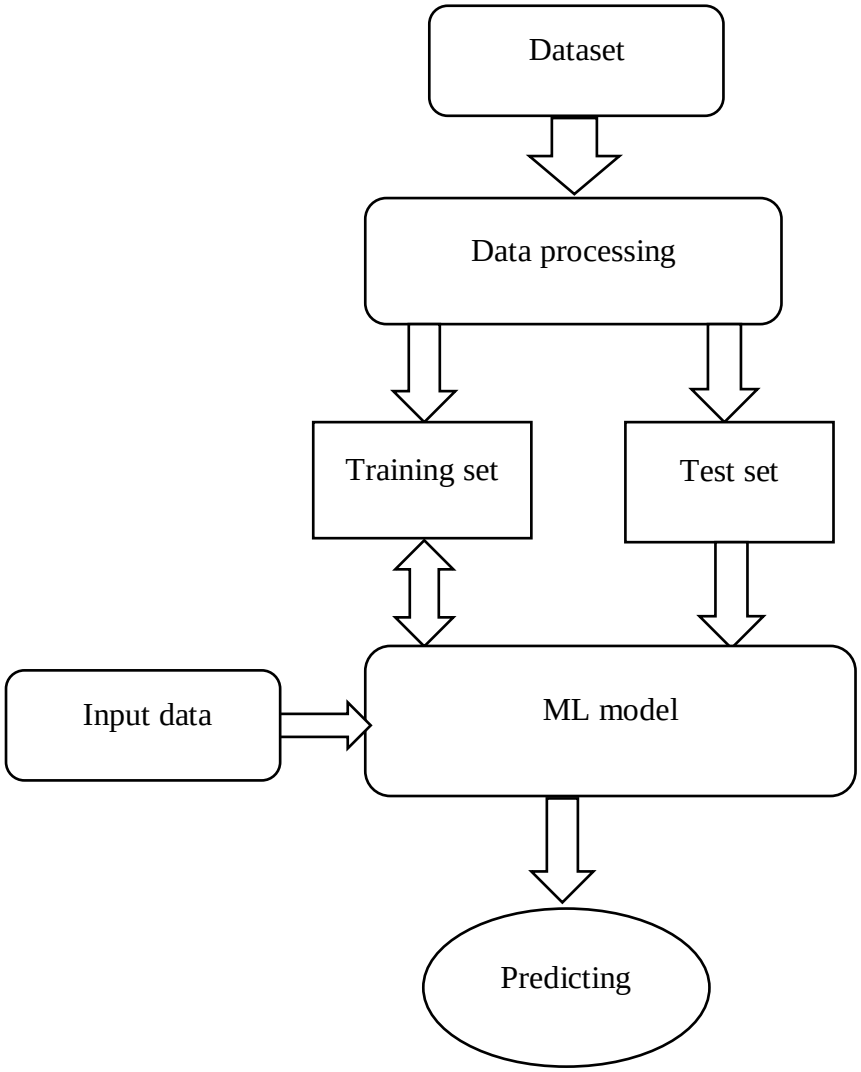

Figure 4.1: Work flow

3. Build machine learning model

In order to build the model, we need data sets and machine learning algorithm(ML), after we have created datasets, we test and select the appropriate algorithm, which gives the best accurate prediction, for this reason we used prediction algorithms: random forest, decision tree, linear regression and XGbosst regression algorithm.

To train the model by the algorithms, we followed those steps:

- Load dataset.

- Preparing data for training: two tasks will be performed in this step. The first task is to divide data into 'attributes' and 'label' sets. The resultant data is then divided into training and test sets
- Divide the data into training and testing sets: 80% training set and 20% testing set.
- Training the model.
- Testing the model by testing set.
- Evaluate the performance of the algorithm.

➢ The prediction accuracy after using random forest algorithm was 99.81 %

➢ The prediction accuracy after using decision tree algorithm was 99.29 %

➢ The prediction accuracy after using linear regression algorithm was 96.68 %

➢ The prediction accuracy after using XGBoost regression algorithm was 99.88 %

After calculating the accuracy of each algorithm and compare them, we have selected XGBoost regression algorithm, because it got the best accuracy value.

The architecture of XGBoost regression

Xgboost regression it is a machine learning algorithm. Use decision tree-based ensemble learning to boost the performance of models. It Make a top choice for regression tasks in various industries. (as shown in the figure4.2)

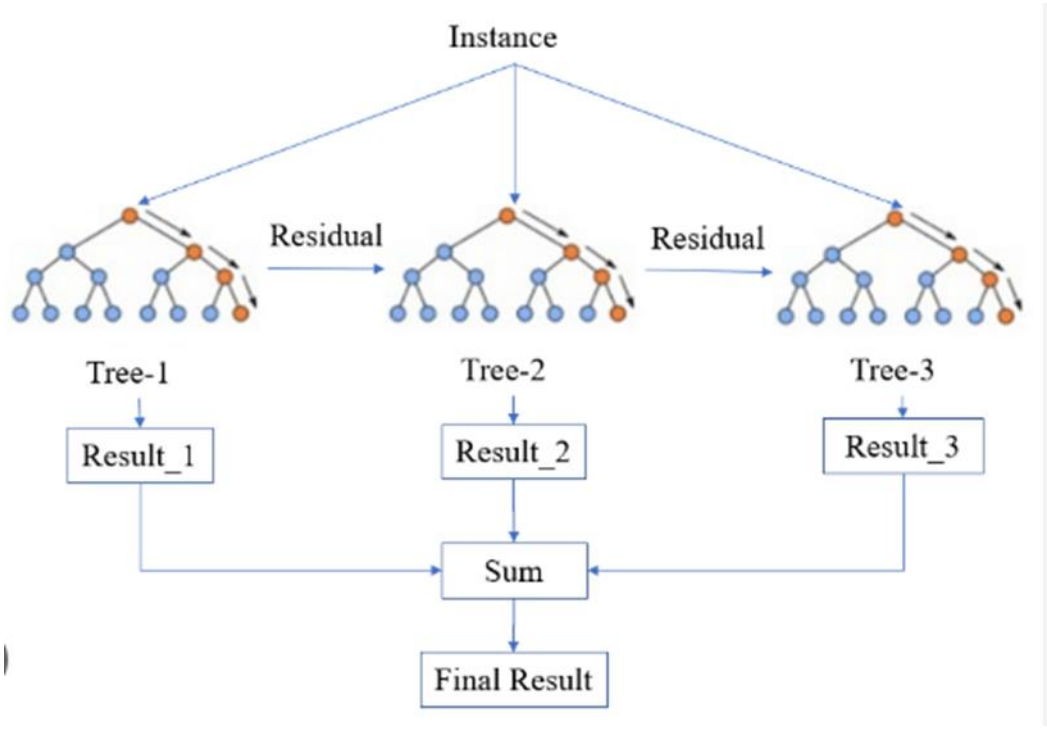

Figure 3.2 XGBoost regression architecture *[43]*

The architecture of XGBoost regression in the context of calorie burn prediction involves building an ensemble of decision trees that are trained to predict the number of calorie burned during physical activity. The input to the algorithm consists of features such as age, weight, height, and activity level, which are used to generate a set of decision rules. Each decision rule is represented by a binary tree, where each node corresponds to a feature and each branch corresponds to a decision based on the value of the feature. The output of the algorithm is a weighted sum of the predictions of the individual decision trees. The weights are determined by the training process, which minimizes a loss function that measures the difference between the predicted and actual calorie burn values. The algorithm uses a combination of regularization, sparsity, and parallelism techniques to improve its accuracy and speed.

4. Application modeling:

Uses case diagram

the main uses case in the system are:

1) fill the input by the user.
2) compare the inputs with the features in dataset.
3) predict the calories.

To understand our uses case diagram, see the figure (as shown in figure4.3).

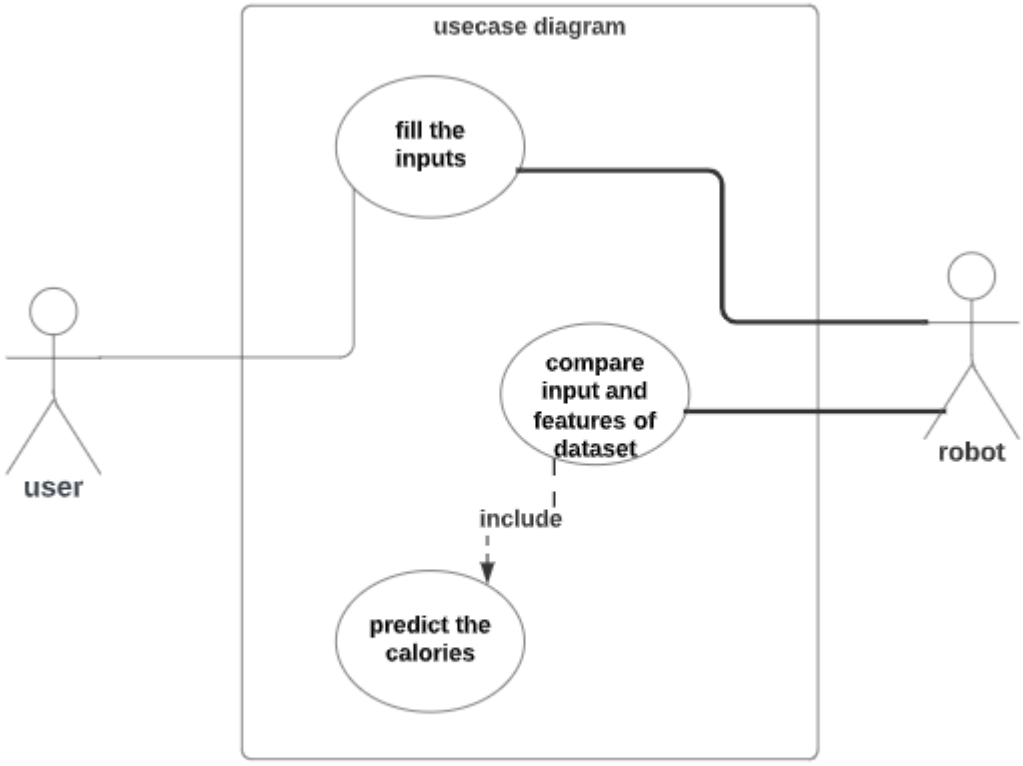

Figure 4.3: Uses case diagram

Flowchart diagram

It is a diagram showing the operational mechanism of the application and its sequential steps which are:

1) fill the inputs by the user.
2) testing the missing values by the system.
3) if there is a missing value it's return error, else it compares the inputs with the features in dataset.
4) predict the result and return it.

And this is the flow chart diagram (as shown in figure4.4)

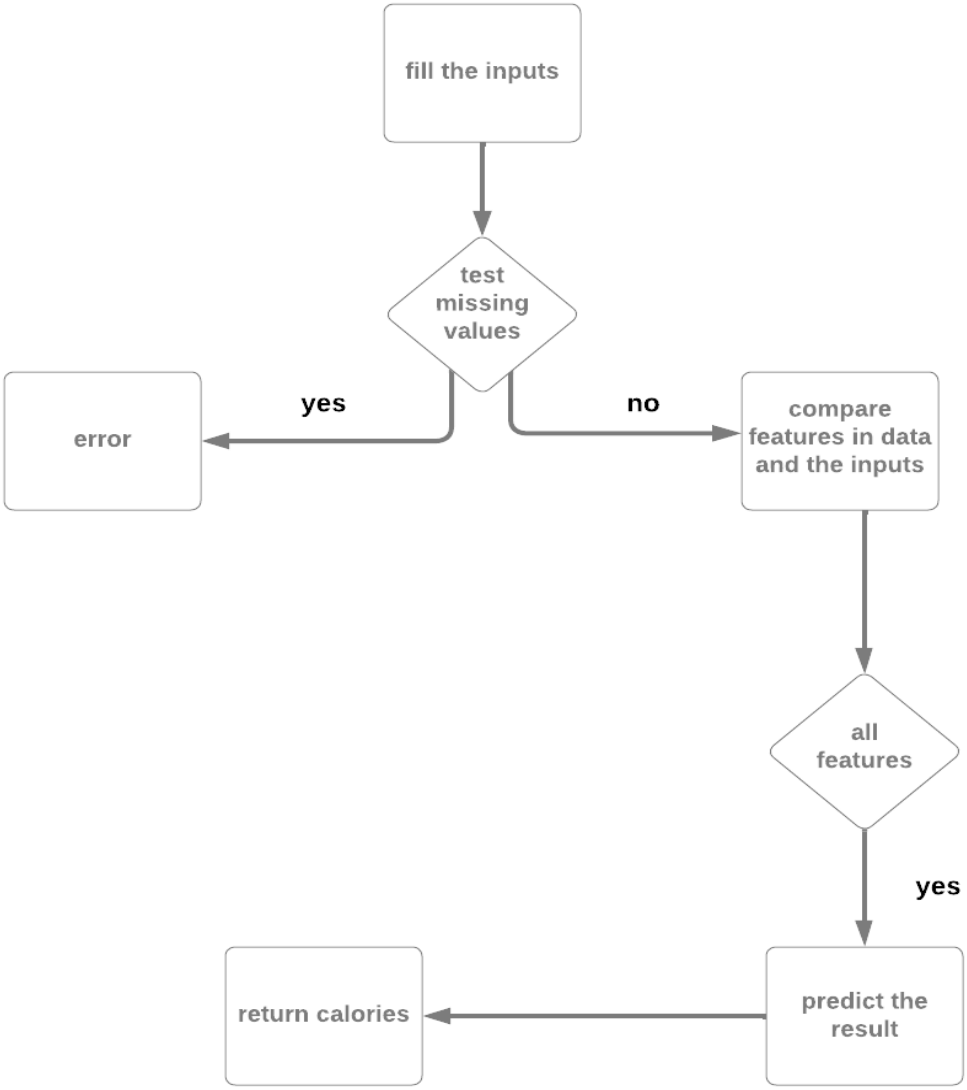

Figure 4.4: Flow chart diagram

Conclusion

In this chapter, we presented in detail our methodology steps to determine the calorie burn prediction and our application modeling via different UML models. In the next chapter, we will present how to implement and experiment our application.

CHAPTER V

IMPLEMENTATION & RESULTS

Introduction

In this chapter, we provide a detailed overview of the programming environment for our project, including the programming language, libraries, and tools used to develop the application. We then describe how we built the ML model for calorie burn prediction. Also we present the implementation details of our application and the results.

1. Anaconda environment

1.1. Definition

Anaconda is a Python package that includes a number of different packages. Libraries, including key libraries used extensively in data science. The biggest benefit This version has the advantage of not requiring sophisticated settings and being able to operate on a variety of platforms. Operating systems and platforms (particularly Windows), which are prone to failure. While attempting to install specified Python packages [44]

Anaconda includes a fantastic integrated development environment (IDE), Spyder (Scientific Python Development Environment), as well as other useful tools like jupyter notebooks, the python console, and the excellent package management tool, conda, which allows us to install, remove, or upgrade any Anaconda package with a single command in Anaconda Prompt. Anaconda 3, which is compatible with Python 3.9.13, is used in our project. [45]

Installation of Anaconda (Python 3.9.13)

Anaconda is easy to install just download Anaconda 64-bits, execute the installer and follow the instruction:

- Search Google for “Anaconda Python” ➔ Python V 3.9.13 for Windows 64 bits
- Install the Anaconda 64-bit installation file for Windows
- Install for ➔ option: Just for me
- Modify anaconda packages by typing the two commands:

- conda update conda
- conda update –all

2. Python language

2.1. Definition

The Python programming language is one of the most widely used languages, with a wide range of applications. It's an interpreted, interactive, object-oriented programming language that's becoming more popular in academia and industry. [46]

Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. [47]

Python for ML and DL Applications is a set of libraries that allow developers to quickly extract and convert data, execute data wrangling operations, apply existing robust ML algorithms, and create bespoke algorithms. These are the libraries that are available. Numpy, Pandas, Scikit-Learn, tensorflow, Keras, and more Python libraries. [48]

2.2. Brief history

Python was developed at the Dutch Mathematics and Informatics Institute (CWI)

- It was first announced in 1991
- Python 2.0 was released on 16 October 2000
- Python 3.0 was released on 3 December 2008

2.3. Installing python

We use in our project the version of python 3.9.13 because its compatible with our devise and contains the libraries that we need in our project.

Just go to (www.python.org) and download.

- Installing Python packages with Conda:
  - C:\>conda install jupyter ➔ jupyter 1.0.0.
  - C:\>pip Install sklearn ➔ la bibliothèque de machine Learning.
  - C:\>pip Install pandas ➔ la bibliothèque de machine Learning.
  - C:\>pip Install pandas-datareader ➔ outils de lecture de la bibliothèque ML pandas.

- C:\>pip Install matplolib ➔ Outils complémentaire de ML.
- C:\>pip Install requests ➔ Outils complémentaire de ML.
- C:\>pip Install tensorflow/tensorflow ==2.11.0 ➔ Installer la bibliothèque de machine Learning et l'interface Google tensorflow.
- C:\>pip Install keras/keras==2.11.0 ➔ Installer la bibliothèque de machine Learning et l'interface keras.

2.4. Jupyter notebook

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter. Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, [49]

We use in our project jupyter 1.0.0

2.5. Python libraries

2.5.1. Numpy

Numpy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more. [50]

2.5.2. Pandas

Pandas is a well-known Python-based data analysis package that offers a wide range of functions, from parsing numerous file formats to transforming an entire data table into a numpy matrix array. As a result, pandas is a reliable partner in data science and machine learning. [51]

2.5.3. Matplotlib

Matplotlib is a Python library that provides interactive and non-interactive 2D charting, as well as the ability to save pictures in a variety of formats. It has a wide range of plots to choose from (lines, bars, pie charts, histograms...etc.). Furthermore, it is adaptable and simple to use. [52]

2.5.4. Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. [53]

2.5.5. Scikit-learn

Scikit-learn is an open source data analysis library, and the gold standard for Machine Learning (ML) in the Python ecosystem. [54]

2.5.6. Request

Python requests is a library for making HTTP requests. It provides an easy-to-use interface that makes working with HTTP very simple, which means it simplifies the process of sending and receiving data from websites by providing a uniform interface for both GET and POST methods. Some of the benefits of using python requests are that they're fast, support multiple languages, and can be piped into other programs to make processing tasks easier. [55]

2.5.7. Tensorflow

TensorFlow is a Brain API and machine-learning library that works with Python machines.

In 2015, Google created a team. Python programmers may now take advantage of this library with ease. Because AI and machine learning are supposed to make usage easier. And Tensor Flow is built on c, C++ making it very fast. It provides primitives for defining functions on tensors and automatically computing their derivatives in domains like physics and engineering, tensors are basic mathematical constructions. Tensors, on the other hand, have historically made less gain in computer science, which has generally been dominated by vectors. Logic and discrete mathematics are more closely related. This situation is beginning to shift.

With the invention of the machine, this has changed dramatically. We utilized tensorflow 2.11.0 [56]**.** [57]

2.5.8. Keras

Keras is a Python-based deep learning API whose goal is to make constructing deep learning models as simple and quick as feasible. We utilized Keras 2.11.0 [58]

2.6. Flask

It is a micro web framework written in Python is play a major role as a communication channel between different services. It has become the de facto standard of passing information across multiple systems in the JSON format [59]

In order to build the API, we use flask.

2.7. Flutter

Flutter is a cross-platform tool to build desktop, mobile, or web applications. Flutter apps are pixel-perfect. Flutter paints the same UI on each app irrespective of the target platform. This is because each Flutter app contains the Flutter engine. This engine renders the Flutter UI code. Flutter provides a canvas for each device and allows you to paint as you want. The engine communicates with the target platform to handle events and interactions. [60]

We use flutter to build the front end for the application.

3. Dataset

3.1. Data source

We use kaggle to download our dataset [61], there are two csv files which present the data, contains at total 15000 instant and 7 attributes, the “kaggle” repository’s contains attributes information about a variety of people including their age, height, weight, gender, body temperature, heart rate, workout duration. This dataset is taken as the training data. And the second calorie dataset comprises the calorie burned by corresponding person.

Table 1: Attributes and their values

Input attributes	**Function**
Gender	Gender (male : 0 , female : 1)
Age	Age mentioned in years (20<age<80)
Height	Height of the person (140<height<220)
Weight	Weight of the person (40<weight<120)
Duration	The time taken to complete the exercising minutes (1<duration<30)

Heart rate	Average heart rate during the workout (70<heart rate<120)
Body temperature	Body temperature in the course of the workout (37<body temperature<40)
Calories	Total calorie burned during the exercise

There are two dataset csv files which are necessary to put their links in the jupyter notebook which used for processing.

For analysis and processing we use data frames. This obtain some statistical measures of the data.

The description of the dataset can be seen in this figure (as shown in figure5.1).

.

```
Entrée [11]: # get some statistical measures about the data
             calories_data.describe()
```

Out[11]:

	User_ID	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
count	1.500000e+04	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000
mean	1.497736e+07	42.789800	174.465133	74.966867	15.530600	95.518533	40.025453	89.539533
std	2.872851e+06	16.980264	14.258114	15.035657	8.319203	9.583328	0.779230	62.456978
min	1.000116e+07	20.000000	123.000000	36.000000	1.000000	67.000000	37.100000	1.000000
25%	1.247419e+07	28.000000	164.000000	63.000000	8.000000	88.000000	39.600000	35.000000
50%	1.499728e+07	39.000000	175.000000	74.000000	16.000000	96.000000	40.200000	79.000000
75%	1.744928e+07	56.000000	185.000000	87.000000	23.000000	103.000000	40.600000	138.000000
max	1.999965e+07	79.000000	222.000000	132.000000	30.000000	128.000000	41.500000	314.000000

Figure 5.1: Dataset

The count of the gender is equally distributed in the dataset which can be seen in the below figure (as shown in figure5.2)

```
Entrée [13]: # plotting the gender column in count plot
             sns.countplot(calories_data['Gender'])

C:\Users\MXP TAIBECHE\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a ke
yword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an expl
icit keyword will result in an error or misinterpretation.
  warnings.warn(

Out[13]: <AxesSubplot:xlabel='Gender', ylabel='count'>
```

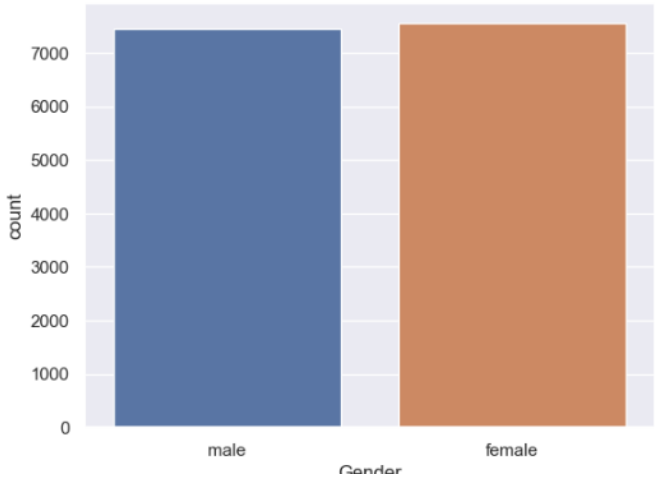

Figure 5.2: Gender plot 1

As well, we have the mean value for age in the figure 3. (as shown in figure5.3).

```
Entrée [14]: # finding the distribution of "Age" column
             sns.distplot(calories_data['Age'])

C:\Users\MXP TAIBECHE\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated func
tion and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with simi
lar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

Out[14]: <AxesSubplot:xlabel='Age', ylabel='Density'>
```

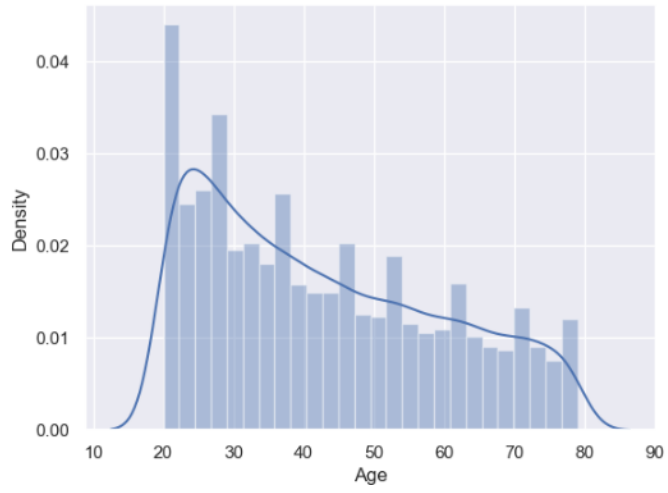

Figure 5.3: Mean age

For age, more values between the age group of 20 and 30 can be seen there is a peek in the curve means from 15000 instances. The decrease in the curve can be seen as people tend to not workout at an older age.

Also, we have value for height and weight present in this figures (as shown in figure5.4).

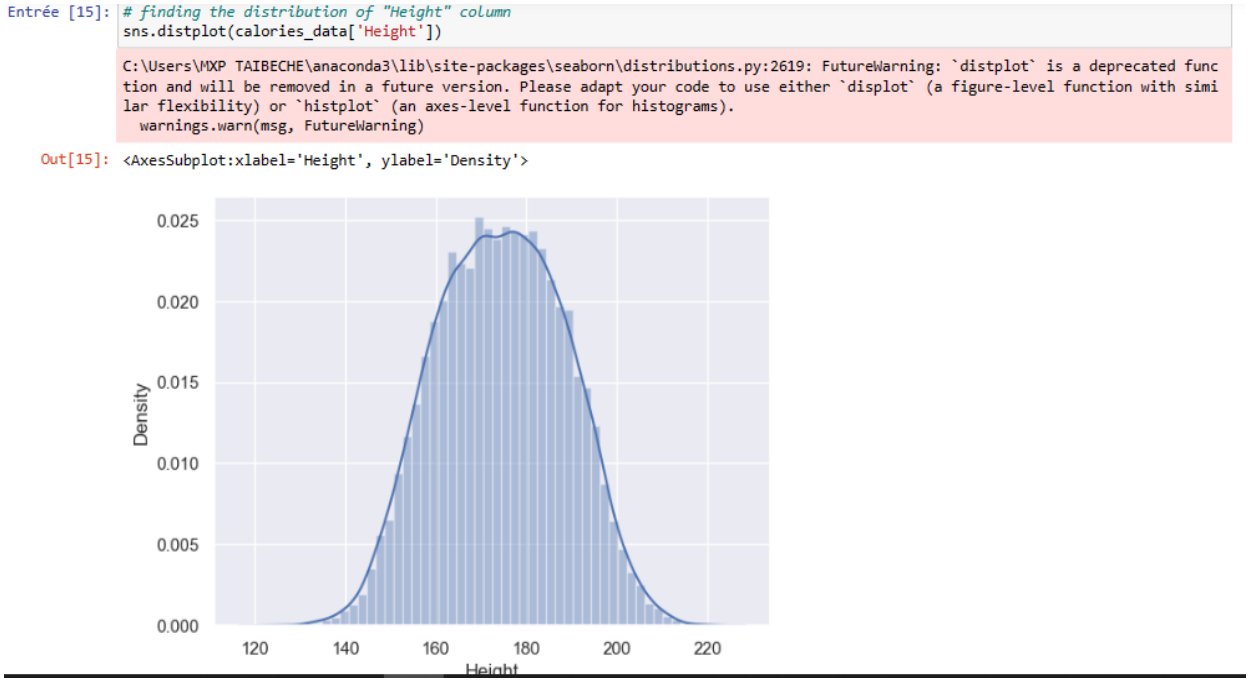

Figure 5.4: Mean height

```
Entrée [16]: # finding the distribution of "Weight" column
             sns.distplot(calories_data['Weight'])

C:\Users\MXP TAIBECHE\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

Out[16]: <AxesSubplot:xlabel='Weight', ylabel='Density'>
```

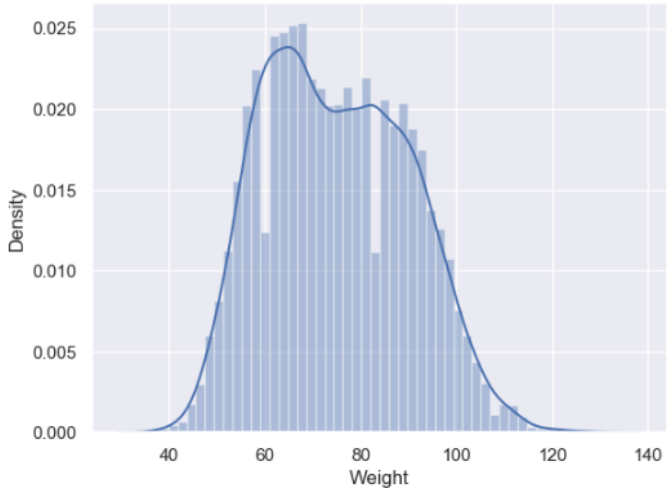

Figure 5.5: Mean weight

The relationship between the different records is then examined. There are two forms of correlation: positive correlation and negative correlation. The quantity of calorie burned will increase as exercise duration increases. These values are therefore proportionate, in the same direction, and connected. (as shown in figure5.6) (as shown in figure5.7).

```
Entrée [17]: #Finding the Correlation in the dataset
             correlation = calories_data.corr()

Entrée [18]: # constructing a heatmap to understand the correlation
             plt.figure(figsize=(10,10))
             sns.heatmap(correlation, cbar=True, square=True, fmt='.1f', annot=True, annot_kws={'size':8}, cmap='Blues')
```

Figure 5.6: Correlation

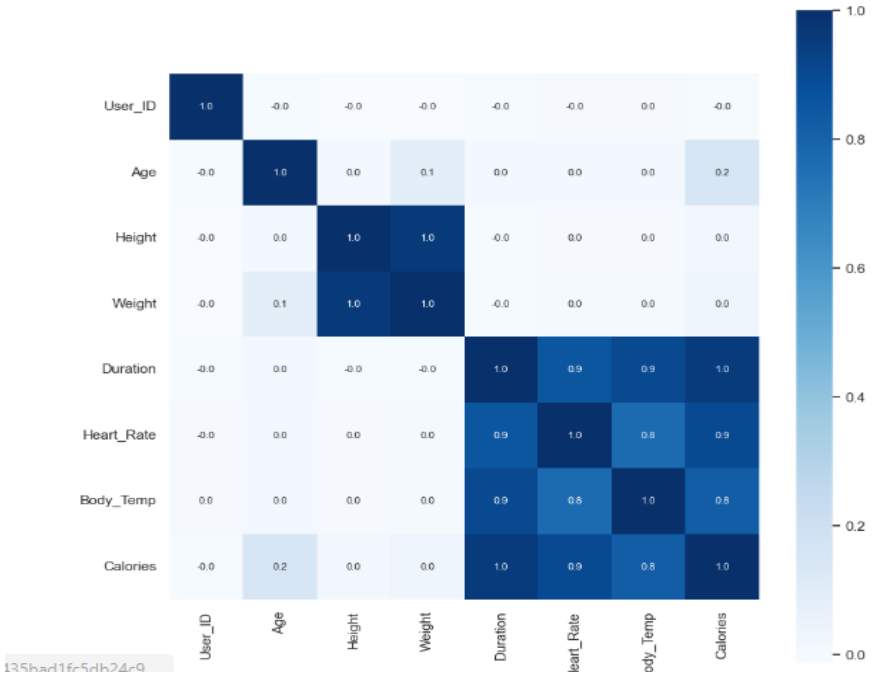

Figure 5.7: Correlation of attribut

4. Machine learning modelling

4.1. Collecting data set

The first step is data retrieval. we use the data repository kaggle. The information collected is category and numerical.

4.2. Data pre-processing

15000 instances and 7 attributes. Each person's attributes are included in kaggle dataset. Including height, gender, age, weight, body temperature, heart rate and exercise duration.

4.3. Data analysis

Jupyter Notebook, the platform used for processing, requires the upload of two dataset csv files ("exercise.csv" and "calorie.csv"). The average body temperature is 40. Those who are exercising will have a higher body temperature. The coronary heart rate and temperature are the most important findings for this analysis. The data must then be visualized using a few charts and graphs. The two types of correlation—positive and negative correlation—are then studied between the various records. After that, load the XGB Regressor model and assess the prediction using test data. This test data and calorie burned for the X test are run through the model.

4.4. Machine learning model

At this step we apply four algorithms (Random forest/Decision tree/Linear regression/XGBoost regressor) then we have chosen XGBoost regressor due to its superior accuracy compared to other algorithms. The XGBoost regressor algorithm has proven to be an effective method in prediction of calorie burn.

Algorithms	Accuracy
XGBoost regressor	99.88%
Decision tree	99.2%9%
Random forest	99.81%
Linear regression	96.68%

4.5. Evaluation

This dataset was analyzed to make predictions about calorie burned based on factors during the workout. We searched for a machine learning model with low mean absolute error that would produce the most accurate results using these machine learning methods.

5. Result

1) Dataset's first five rows:

This table views the first five records in the dataset (as shown in figure5.8)

.

```
Entrée [4]: # loading the data
            exercise_data = pd.read_csv('C:/Users/MXP TAIBECHE/Desktop/exercise.csv')

Entrée [5]: # print the first 5 rows of the dataframe
            exercise_data.head()
```

Out[5]:

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8

Figure 5.8: Data frame

2) Converting the text data to numerical values, see the figure bellow (as shown in figure5.9)

```
Entrée [19]: #Converting the text data to numerical values
             calories_data.replace({"Gender":{'male':0,'female':1}}, inplace=True)

Entrée [20]: # print the first 5 rows
             calories_data.head()
```

Out[20]:

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	0	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	1	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	0	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	1	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	1	27	154.0	58.0	10.0	81.0	39.8	35.0

Figure 5.9: Gender data conversion

3) Splitting the data into training data and test data (as shown in figure5.10)

```
Entrée [24]: #Splitting the data into training data and Test data
             X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)

Entrée [25]: print(X.shape, X_train.shape, X_test.shape)

             (15000, 7) (12000, 7) (3000, 7)
```

Figure 5.10: Splitting of data

4) Model training (as shown in figure5.11)

```
Entrée [26]: #Model Training
             # loading the model
             model = XGBRegressor()

Entrée [27]: # training the model with X_train
             model.fit(X_train, Y_train)

Out[27]: XGBRegressor(base_score=None, booster=None, callbacks=None,
                      colsample_bylevel=None, colsample_bynode=None,
                      colsample_bytree=None, early_stopping_rounds=None,
                      enable_categorical=False, eval_metric=None, feature_types=None,
                      gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
                      interaction_constraints=None, learning_rate=None, max_bin=None,
                      max_cat_threshold=None, max_cat_to_onehot=None,
                      max_delta_step=None, max_depth=None, max_leaves=None,
                      min_child_weight=None, missing=nan, monotone_constraints=None,
                      n_estimators=100, n_jobs=None, num_parallel_tree=None,
                      predictor=None, random_state=None, ...)
```

Figure 5.4: Training of data

5) Prediction on test data (as shown in figure 5.12)

```
Entrée [28]: #Prediction on Test Data
             test_data_prediction = model.predict(X_test)
             print(test_data_prediction)

             [127.823784 226.00154   38.66253  ... 144.3636    22.767195  89.87375 ]
```

Figure 5.5: Prediction on test data

6) Mean absolute error and accuracy (as shown in figure 5.13)

```
Entrée [29]: #Mean Absolute Error
             mae = metrics.mean_absolute_error(Y_test, test_data_prediction)

Entrée [30]: print("Mean Absolute Error = ", mae)

             Mean Absolute Error =  1.4807048829992613

Entrée [31]: #accuracy
             rss = metrics.r2_score(Y_test, test_data_prediction)
             print('R square score = ', rss)

             R square score =  0.9988455491362879
```

Figure 5.6: Mean absolute error and accuracy

7) Predictive system (as shown in figure 5.14)

```
Entrée [32]: #predictive system
             input_data = np.asarray([0, 68, 190.0, 94.0, 29.0, 105.0, 40.8]).reshape(1,-1)
             model.predict(input_data)[0]

Out[32]: 230.88486
```

Figure 5.7: Predictive system

6. Application interface

6.1. First page

In this page contain one button that had other separate page

The first page represents at this figure. (as shown in figure 5.15)

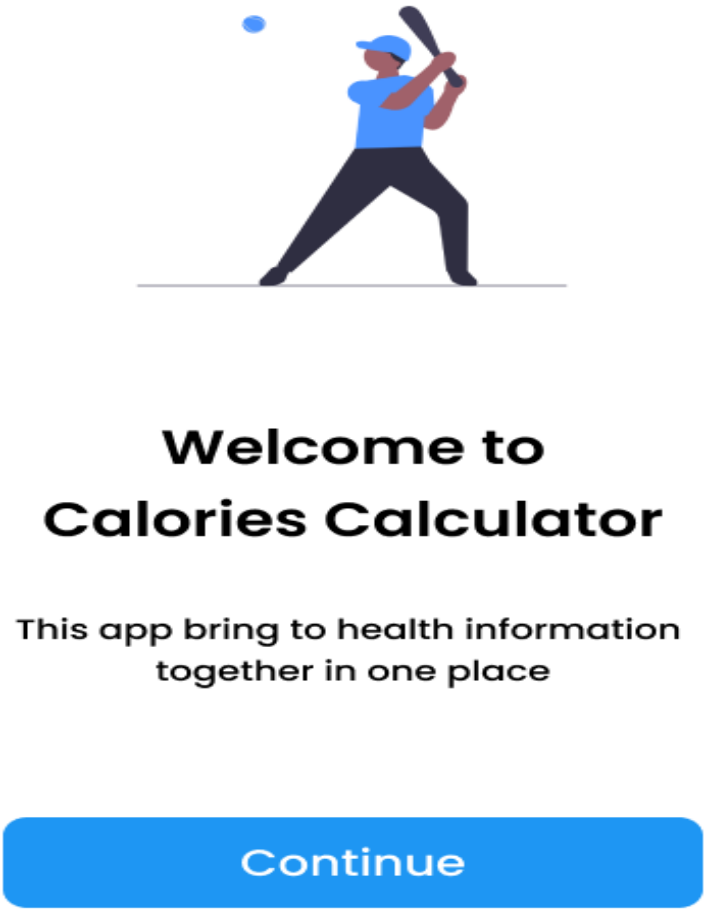

Figure 5.8: Welcoming page

6.2. Second page

This page contains three buttons, each of them has a separate page, that showing in the following figure (as shown in figure 5.16)

Figure 5.9: Home page

6.2.1. Prediction page

This is the first one “calculate calories”, the user must fill the gaps to see the result of burned calorie by clicking in the button “calculate” (as shown in figure 5.17) (as shown in figure 5.18).

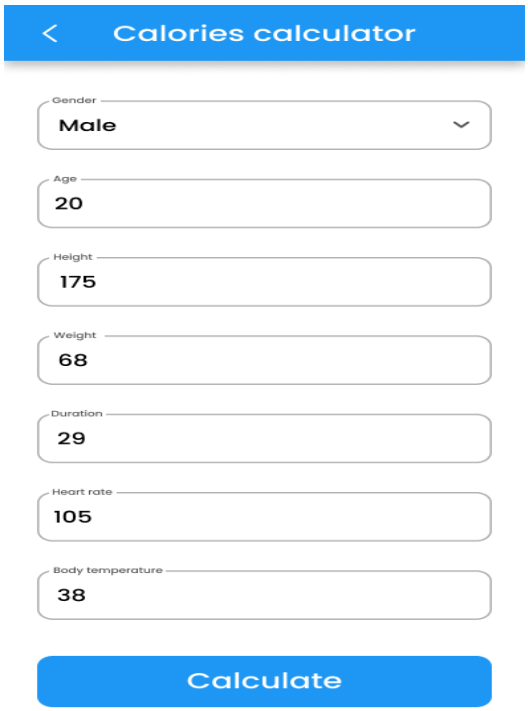

Figure 5. 10 : *Calories calculator*

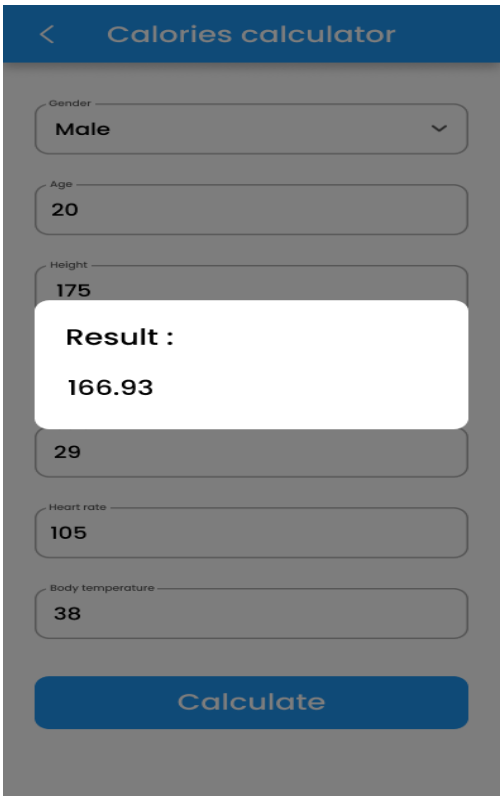

Figure 5.11 : Result page

6.2.2. Coaches List page

This page would help the user to contact coaches to help him for his diet, see the figure bellow (as shown in figure 5.19).

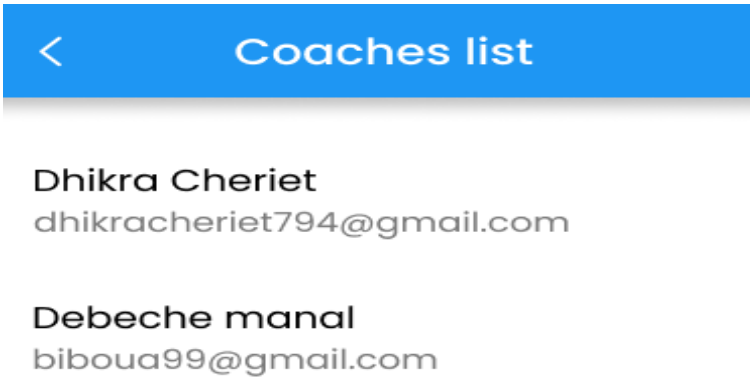

Figure 5.12: Coaches List

6.2.3. Contact us page

This page for if someone want to know something about our application or AI model of predicting (as shown in figure 5.20).

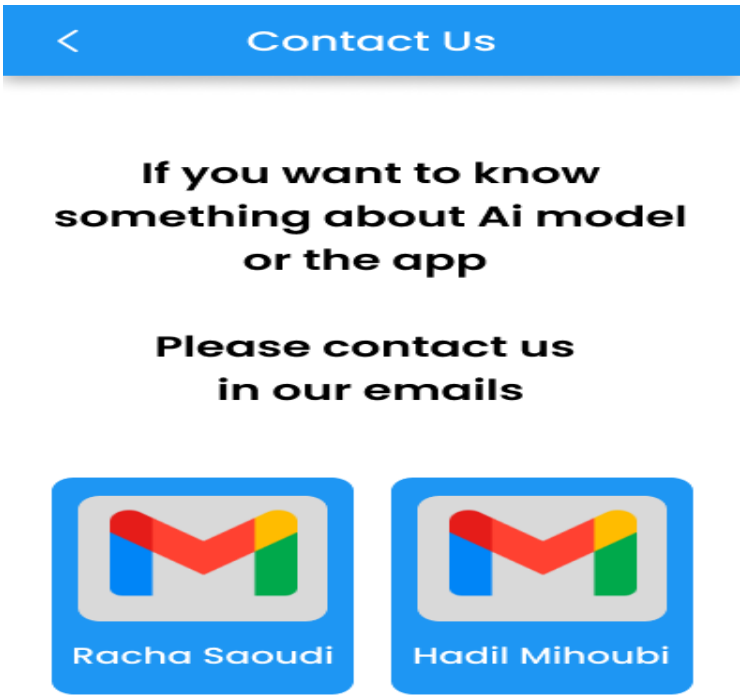

Figure 5.13: Contact us page

Conclusion

In this chapter, we have presented all the techniques and tools that we have utilized to build the android application based on the XGBoost regression machine learning model, which predicts the number of calorie burned during the workout. After conducting extensive experimentation, we can confidently say that our application has produced good, useful, and operational results.

GENERAL CONCLUSION

General conclusion

The main goal of our thesis was to predict the calorie burned during the workout of different people and chose an algorithm in machine learning by looking through the data set.

We deduced from the analysis that the XGB regressor produces more accurate findings. Mean absolute error imply absolute error should be as minimal as possible we find 1.48, it is a good value for the mean absolute value that the XGB regressor gives us. Therefor we can conclude that the best model for the calorie burn prediction is XGBoost Regressor.

In our project we have concentrated on the seven primary factors that influence how many calories our body burns. It’s also important to understand how many calories we are consuming if we want a healthy and fit lifestyle.

Additionally, we build a UI (user interface) is also required so that users may input their values and obtain results that show how many calories they have burned.

Finally, we add a space for personal coaches that could the user contact theme, our emails if they want to know something about the application or the AI model.

Perspectives

Our approach focused more on predicting the burned calories. It could be improved by adding an exercises regimen, a special diet according to the needs of each body. We are able to create a completely functional application with all of these features.

BIBLIOGRAPHY

[1] «understanding calories,» 17 april 2023. [En ligne]. Available: https://www.nhs.uk/live-well/healthy-weight/managing-your-weight/understanding-calories/.

[2] w. bumgardner, «what is a calorie,» 12 octobre 2021. [En ligne]. Available: https://www.verywellfit.com/what-is-a-calorie-3432508.

[3] d. bubnis, «medical news today,» 01 january 2020. [En ligne]. Available: : https://www.medicalnewstoday.com/articles/319731#factors-influencing-daily-calorie-burn-and-weight-loss.

[4] «NIH,» 12 february 2013. [En ligne]. Available: https://www.nhlbi.nih.gov/health/educational/wecan/healthy-weight-basics/balance.htm#:~:text=What%20is%20Energy%20Balance%3F,physical%20activity%20is%20ENERGY%20OUT.

[5] «What Is Energy Balance and How to Use It To Gain or Lose Weight,» [En ligne]. Available: https://www.digitalbarbell.com/blog/2020/9/10/energy-balance.

[6] «BMR Calculator – Basal Metabolic Rate,» [En ligne]. Available: https://www.inchcalculator.com/bmr-calculator/.

[7] M. C. L. F. M. M. G. D. Magrini ML, «wearable devices for caloric intake assessment,» state of the art and future developments. open nurs , 2017.

[8] J. A. Achten J, «heart rate monitoring: applications and limitations.sport med,» 2003.

[9] H. W. W. M. e. a. Ainsworth BE, «compendium of physical activities an update of activity codes and MET intensities. med sci sports exerc,» 2000.

[10] j. chertoff, «byridie,» 20 10 2021. [En ligne]. Available: https://www.byrdie.com/how-many-calories-should-you-burn-in-a-workout-5093459#citation-5.

[11] «Ito S. High-intensity interval training for health benefits and care of cardiac diseases - The key to an efficient exercise protocol,» World J Cardiol, 2019.

[12] «care health insurance,» 27 septembre 2021. [En ligne]. Available: https://www.careinsurance.com/blog/health-insurance-articles/benefits-and-ways-of-burning-calories-for-good-health#:~:text=Benefits%20of%20burning%20Calories,-Health%20is%20much&text=It%20ensures%20your%20blood%20vessels,stress%20and%20feeling%20of%20an.

[13] d. chopra, «popular presentation on artificial intelligence,» 26 september 2022. [En ligne]. Available: https://analyticsdrift.com/popular-presentations-on-artificial-intelligence/.

[14] r. roy, «AI, ML, and DL: How not to get them mixed,» towards data science , 29 april 2020. [En ligne]. Available: https://towardsdatascience.com/understanding-the-difference-between-ai-ml-and-dl-cceb63252a6c.

[15] S. Gadri, «recueil de cours pour la classe master 02 AI. universite m'sila,» 2022. [En ligne]. Available: https://elearning.univ-msila.dz/moodle/pluginfile.php/553914/mod_resource/content/1/Chap%201.%20Apprentissage%20Automatique%2022.pdf.

[16] D. d. l. Fuente, «Dynamic Scheduling of Manufacturing Systems with Machine Learning.,» [En ligne]. Available: https://www.researchgate.net/figure/Machine-learning-general-scheme_fig1_220180151.

[17] V. Kanade, «What Is Machine Learning? Definition, Types, Applications, and Trends for 2022,» 30 august 2022. [En ligne]. Available: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/.

[18] [En ligne]. Available: https : //us.hudson.com/legal/blog/postid/513/predictive-artificial-intelligence-science-fiction-e-discovery-truth.

[19] P. Baheti, «Supervised and Unsupervised Learning [Differences & Examples],» 1 october 2021. [En ligne]. Available: https://www.v7labs.com/blog/supervised-vs-unsupervised-learning.

[20] d. castillo, «seldon blog,» 29 october 2021. [En ligne]. Available: https://www.seldon.io/machine-learning-regression-explained.

[21] a. molodoria, «mobidev,» 18 aug 2022. [En ligne]. Available: https://mobidev.biz/blog/essential-machine-learning-algorithms-for-business-applications.

[22] j. brownlee, «machine learning mastery,» 08 april 2020. [En ligne]. Available: https://machinelearningmastery.com/types-of-classification-in-machine-learning/.

[23] Akanksha_Rai, «geeks for geeks,» [En ligne]. Available: https://www.geeksforgeeks.org/supervised-unsupervised-learning/.

[24] A. Jeffares, «Supervised vs Unsupervised Learning in 3 Minutes,» 24 jul 2018. [En ligne]. Available: https://towardsdatascience.com/supervised-vs-unsupervised-learning-in-2-minutes-72dad148f242.

[25] a. gupta. [En ligne]. Available: https://www.geeksforgeeks.org/ml-semi-supervised-learning/?ref=rp.

[26] «Semi-Supervised,» [En ligne]. Available: http://primo.ai/index.php?title=Semi-Supervised.

[27] p. bajaj, «geeks for geeks,» [En ligne]. Available: https://www.geeksforgeeks.org/what-is-reinforcement-learning/.

[28] «Reinforcement Learning Algorithms and Applications,» [En ligne]. Available: https://techvidvan.com/tutorials/reinforcement-learning/.

[29] L. Brahim, «Comparative Analysis of Machine Learning for Predicting Air Quality in Smart Cities,» [En ligne]. Available: https://www.researchgate.net/figure/Graph-of-Linear-regression_fig2_361730374.

[30] G. sharma, «Analaytics vidhya,» data science blogathon, 26 may 2021. [En ligne]. Available: https://www.analyticsvidhya.com/blog/2021/05/5-regression-algorithms-you-should-know-introductory-

guide/#:~:text=Linear%20Regression%20is%20an%20ML,the%20other%20given%20independent%20variables..

[31] V. Jatana, «Dive Into machine learning,» SRM Institute of Science and Technology, November 2018 2018. [En ligne]. Available: https://www.researchgate.net/publication/328614973_Dive_Into_Machine_Learning.

[32] «java svm,» [En ligne]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm.

[33] Z. Kaifa, «Advanced Big Data Analysis Technologies for Environmental Monitoring Data,» [En ligne]. Available: https://www.hindawi.com/journals/jeph/2022/2288321/.

[34] K. D. Chaudhuri, «Building Naive Bayes Classifier from Scratch to Perform Sentiment Analysis,» [En ligne]. Available: https://www.analyticsvidhya.com/blog/2022/03/building-naive-bayes-classifier-from-scratch-to-perform-sentiment-analysis/.

[35] «simpli learn,» 23 feb 2023. [En ligne]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm.

[36] «Random Forest Algorithm,» [En ligne]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm.

[37] N. s. chauhan, «Decision tree algorithm, Explained,» kdnuggets, 09 february 2022. [En ligne]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html.

[38] «Decision Tree Classification Algorithm,» [En ligne]. Available: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.

[39] A. G. A. J. N. M. M. A. F. M. J. A. B. S. a. J. S. G. M. Nipas, «Burned calories prediction using supervised machine learning:regression algorithm,» *Second international conference on power control and computing technology,* 2022.

[40] V. S. Suvarna Shreyas Ratnakar, «calorie burn prediction using machine learning,» international advanced research journal in science engineering and technology , 06 june 2022. [En ligne]. Available: https://iarjset.com/wp-content/uploads/2022/07/IARJSET.2022.961125.pdf.

[41] R. k. s. v. Gupta, «calorie burn prediction using machine learning,» international advanced research journal in science engineering and technology , 5 may 2022. [En ligne]. Available: https://ijarcce.com/wp-content/uploads/2022/06/IJARCCE.2022.115105.pdf.

[42] «calorie calculator,» [En ligne]. Available: https://www.calculator.net/calorie-calculator.html.

[43] [En ligne]. Available: https://www.researchgate.net/figure/Simplified-structure-of-random-forest_fig1_348025909.

[44] D. B. R. A. S. T. N. Sarkar, « Practical Machine Learning With Python,» p. P29.

[45] B. A. R. V. Alvarado, *Python Con Anaconda.,* vol. Researchgate, p. P6, 2018.

[46] «What is Python?,» Executive Summary,python organization, [En ligne]. Available: https://python.sdv.univ-paris-diderot.fr/..

[47] [En ligne]. Available: https://www.python.org/doc/essays/blurb/.

[48] S. A. M. V. N. Raschka, «Python Machine Learning.,» pp. P,58.

[49] m. driscoll, «jupyter notebook introduction,» [En ligne]. Available: https://realpython.com/jupyter-notebook-introduction/.

[50] «what is numpy?,» [En ligne]. Available: https://numpy.org/doc/stable/user/whatisnumpy.html.

[51] «what is pandas in python,» [En ligne]. Available: Https://Www.Educative.Io/Edpresso/What-Ispandas-In-Python.

[52] S. Tosi, « Matplotlib For Python Developers. Birmingham, U.K,» pp. P,8, 2009..

[53] [En ligne]. Available: https://seaborn.pydata.org/#:~:text=Seaborn%20is%20a%20Python%20data,introductory%20notes%20or%20the%20paper..

[54] «what is scikit learn in python,» 05 august 2022. [En ligne]. Available: https://www.activestate.com/resources/quick-reads/what-is-scikit-learn-in-python/#:~:text=Scikit%2Dlearn%20is%20an%20open,categorizing%20data%20based%20on%20patterns..

[55] «python request,» 23 Feb 2023. [En ligne]. Available: https://www.simplilearn.com/tutorials/python-tutorial/python-requests.

[56] V. B. K. Rudraswamimath, «Handwritten Digit Recognition Using Cnn'',» *International Journal Of Innovative Science And Research Technology.,* 2009.

[57] Bharath Ramsundar, « Tensorflow Tutorial,» 2015.

[58] «about keras,» [En ligne]. Available: Https://Keras.Io/About/.

[59] [En ligne]. Available: https://flask.palletsprojects.com/en/2.0.x/.

[60] o. nwabude, «Flutter For Front-End Web Developers,» 05 sep 2022. [En ligne]. Available: https://css-tricks.com/flutter-for-front-end-web-

developers/#:~:text=Flutter%20is%20a%20cross%2Dplatform,app%20contains%20the%20Flutter%20engine..

[61] kaggle. [En ligne]. Available: https://www.kaggle.com/code/pragathiputhran/calories-burnt-prediction.

[62] [. l. A. :. https://www.nhs.uk/live-well/healthy-weight/managing-your-weight/understanding-calories/..

[63] D. d. l. Fuente, «Dynamic Scheduling of Manufacturing Systems with Machine Learning.,» [En ligne]. Available: https://www.researchgate.net/figure/Machine-learning-general-scheme_fig1_220180151.

[64] D. d. l. Fuente, «Dynamic Scheduling of Manufacturing Systems with Machine Learning.,» [En ligne]. Available: https://www.researchgate.net/figure/Machine-learning-general-scheme_fig1_220180151.

[65] [En ligne]. Available: https://www.researchgate.net/figure/Simplified-structure-of-random-forest_fig1_348025909.

Abstract

Life is all about finding balance. and that's most important when it comes to our body. However, staying fit and healthy necessitates frequent physical activity. The variety of burned energy in daily life is directly related to weight maintenance, weight gain, or weight loss. people need to know how many

calories they burned each day. Our project is predicting the calorie burned during the workout with the use of machine learning algorithm XGBoost regressor model approach to produce accurate results. the model is fed with more than 15000 data and its mean absolute error is 1.48. Therefore, we built a mobile application that help the users easily by put their values obtain results of burned calorie.

Key words: XGBoost regressor, machine learning, accurate.

الملخص:

الحياة تدور حول إيجاد التوازن. هذا هو الأكثر أهمية عندما يتعلق الأمر بجسمنا، ومع ذلك فإن الحفاظ على لياقتك وصحتك يتطلب نشاطًا بدنيًا متكررًا. يرتبط تنوع الطاقة المحروقة في الحياة اليومية ارتباطًا مباشرًا بالحفاظ على الوزن أو زيادة الوزن أو فقدانه فيحتاج الناس إلى معرفة عدد السعرات الحرارية التي يحرقونها كل يوم. يتنبأ مشروعنا بالسعرات الحرارية المحروقة أثناء التمرين باستخدام نهج نموذج الانحدار XGBoost لخوارزمية التعلم الآلي لتحقيق نتائج دقيقة. يتم تغذية النموذج بأكثر من 15000 من البيانات ومتوسط الخطأ المطلق هو 1.48. لذلك، قمنا ببناء تطبيق هاتف المحمول يساعد المستخدمين بسهولة، بإدخال معلوماتهم للحصول على نتائج السعرات الحرارية المحروقة.

الكلمات المفتاحية :التعلم الآلي، الدقة.