B.Sc. (Hons) in Software Development

Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

Biometric Data Analysis in Digital Game Scenario

By
Rodrigo Almeida

April 24, 2024

Minor Dissertation

Department of Computer Science & Applied Physics,
School of Science & Computing,
Atlantic Technological University (ATU), Galway.

Contents

List of Figures

List of Tables

Chapter 1

Abstract

Background and Objective The aim of this project is to develop a robust and enduring system using live data to analyze the correlation between a gamer's biometrics data and performance in a digital gaming scenario and corroborate the correlation by finding a function that approximates the relationship using machine learning algorithm.

Users Biometric data were collected using a wearable device, Polar Vantage V2 and performance metrics were captured with a test game designed to test users performance in a first-person shooter gaming scenario. Volunteers were selected and enrolled to gather biometric and test data. Heart Rate Variability, Heart Rate Average and Quality of Sleep were the biometric data collected and were used as the independent variables for the machine learning algorithm. The dependent variables to be predicted are the user's performance metrics in the test game. The performance metrics chosen were Fine Motor Average Tracking Time, Fine Motor Accuracy, Visual Average Response Time, Visual Average Tracking Time, Visual Accuracy and Audio Average Response Time.

An on-cloud solution was proposed and implemented to access users data and do real-time analysis. The solution was implemented using Amazon Web Services (AWS) as the cloud provider, Amazon Relational Database Service (RDS) as a Platform as a Service (PaaS) provider and Amazon Elastic Compute Cloud (EC2) as an Infrastructure as a Service (IaaS) provider.

Using already predefined algorithm, a real-time model selection, hyperparameter tuning and model training pipeline was implemented to analyze the collected data. The following observations and assertions were made from the analysis of the data.

Chapter 2

Introduction

First person shooter games represents the class of games where the player views the environment through a viewport and can perform such actions as looking around, moving around, aiming and firing of weapons. these actions are accomplished using various button or combination of button.

During a typical gameplay, players are confronted with other opposing players and are required to eliminate their opponents using various weapons available while evading enemy fire. To successfully compete in such a scenario, players are expected to react fast, effectively track targets, accurately hit targets, perceive sound and accurately map them to a location within their environ.

The primary objective of this research work is to formulate a metrics that can accurately measure users performance in the such scenario and compare their performance with their fitness data with the aim of finding a correlation between performance and physiological state.

2.1 Background

PUBG: Battlegrounds (previously known as PlayerUnknownś Backgrounds) is a battle royale style player versus player (PvP) shooter game developed by PUBG Studio. Players face-off with each other using various types of battlefield weapons in a last man standing deathmatch and the last person to remain alive wins. The game is available in all major platforms and as of March 2021, the mobile version of the game has accumulated more than a billion download outside of China with revenue of over $9billion while the PC and console versions have accumulated a total revenue of $4billion [?].

Since its first release in 2017, the game has since become one the fans favorite and has over '350,000' peak concurrent monthly users [1]. As a multiple award-

[1] statista

winning game with proven longevity records and a large community. Interest in the game cut across diferent demography and is equally far-reaching across the globe. The game playing scenario requires players to face-off with other players and there is where some skills like 'eye-hand-coordination', 'ear-hand-coordination', 'fine-motor' skills, etc: are required to compete favorably against other players. Players have access to a varieties of weapons with different capabilities and can make in-game adjustments to their control to suite their various preferences.

This project is a continuation of research work previously done by Fourth Year Software Design Students titled 'Biometric Data Collection for Performance Optimization in a Digital Game Scenario' in collaboration with the Department of Sports & Excercise Science, Atlantic Technological University. The originiating project titled **'Biometric Data Collection for Performance Optimization in a Digital Game Scenario'**, posed the question 'can a playerś biometric data be used to optimize their performance in a first-person shooter game'? And a subsequent follow up project which sought to create a Chart API capable of displaying all relevant information previously displayed on different pages on a single page.
The former research was geared towards creating a platform for collecting performance data in a similar scenarios (Weapons, controls, user perspective, ect:) obtainable in PUBG :Battlegrounds in the form of a Unity Desktop `Application`. Collection and storage of Biometric data from an `Activity Monitor` in the form of a `Smart Watch`. With the eventual goal of finding correlation between their performance and their Biometric data.

2.2 Performance Metrics

For the purpose of measuring users performance in in a first-person shooter game scenario, three categories of metrics where developed to measure users performance. They are listed as follows: Fine-Motor Control, Visual metrics and Audio metrics.

2.2.1 Fine Motor Control

Fine motor refers to the controlled and coordinated incremental movements made by the hand when handling items. This metrics assesses how quickly users are able to adjust their aim to targets appearing on the screen. Two take away from this classification is the Response Time (seconds) which quantifies reaction time and Accuracy (%) which quantifies users fine adjustment capabilities.

2.2.2 Visual Reflexes

The game scenario under study for most of the time involves users having multiple legitimate targets to shoot, and having to engage them simultaneously. This metrics accesses users visual reflexes by considering the Target Accuracy which is the measure of the number of successful targets hit with the total number of target spawned expressed in percentage. Shot Accuracy which is the measure of the number of successful shots to the number of targets hit. Finally the Average Response Time which is the average time it takes for a target to appear before being hit.

2.2.3 Audio Reflexes

User audio reflexes are accessed using the Response Time, which measures the time it takes to identify the source of a sound within the users environment

2.3 Biometric Data

The collection and analysis of biometric data play a significant role in understanding the relationship between physical fitness and gaming performance. This research aims to investigate how various physical parameters, such as HRV (Heart Rate Variability), Heart Rate Max, Heart Rate average, Active Sleep and Quality of Sleep, influence a user's performance in a digital game, in this case PUBG: Battlegrounds. The research will also explore the potential of using biometric data to suggest the most suitable settings for different game scenarios. By using wearable technology, which can monitor metrics such as heart rate, sleep, and activity, this study seeks to establish a correlation between the player's physical condition and their performance in a digital game scenario. The importance of fitness data in this context cannot be overstated. Heart Rate Variability (HRV), maximum heart rate, active sleep, and sleep quality are all essential indicators of a person's physical condition. THese metrics offers a comprehensive insight into the user's physical condition, which can be used to improve their gaming performance.

2.3.1 Heart Rate Variability (HRV):

is a measure of the variation in time between each heartbeat, and it is closely linked to the body's stress levels. HRV is widely used as a measure of the body's autonomic nervous system, which controls the body's stress response. A higher HRV is associated with a lower stress level, while a lower HRV is associated with a higher stress level. A research from E. Ortega[2] highlights the importance of HRV in sports science, as a way of understanding the psychological state of

atheletes before competions. It found a positive correlation between HRV, self-efficacy, and performance among sport shooters. Advanced shooters demonstrated lower average heart rates and employed mental skills more effectively than less experienced shooters. This suggests HRV as a valuable asset, when tranferring the psychological state of the athlete to the performance in a digital game. It allows for a better understanding of the player's stress levels, and how it affects their performance. This personalized approach can enhance player experience, and potentially improve their performance in the game.

The Heart Rate Variability measurements typically requires a chest strap, and measurements typically span from 5 minutes to 24 hours, it is commonly used in clinical settings to evaluate cardiac conditions.[3] Short-term HRV analysis lasting less then 5 minutes have been proven to provide more accurate estimations compared to longer measurements.[4] Based on this, the research will focus on short-term HRV measurements, as it is more practical for the user, and it provides accurate estimations. It will be more detailed in the methodology section.

Heart Rate Max: is the maximum number of times the heart can beat in a minute, and it is a measure of the body's cardiovascular fitness. It is an important indicator of the body's physical condition, and it is used to evaluate the body's ability to perform physical activities, reflecting their cardiovascular fitness. In this research, understanding HRmax is vital for assessing participants fitness and endurance. It could indicates how well players can handle stress and maintain their concentration over extended periods of time. By analysing alongside other biometric data this research could uncover valuable insight into optmizing players performance in a digital game scenario.

Heart Rate Average: is the average number of times the heart beats in a minute, and it is a measure of the body's physical activity. It is also an important indicator of the body's physical condition, and it is another measure of the body's cardiovascular fitness. It is used to evaluate the body's ability to perform physical activities, reflecting their cardiovascular fitness.

Active Step: Active steps, as a mesure of physical activity, is an important indicator of the body's physical condition. It is used to evaluate the body's ability to perform physical activities, reflecting their cardiovascular fitness. Counting the numbers of steps taken daily can help track overall physical activity, and it is a good indicator of the body physical condition. It is also used to evaluate the body's ability to perform physical activities, reflecting their cardiovascular fitness.

Quality of Sleep: are important indicators of the body's recovery and readiness for physical activity. Active sleep is a measure of the body's physical activity during sleep, and it is an important indicator of the body's recovery and readiness for physical activity.High-quality sleep, marked by sufficient duration and minimal disruptions, is fundamental for overall health impacting mood, cognitive

function. A research that evaluated Sleep and performance in Eathletes [5] uderscores the critical role of sleep in Eathletes performance in Esports, showing its impact on cognitive functions as a crucial factor for competitives success. Adequate sleep improves information processing, visual motor functioning, attention, working memory, and other functions essential for decision-making and reaction time. On the other hand, sleep depravation can significantly impair these cognitive abilities, potentially to poor performance. For this research it proves the importance of sleep in the context of gaming performance, and how it can be used to predict the player's readiness and tune their performance in a digital game scenario.

2.4 Classification Model

2.5 Regression Model

Chapter 3

Methodology

This research is geared towards finding a correlation between users' biometric data to their performance in a first-person shooter gaming scenario. The quality of the user data collected was critical in ensuring that the research aims were achieved. For Quality Assurance purposes, a systematic step-by-step approach has been designed for the data-capturing process. An activity monitoring device was distributed to the volunteers in the form of a Polar Watch. The individuals were instructed to use these devices at pre-designated times prior to undertaking a test in the form of a First-Person Shooter Game. Results from the test, was than paired with their biometric data for further analysis.

3.1 Requirements Engineering

The requirements engineering process was carried focusing on the research objectives. It was divided into the following stages:

3.1.1 Requirement Input

The research objectives were determined by the client. The research objectives were to find a correlation between users' biometric data to their performance in a first-person shooter gaming scenario. The research objectives were used as the input for the requirements engineering process.

3.1.2 Requirement Analysis

The feasibility study was carried out to determine the resources that would be required to implement the research objectives and to determine the time frame for the research.

3.1.3 Requirement Specification

After analyzing the requirements, the final specifications were defined for the research. The final specifications were used as the input for the design process.

3.2 Analysis

The analysis phase involved examining the collected biometric data and gaming performance results to identify patterns and correlations. The analysis was carried out in the following stages:

3.2.1 Recruitment of Volunteers

The process of recruiting the volunteers was carried out in a way that ensured they satisfied the conditions of the research. The volunteers were required to be students of the Atlantic Technological University (ATU) Galway campus. The volunteers were also required to be over the age of 18 and have no known medical conditions that could affect their performance in the test. A Microsoft form was created and distributed to the students of the university to register their interest in the research. A poster 3.1 was created which contained information about the research and a QR code that linked to the Microsoft form.

Figure 3.1: Volunteers Recruitment Poster

The form can be found in the appendix 8. For a volunteer to be considered,

they were required to fill out the form and submit it. The form was designed to capture the following information from the volunteers:

- How often the participant practices exercises.
- The average duration of the exercises sessions.
- If the participant owns a activity monitoring device, if so which brand (Polar, Garmin, Apple, FitBit, or other).
- If the participant plays video games, if so which platform (PC, Xbox, Playstation, or other).
- Which game genre the participant enjoy the most. (First Person Shooter, Third Person Shooter, Soccer, Car Racing, or other).
- Participant should also sign the form as a consent to participate in the research.

3.2.2 Data Collection

For the purpose of the research, different categories of data was collected from the volunteers through the monitoring device to help achieve the stated goals. As personal data was being collected, an application to the Ethics Research and Ethics Committee was made to ensure that the data collection process was in compliance with the General Data Protection.

The Taught Programme Research Ethics Approval Application Form can be found on appendix 8. It was an extensive and iterative process with the Ethics Research and Ethics office to ensure that the data collection process was in compliance with the General Data Protection Regulation (GDPR). Unfortunately, it took longer than expected to get the approval from the Ethics Research and Ethics Committee, which lead to an unexpected delay in the data collection process. The data collection process was divided into two main categories:

Biometric Data

For the very first time processing a volunteer, some user information was needed to have them registered with the activity monitoring device and their data was saved to the manufacturer' repository. The Activity Monitoring Device used on this research was the Polar Vantage Smart Watch. The watch is capable of monitoring and capturing user's biometric data and saving the data to a repository where it can be accessed online via an API. The device is commercially available and accessible to the public. The rationale for using the Polar Vantage Watch is that

it has been widely used in both academic and industrial research projects and validation. Most importantly, these devices were available to us in such quantity that could satisfy our research needs.

Physical Information

The following physical information was collected from the volunteers, and stored with the manufacturer (Polar) which is GDPR compliant. The data was then accessed through the Polar API.

- Sleep Data: According to the manufacturer's manual [6], the watch can record both the quality and quantity of sleep, providing insights into the duration spent on the stages of sleep, such as REM Sleep, Deep Sleep and Light Sleep, Deep Sleep, and REM Sleep, along with the respective durations of each sleep stage. For the purpose of the research, an aggregated total of the various categories of sleep was recorded and used. Volunteers were instructed to wear the watch to sleep on the previous night before their scheduled test. Unit of measure used for the sleep data is minutes.

- Daily Activities: According to the Activity monitor manufacturers' manual [6], "Polar device uses an internal 3D accelerometer to record the wrist movements. It analyses the frequency, intensity and regularity of the movements together with the physical information." Calories, active calories, active steps and their respective durations are collected from through the device. For the purpose of the research, the active steps was used. Volunteers were expected to wear the device on the previous day for this data to be available. Unit of measure used for the daily activities is count of steps.

- Nightly Recharge: From the Activity monitor manufacturers' manual [6], the Nightly Recharge is recorded as follows: " is an overnight recovery measurement that shows how well your body has coped with overall stress you have experienced lately." The parameters, measured during roughly the first four hours of your sleep are heart rate, heart rate variability and breathing rate. For the purpose of the research, the following parameters and derived parameters was used:

  - Heart Rate Average (bpm)
  - Heart Rate Maximum (bpm)
  - Heart Rate Variability (HRV)

Test Data Collection

The test data is generated at the completion of the test session in the Unit Test Application. A test session consists of three categories of tests, which is undertaken sequentially in different stages. For quality assurance, the test was performed in a controlled environment using the same hardware in similar conditions over the course of the trials. For the purpose of the trial, a special designated room has been reserved for the data collection effort. The following metrics for the different tests was then captured and subsequently used for the research effort.

- Audio Test: designed to test users' audio reflexes.
- Audio Average Response Time: this is a measure of how quickly a user is able to react and determine the origin of a sound based on the audibility variance (in decibels) in a 3-Dimensional space.
- Visual Test: designed to test users' visual perception. At the end of this category of test, the following metrics were designed to measure users' performance.
- Visual Average Response Time: a measure of how quickly a user can identify and engage targets.
- Shot Accuracy: a measure of the percentage of successful shots to the number of targets spawned.
- Target Accuracy: a measure of the percentage of successful shots taken to a number of targets hit.
- Fine Motor Test: designed to test users' perception of depth and eye to hand coordination.
- Average Tracking Time: measure of average time it took for a player to successfully track, engage, and eliminate a target.
- Accuracy: measure of the percentage of shots fired to the number of targets hit.

Data Collection Procedure

The data collection process followed the below-listed steps to maximize throughput and minimize the average time spent processing each participant. The whole procedure can be found in the appendix 8.

For this study, participants were required to wear the Polar watch which is referred sometimes as "Activity Monitoring Device" in this text, a day prior to undertaking a gaming test.

Location:

The data collection effort was carried out at the Gym on the Atlantic Technological University (ATU) Galway campus.

Time:

The following weekly schedule was available for the volunteers at various times that best suit their personal schedules.

- Tuesdays: 12:00 - 13:00, 15:00 - 16:00
- Wednesdays: 12:00 - 13:00, 15:00 - 16:00
- Thursdays: 12:00 - 13:00, 15:00 - 16:00
- Fridays: 13:00 - 15:00

Results of the test were then paired with their biometric data for further analysis.

3.3 Design and Development

As a software development methodology for this project, many factors were considered. Initially, a comparison of the different software development methodologies was carried out to determine which would be the most suitable for the project. The comparison was based on the following factors:

- Continuous Stakeholder Involvement: The methodology should allow for continuous stakeholder involvement in the development process. Maintaining a continuous feedback loop with the stakeholders was essential for the project.
- Flexibility in Requirements: The methodology should allow for flexibility in the requirements. The requirements were not completely established at the start. The methodology should allow for changes to be made to the requirements as the project progresses.
- Iterative Development: The methodology should allow for iterative development. As the project was divided into different phases, each phase should be developed iteratively, allowing for the project to be developed in a series of small, manageable steps.

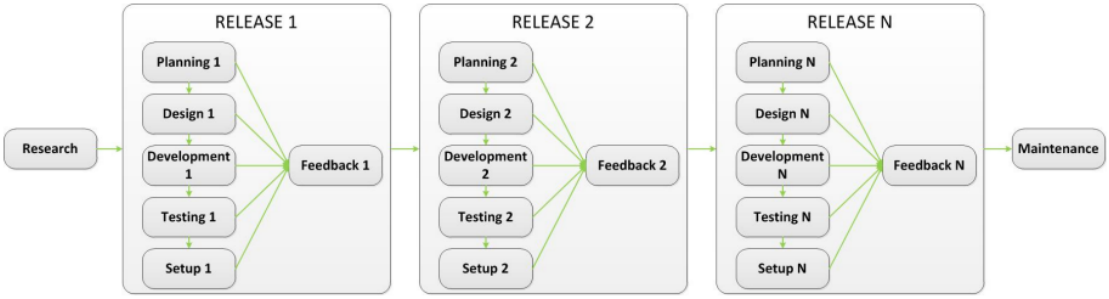


Figure 3.2: Extreme programming methodology [1]

After all the considerations, Agile Methodology [1] appeared as the preferred choice over Waterfall, as it allows for continuous stakeholder involvement, flexibility in requirements and iterative development. Unlike the rigid Waterfall methodology [1], Agile allows for changes to be made to the requirements as the project progresses. The Agile methodology also allows for the project to be developed in a series of small, manageable steps. This iterative nature enabled continuous refinement and adjustment, ensuring that the stakeholders were satisfied with the progress.

Agile - Extreme Programming (XP)

Extreme Programming (XP) is an Agile framework that moves from the traditional software development methodologies by breaking down the development process into smaller and more manageable increments. Instead of extensive planning, analysis, and design upfront, XP focuses on continuous and iterative development. To minimize errors and defects, XP places a strong emphasis on testing. This test-driven development approach ensures that the code is always working and that any defects are caught early in the development process. XP also places a strong emphasis on customer satisfaction. The customer is involved in the development process and is able to provide feedback on the project as it progresses. [1] In general, Extreme Programming presents a flexible and iterative method for software development that aligns well with the requirements of this project.

Figure 3.2 shows the Extreme Programming methodology. The methodology is divided into four main phases: Planning, Designing, Coding, and Testing. Each phase is developed iteratively, allowing for the project to be developed in a series of small, manageable steps.

3.3.1 Testing

As part of testing the system, the following steps were taken to ensure that the system met the stakeholders' requirements:

- Mocha Testing: this test was carried out to ensure that the API endpoints were working correctly and that the data was being retrieved correctly.
- Integration Testing: this test was responsible for testing how the different components of the system worked together.
- System Testing: this test was carried out to ensure that the entire system worked as expected and that it met the stakeholders' requirements.

3.3.2 Deployment

After the system was accepted by the stakeholders, it was deployed to the server where it could be accessed.

3.3.3 Maintenance and Update

Following deployment, the system was maintained and updated as needed to ensure that it continued to meet the stakeholders' requirements. Regular updates were made to the system to ensure that it remained up-to-date and that it continued to meet the stakeholders' needs.

3.4 Meetings

In this section we will discuss the different types of meetings that were held throughout the project. The meetings were held to ensure that the project was on track and that the stakeholders were satisfied with the progress.

3.4.1 Supervisor Meetings

Weekly meeting was setup with the supervisor to discuss the progress of the project. The meetings were held in the supervisor's office on campus every Tuesday at 12 pm throughout the duration of the project. The meetings were used to discuss:

- The progress of the project.
- Any issues or challenges that were encountered.

- Any changes that needed to be made to the project.
- Any feedback that the supervisor had on the project.

These meetings were important, especially in the early stages of the project, due to various challenges that were encountered. The feedback from the supervisor was invaluable in helping to overcome these challenges.

3.4.2 Client Meetings

At the beginning of the year we were offered the opportunity to work with a client on a project. In our meeting with the client, we were informed that the project was a research project that aimed to find a correlation between users' biometric data to their performance in a first-person shooter gaming scenario. The aim was to quantify how biometric such as Heart Rate Variation, Heart Rate, active steps taken, quality and quantity of sleep, etc. affect select gaming skills like eye-to-hand coordination, fine-motor skills and reaction time. The client also informed us that the project was a continuation of a previous project and that we would be inheriting an existing infrastructure. For the collection of the biometric data, the client informed us that there was four Polar Vantage Smart Watches available for the project. The meetings were also held to get the devices registered on the Polar API and to get the volunteers enrolled in the project.

The meetings were held online using Teams and also in the client's office in the ATU Gym at Unit D, Racecourse Business Park, Ballybrit, Galway.

3.5 Development Tools

The selection of development tools for this project was determined by their capacity to meet the project's requirements. The tools were selected and used in a way we could have a continuous feedback loop with the stakeholders.

The tools used in the project are as follows:

3.5.1 Postman

Postman was used to test the API endpoints that were used to access the dataset from the Firebase Realtime Database. Postman is a popular API testing tool that allowed us to test the API endpoints and ensure that the data was being retrieved correctly. It was also important when testing the security of the end points, as we were able to test the different authentication tokens that were used to access the data.

3.5.2 Visual Studio Code

Visual Studio Code was used as the primary code editor for the project, it was also used for this own dissertation as it also supports LaTeX. Visual Studio Code is a widely used code editor in the software development industry, it was also used to write the code for the project and to test the code locally before deploying it to the server.

3.5.3 Teams

Microsoft Teams was used as the primary communication tool for the project. Teams was used to communicate with the supervisor, the client, and the volunteers and to schedule meetings, share documents, and communicate with the stakeholders. Teams was also used to communicate with the volunteers and to provide them with updates on the project.

3.5.4 Draw.io

The diagrams shown in this project were designed using Draw.io, a widely recognized tool for diagramming. This platform facilitates the creation of various diagrams, such as flowcharts, UML diagrams, and network diagrams, accommodating all the project diagramming needs.

The diagrams were used to help visualize the project architecture and to communicate the design of the project to the stakeholders.

3.6 Version Control

Git is well known for its distributed version control system. It is not only open-source but also free to use. Git was used as the version control system for the project. Git allowed us to track changes to the code and to collaborate with each other on the project. While developing the project locally on our machines, we were able to push the code to the remote repository on GitHub and later pull the code to the server for deployment, this allowed us to work on the project from different locations and to collaborate with each other on the project.

3.7 Project Management

Project management was an important aspect of the project. The project was divided into different phases, each with its own set of tasks and deliverables. The project was managed using the Agile methodology, which allowed for continuous

stakeholder involvement, flexibility in requirements, and iterative development. Jira was used as the project management tool for the project. Jira allowed us to create tasks, assign tasks to team members, and track the progress of the project. Jira was also used to create sprints, which allowed us to break down the project into smaller, more manageable steps. Sprints were used allowing us to develop the project in a series of small, manageable steps.

Chapter 4

Technology Review

This section will provide an overview of the technologies chosen for the project. It was chosen based on the project requirements. The technologies used in the project are as follows:

4.1 IAAS - Infrastructure as a Service

IAAS is a cloud computing service that provides virtualized computing resources over the internet. It is one of the three main categories of cloud computing services, alongside Platform as a Service (PAAS) and Software as a Service (SAAS). IAAS provides virtualized computing resources, such as virtual machines, storage, and networking, allowing users to build and manage their own IT infrastructure in the cloud.

4.1.1 Amazon Web Services (AWS)

For this project, AWS EC2 was chosen as the IAAS provider. AWS EC2 is a web service that provides resizable compute capacity in the cloud. It allows users to launch virtual servers, known as instances, on the AWS cloud. EC2 instances can be configured with different types of CPU, memory, storage, and networking capacity, providing users with the flexibility to choose the resources that best suit their needs. /citeaws

The table below shows the configurations of the EC2 instances available on AWS[7]:

Instance Type	Description
t2.micro	General-purpose instance type that provides a balance of compute, memory, and network resources.
t2.medium	Provides a higher level of compute, memory, and network resources compared to the t2.micro instance type.
t2.large	Offers a higher level of compute, memory, and network resources compared to the t2.medium instance type.
t2.xlarge	Provides a higher level of compute, memory, and network resources compared to the t2.large instance type.

Table 4.1: AWS EC2 Instance Types

For this project, the t2.micro instance type was chosen as it provides a balance of compute, memory, and network resources, making it suitable for the project's requirements. The t2.micro instance type is also cost-effective, making it an ideal choice for the project's budget constraints.

The complete system architecture for this project is explained in detail in the System Architecture chapter.

4.2 Executive Dashboard

When it came to developing the Executive Dashboard for data analysis it was crucial to select the tools that could transform data into easily comprehensible information. After conducting research and exploring alternatives Angular and Chart.js were chosen for specific reasons.

4.2.1 Angular

Angular is based on TypeScript and often used for building single-page applications (SPAs) or enterprise-level solutions. It is a powerful tool for building client-side web applications.

Main Concepts: Components, Modules and Services

Angular is composed of three foundational blocks: Components, Modules and services.

Components

Components are like Lego blocks of the application. Each component holds a portion of the user interface and its behaviour. Components serve as the bridge

between the application data and what the user experiences on the screen.[8]

Modules

In Angular modules serve as containers that group related components, directives, and services together that can be combined with other modules. It plays an important role in improving maintainability and re-usability, key concepts of Angular development.[9]

Services

Angular services use typescript classes with injectable decorators. The decorator informs Angular that the class function as a service and can be injected into other components that need that service.[10]

Angular architecture

The Model-View-Controller (MVC) has as its components the model, the view, and the controller. With controller orchestrating the communication and interactions between the model and the view.

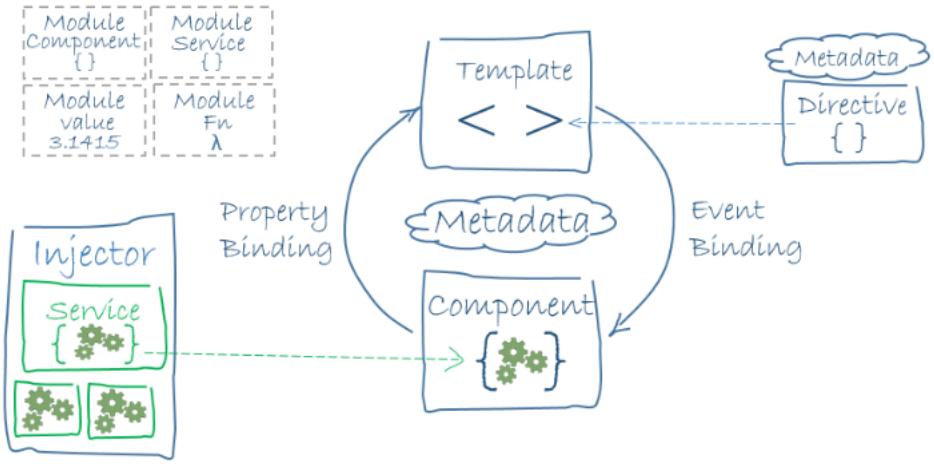


Figure 4.1: Angular application architecture

Advantages of using Angular

Component-Based Architecture

As mentioned earlier in discussions Angular organizes its functionalities into components. These components have the ability to communicate with each other enabling updates to sections without affecting the rest of the application.

Mobile-Friendly Approach

Angular incorporates techniques such, as lazy-loading, which means loading parts of the application (like images) only when they are needed. This ensures that users do not experience long waiting times.

Two-Way Data Binding

Data can be synchronized between the model and the view. The two-way data binding feature in Angular ensures that any changes made to the model are reflected in the view and vice versa.

Asynchronous Programming

By utilizing programming executes code in a non-sequential manner and employs multi-threading to enhance performance. This speeds up operations and prevents system freezes, providing users with a seamless experience.

Single-Page Applications

Angular creates a dynamic single-page application which can be navigated without page reloads, improving the user experience with better user interaction and engagement.

Code Re-usability

The component-based architecture of Angular promotes the re-usability of UI components saving development time.

Dependency Injection

With dependency injection in place, Angular allows for the creation of objects that rely on other objects. This improves modularity and efficiency, within the app.

Angular Material

Angular's documentation offers a range of built user interface components and modules that adhere to Google's Material Design principles. This greatly facilitates the developer's work, simplifying the design process and enabling application development.

Angular CLI

Angular command line interface gives the developer the ability to generate Angular projects, modules, services, and components with a single command,

which helps reduce configuration errors and gives the developer the freedom to dive into creative aspects of the project, focusing on innovation and functionality rather than getting bogged down by initial setup complexities.

4.2.2 Embedded JavaScript (EJS)

EJS is a straightforward templating language that enables the creation of HTML markup using pure JavaScript. It provides a rendering mechanism for dynamically generating HTML content.[11] Its compatibility with Express.js allows for it to render data from the server-side to the client-side, making it an ideal choice for web application development. EJS also allows to reuse of code snippets, increasing the efficiency of web application development. It also allows for the implementation of conditional logic, enabling the rendering of different content based on specific conditions. EJS supports looping constructs, facilitating the iteration over data collections and the generation of repetitive HTML elements. It provides error handling capabilities, ensuring that errors are caught and managed effectively during the rendering process. EJS is highly extensible, allowing for the integration of additional features and functionalities to enhance the rendering process. [11]

Feature	Description
Simple Syntax	EJS uses a simple syntax that allows for the embedding of JavaScript code within HTML markup.
Dynamic Content	EJS enables the dynamic generation of HTML content.
Code Reusability	EJS supports the reuse of code snippets, enhancing the efficiency of web application development.
Conditional Logic	EJS allows for the implementation of conditional logic, enabling the rendering of different content based on specific conditions.
Looping Constructs	EJS supports looping constructs, facilitating the iteration over data collections and the generation of repetitive HTML elements.
Error Handling	EJS provides error handling capabilities, ensuring that errors are caught and managed effectively during the rendering process.
Extensibility	EJS is highly extensible, allowing for the integration of additional features and functionalities to enhance the rendering process.

Table 4.2: EJS Features

4.2.3 AJAX

AJAX stands for Asynchronous JavaScript and XML. It is not a programming language, but a technique used in web development to create interactive web applications. It allows the page to be loaded asynchronously by exchanging data with the server behind the scenes. This means that it is possible to update parts of a web page without reloading the entire page.

Figure 4.2 shows the process of AJAX.

- 1. An event is triggered in a web page (the page is loaded, a button is clicked, etc.).
- 2. An XMLHttpRequest object is created by JavaScript.
- 3. The XMLHttpRequest object sends a request to a web server.
- 4. The server processes the request.
- 5. The server sends a response back to the web page.
- 6. The response is read by JavaScript.
- 7. The web page is updated.

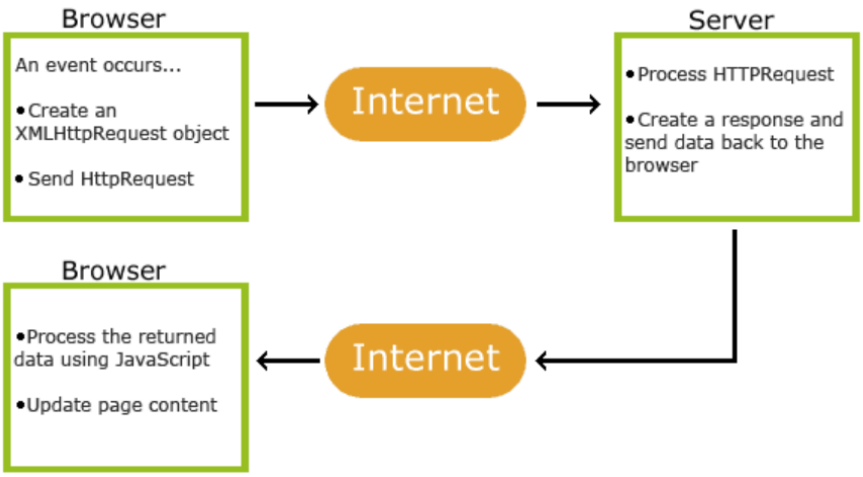


Figure 4.2: AJAX

4.3 Data Visualization

The visual representation of data is a important aspect of the project. It is used to communicate information clearly and efficiently. It allows to identify patterns, and find correlations that would be hard to identify in raw data, standing as an essential step in the data analysis process.

4.3.1 Chart.js

Chart.js is an efficient JavaScript library, it is not only open-source but also offers capabilities where attractive charts and graphs can be generated, and later embedded into web pages. Chart.js became very popular due to its simplicity and ease of use, making it an excellent choice for data visualization tasks. Utilizing HTML5 canvas technology, it ensures that charts are responsive and adjust seamlessly to the container's dimensions. Additionally, it boasts compatibility across all contemporary web browsers and enjoys support from an extensive developer community. Chart.js acts as an artist's palette for developers, equipped with a versatile API that supports an array of chart types, including line, radar, and pie charts, among others. Moreover, it provides extensive customization possibilities, enabling the creation of distinct and eye-catching charts.[12]

Table **??** is showing some of the features of Chart.js[12]:

Feature	Description
Easy to use	Aesthetically charts can be created without the need of extensive configuration. The library follows a declarative approach allowing the developer to define the data and settings of the chart in a single object.
Responsive	Charts generated by Chart.js are responsive, adapting to different devices, and makes sure that the visualization is readable across various platforms.
Customization	Chart.js provides a high degree of customization. Colors, fonts, and other visual elements can be changed creating unique charts.
Interactivity	Chart.js has great support tooltips and animation, enabling the user to explore the data points, adding a layer of engagement to the project.
Cross-browser compatibility	Chart.js is compatible with all most of the browsers.

Table 4.3: Chart.js Features

Chart.js stands as a dependable and powerful solution for visualizing data, with simple and intuitive APIs that facilitate the creation of attractive and informative charts. Tt also benefits from the strong support of an extensive developer community, offering a significant advantage.

4.4 AI module - Data analysis

An amazing fact is that ninety percent of the data generated in the world was generated in the last two years. In the early 2000s, the amount of data being generated exploded exponentially on the same rates as the internet and social media usage. Organizations found themselves facing a massive volume of data that was very hard to process. Then the concept of Big data was created to describe this large volume of data. It refers to data that is so large and complex that traditional methods of data processing are not sufficient to process it. [13]

4.4.1 Scikit-learn

Scikit-learn, also know as sklearn, is an open-source machine-learning library for Python. It offers a range of regression and classification algorithms, such as SVM,

Random Forest, Gradient Boosting, and K-means. Its architecture allows it to be integrated with other Python libraries, such as NumPy, providing comprehensive toolkit for machine learning. [14]

- **Simple and Efficient Tools**: Scikit-learn offers a simple solution for data mining and data analysis.
- **Consistent Interface**: It provides an interface for the user that is consistent across different algorithms, so it is easy to switch between the models.
- **Integration**: Scikit-learn provides a wide range of algorithms and tools for machine learning which integrates seamlessly with other Python libraries.
- **Extensive Documentation**: With a large available community, the library has a huge documentation available, making it easy to find help and resources.

Scikit-learn was used on the AI model to train and evaluate machine learning models. As our dataset was in a structured format, Scikit-learn was the ideal choice for the project, providing algorithms that can be used to work with structured data. It was a regression problem, and for that, Linear Regression, Support Vector Machine, Random Forest, and K-Nearest Neighbors algorithm from this library were used to evaluate the best model for the dataset.

4.4.2 TensorFlow

TensorFlow was developed by the Google Brain team and has been widely adopted in the machine learning community. It mainly used for deep learning tasks, such as neural networks, and machine learning that requires a large amount of data. It is not only powerful but also flexible, allowing for the creation of custom machine learning models. It is also highly scalable, allowing for the training of models on multiple CPUs, GPUs, and TPUs (Tensor Processing Units).[15]

- **Flexible Architecture**:
- TensorFlow provides developers with the capability to execute computations across multiple CPUs or GPUs using a unified API, whether on desktops, servers, or mobile devices, thereby making it versatile for various applications.
- **Comprehensive Library**:
- This platform offers libraries and community support, that enables the developer to create and implement applications powered by ML.

- **High-Level APIs**: With high-level APIs like Keras, TensorFlow is very accessible to any developer, simplifying the build of the models. It also provides a low-level API for more advanced users, allowing for greater flexibility and customization.
- **Visualization with TensorBoard**: As part of TensorFlow, there is a tool called TensorBoard, which provides many visualization tools that are great to aid the understanding of the model, debugging, and optimization of complex neural networks.
- **Scalability**: It has the capability of running computation on multiple CPUs or GPUs, which makes it perfect for big machine learning tasks.
- **Large Community and Support**: Having a vast community, TensorFlow benefits from a plethora of tutorials, documentation, and active community support which aids in solving problems and improving the framework.

Keras

Keras is a high-level neural networks API, written in Python, that runs on top of TensorFlow. It is designed to be user-friendly, modular, and extensible, allowing for rapid prototyping of deep learning models.[16]

- **Modularity and Composability**: Keras models are assembled by connecting configurable building blocks together, with few restrictions. This modularity enables fast experimentation and research, allowing researchers and developers to build and test new ideas quickly.
- **Pre-built Layers and Models**: Keras comes with numerous pre-built and pre-trained models, such as neural networks, convolutional networks, and recurrent networks. The time and effort required to build and train new models from scratch are significantly reduced.
- **Support for Multiple Backends**: Originally designed to run on top of TensorFlow, Keras also supports other backends, such as Microsoft Cognitive Toolkit (CNTK) and Theano.
- **Seamless Integration with TensorFlow Features**: Being fully integrated into TensorFlow (‘tf.keras‘), Keras provides full support for TensorFlow functionality, including TensorFlow’s eager execution, ‘tf.data‘ pipelines, and ‘tf.distribute‘ for multi-GPU training. This integration ensures that Keras models can leverage the power and scalability of TensorFlow without compromise.

- **Broad Adoption and Community Support**: Keras enjoys wide adoption in both academia and industry, making it one of the standard APIs for developing deep learning models. This popularity ensures a large community for support, sharing, and collaboration.

Keras simplify the process of creating the neural network structure, as well as training and evaluating the model.

Figure ?? shows the architecture of TensorFlow and Keras. TensorFlow is the core library for numerical computation, while Keras serves as a high-level API that runs on top of TensorFlow.

4.5 Backend Technologies

4.5.1 Node JS

NodeJS enables the execution of JavaScript code outside the limitations of a web browser. It is cross-platform JavaScript runtime that allows for the creation of fast and scalable network applications. Because of its architecture, that is asynchronous and event-driven, it handles concurrent operations efficiently, preventing the halt of other processes. It suits real-time applications that are data-heavy and operate on a distributed network of devices. [17]

Advantages of Node.js

NodeJS is designed as an asynchronous, event-driven JavaScript runtime ideal for crafting scalable network applications. Its efficiency and lightweight nature make it an excellent option for data-heavy, real-time applications.[18] A significant benefit of NodeJS is its use of a single-threaded event loop, enabling it to manage multiple operations concurrently without hindering the execution of additional tasks.

A key feature of NodeJS is its extensive library ecosystem. The Node Package Manager (NPM) hosts over a million packages, making it the world's largest collection of open-source libraries.[19] This abundance of resources significantly streamlines the development process, allowing developers to leverage existing modules and focus on innovation.

NodeJS also benefits from its single programming language across both the server and client sides. This allows for the sharing of code and data between the server and the client, which is a great advantage, reducing the learning curve associated with NodeJS development. [18] Furthermore, NodeJS's uniform programming language for both server and client sides simplifies code and data sharing across them, offering a streamlined learning process for development. In terms of

performance, NodeJS excels due to its foundation on Chrome's V8 JavaScript engine, a high-performance engine that ensures speed and efficiency, especially in scenarios requiring real-time data processing. [18]

Disadvantages of Node.js

NodeJS is not without its criticisms. Critics often point out the callback hell, a situation where the code becomes unreadable due to the excessive use of callbacks. Although it has been largely mitigated by the introduction of Promises and async/await syntax, it remains a valid criticism. [20]

Table **??** is summarizing the advantages and disadvantages of Node.js[18]:

Advantages of Node.js	**Disadvantages of Node.js**
Non-blocking and event-driven: Allows for scalabilityand concurrency, making it ideal for real-time applications.	Single-threaded: it is single-threaded, and can lead to blocking if not handled correctly, impacting CPU-bound tasks.
JavaScript as a single language: It can be used for the server-side and client-side development, reducing context switching.	Callback hell: the code could be hard to read and maintain as there are nested callbacks (Promises and async/await can minimize this problem).
Large and active community: The Node Package Manager makes thousands of open-source libraries and modules available.	Limited support for multi-core processors: Node.js does not fully utilize multi-core CPUs out of the box.
Speed: it is known by its high-performance.	Not weel suited for CPU-intensive tasks: The event-driven and single-threaded characteristics.
Lightweight and fast startup: Node.js applications typically have lower memory consumption and quicker startup times.	Maturity and stability: Some developers argue that Node.js, compared to more established platforms, may have less maturity and stability in certain use cases.

Table 4.4: Advantages and Disadvantages of Node.js

4.5.2 PM2

PM2, or Process Manager 2, is a process manager for Node.js applications. With a rich feature set, including monitoring, load balancing, and error handling, its

primary function is to keep applications alive forever, restart them without downtime, and simplify common system administration tasks. The standout feature of PM2 is its ability to keep processes running in the background indefinitely. This is particular important for web applications that requires constant availability. PM2 automatically resurrects crashed applications, ensuring that system glitches do not result in prolonged downtime.[21] This automatic restart capability is a safeguard against potentially costly application crashes that could otherwise lead to a poor user experience or even lost revenue. [21] PM2 also provides a built-in load balancer that can distribute incoming requests across multiple instances of the application. This improves performance and scalability, allowing the application to handle more requests, evenly distributing traffic across the instances, which can be particularly beneficial when running on multi-core systems. [22] However, PM2 is not one-size-fits-all solution. It is designed for NodeJS applications, and is not suitable for applications written in other languages. It is also not suitable for applications that require a high degree of customization. [21] Despite the considerations, PM2's benefits have outweighed its limitations, and it was chosen as the technology for the project. It provides many features, including monitoring, load balancing, and error handling. Its installation and setup process is straightforward, requiring minimal configuration. All these features make PM2 an ideal choice for the project, as it provides the necessary tools to manage and monitor the NodeJS application in a production environment.[18]

4.5.3 Docker

Docker is a tool that enables the creation, deployment, and running of applications in containers. Comparing with virtual machines (VMs), containers are lightweight, portable, and efficient, making them an ideal solution for deploying applications across different environments, having the application and its dependencies packaged into a single container. The containerization process encapsulates the application's code, the libraries, and the dependencies into a single object, ensuring that the application runs reliably and consistently across different environments. [23] Due to its simplicity and portability, Docker has become the de facto standard for containerization. Its lightweight nature compared to traditional virtual machines makes it ideal for cloud computing. The memory footprint is significantly reduced, and the performance is improved, making it an ideal solution for deploying applications in the cloud. [23]

However, Docker's container model is not without challenges. In the context of this project, the average size of a single Docker image exceeding 400MB presents a concern regarding bandwidth utilization. Continuous build and deployment processes involving large container images can consume substantial network resources, leading to potential cost overruns. [23] Another challenge is the security of Docker

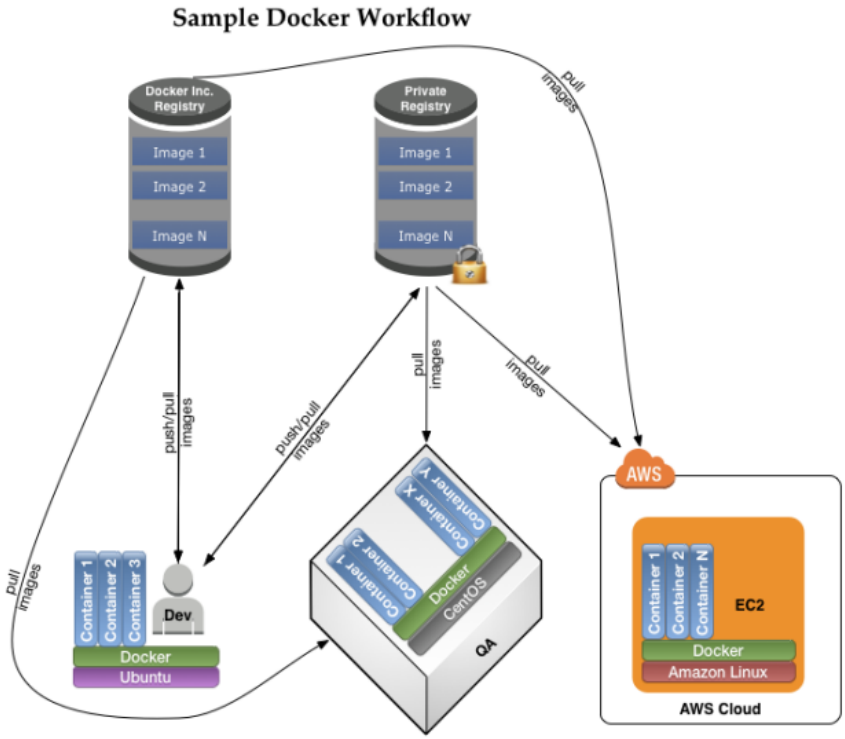


Figure 4.4: Development workflow with Docker

containers. Docker containers share the same kernel, which means that a vulnerability in the kernel can affect all containers. [24] While Docker's layering and image caching mechanisms can mitigate some of the network overhead, it is still a concern. In conclusion, Docker is a powerful tool that offers numerous benefits for application deployment, from development to production. However, it is not without its challenges. The large size of Docker images can lead to bandwidth utilization issues, and the shared kernel model can lead to security concerns. Considering the scope of the project, time constraints and budget, Docker wasn't chosen as the technology for the project as the cost would be too high.

4.5.4 NGINX

NGINX works as a reverse proxy server, load balancer, and HTTP cache. It is a web server able to handle high volumes of concurrent connections, which makes it an ideal solution for applications that require high availability and scalability. It is also lightweight and efficient and highly configurable, allowing for the customization of its behaviour to suit the needs of the application.

It was created in 2004, with the goal of solving the C10K problem, refering

to the challenge of handling 10,000 concurrent connections. [25] Due to its vast capabilities, and the ability to handle high volumes of concurrent connections, NGINX has become one of the most widely used web servers in the world, often used as a load balancer and as a reverse proxy in addition to its role as a web server.

Table **??** is showing some of the features of NGINX[25]:

Aspect	**Description**
Type	Web server software
Primary Use	Serving web content, load balancing, reverse proxy, and more
Performance	Highly efficient in handling high concurrency with low memory footprint
Architecture	Event-driven, asynchronous, and non-blocking which contributes to its ability to handle a large number of simultaneous connections easily
Scalability	Scalable to support growth in traffic and applications
Security Features	Offers robust security features including rate limiting, client request filtering, and SSL/TLS termination
Flexibility	Highly configurable for many type of servers
Open Source/Commercial	Available in both open-source and commercial versions (NGINX Plus)

Table 4.5: NGINX Features

Due to its vast capabilities NGNIX was chosen for our project as reverse proxy and load balancer. It is an ideal solution for applications that require high availability and scalability with a high number of concurrent connections. It is also highly configurable, allowing for the customization of its behaviour to suit the needs of the application.

4.5.5 MySQL

MySQL uses structured query language (SQL) for database access, providing high performance and reliability. It interacts with the database using SQL, which is a standard language for accessing and managing databases.

Table **??** is showing some of the features of MySQL[26]:

Aspect	Description
Data Handling	Efficiently manages large datasets, crucial for training machine learning models
Query Performance	Fast query execution, beneficial for data retrieval and preprocessing in machine learning workflows
Scalability	Easily scales with data volume and complexity, supporting the growing needs of machine learning applications
ACID Compliance	Ensures data integrity and consistency, vital for the accuracy of machine learning outputs
Advanced Analytics	Supports SQL extensions for advanced analytics, facilitating machine learning data processing tasks
Data Storage Options	Offers various storage engines, allowing optimization based on the specific needs of machine learning models
Integration Capabilities	Easily integrates with popular machine learning frameworks and languages, streamlining the development process

Table 4.6: MySQL

In the context of this project, the dataset was originally in a JSON format in FireStore which were found to be of a high complexity while performing compound querying. To facilitate querying and the difficulties of FireStore, MySQL was chosen to be the database for the project, where all the data for analysis is sourced directly from it, having a replica of the database from the Legacy System.
By regularizing the data in MySQL, a more standardized and organized dataset was achieved. This standardization significantly simplified the data preprocessing steps in the machine learning workflow.

4.5.6 Express JS

Express.js, often simply called Express, is a streamlined and adaptable framework for Node.js designed to enhance web and mobile application development. It comes equipped with tools and features to suit every developers needs, while creating sophisticated applications. It is often used to build server-side applications and APIs, due to its simplicity, performance, and scalability.In this project, Express.js was chosen for several strategic reasons, which are outlined below.[27]

Feature	Advantage for Our Project
Minimalist Framework	Express.js provides essential web application features without dictating any specific architecture, allowing for flexibility and customization in our project.
Middleware Support	The use of middleware modules enables us to extend the functionality of our application easily and efficiently.
Routing System	Its powerful routing system helps manage requests and responses effectively, a crucial aspect for our project's RESTful API design.
High Performance	Known for its high performance, Express.js enhances the responsiveness and speed of our web application.
Community Support	As it is a very popular framework, it has strong community support, with many resources available.
Easy Integration	Express.js seamlessly integrates with other technologies and databases, which is vital for the diverse tech stack of our project.
Simplicity	Its simplicity and ease of use accelerate development and reduce the learning curve for new team members.

Table 4.7: Express.js Features

It was chosen as a back end web application for this project due to its simplicity, performance, and scalability. Being widely used for building APIs and server-side applications, it offers a great advantage.

4.6 Testing Technologies

4.6.1 MochaJS

Mocha is a JavaScript testing framework that runs in the browser and on Node.js. It executes sequentially, allowing for flexible and accurate testing of code. It accurately associates test results with the code that was tested, making it easier to identify and resolve issues. This testing framework has numerous features, offering functionalities such as asynchronous test execution, test coverage reports, and the ability to run tests in parallel. Mocha is highly extensible, allowing for the integration of additional libraries and plugins to enhance the testing process. It supports a variety of assertion libraries, giving the developer the option to choose the one that best suits.

Mocha is widely used in the industry and is widely supported by the developer community. [28]

Feature	**Advantage for Our Project**
Asynchronous Testing	Mocha's support for asynchronous testing was invaluable for testing our application's asynchronous operations, ensuring accurate and efficient tests.
Flexible Reporting	Its variety of reporting options allowed us to choose the best format for our testing outcomes, enhancing readability and understanding of test results.
Rich Assertion Library	Integration with various assertion libraries gave us the flexibility to write test cases in a style that best suited our project's needs.
Before/After Hooks	Mocha's before and after hooks simplified the setup and teardown processes for our test suites, making our tests cleaner and more reliable.
Wide Adoption	Mocha's popularity and wide adoption provided us with extensive resources for learning and troubleshooting, contributing to a smoother testing process.
Easy Integration	Its seamless integration with other tools and libraries in our project's ecosystem allowed for a more streamlined development and testing workflow.
Customizable and Extensible	Mocha's customizable and extensible nature enabled us to adapt the testing to meet our project's specific requirements, enhancing test efficiency and effectiveness.

Table 4.8: Mocha Features and Advantages

Mocha was selected for the testing of our project due to its flexibility, ease of use, and vast number of features suited for both synchronous and asynchronous tests.

Throughout the development cycle, Mocha facilitated a robust testing environment that improved code quality and reliability. By utilizing Mocha's features, we were able to implement testing that covered a range of functions, allowing to identify and resolve issues early in the development process. Also, the ability to

use different assertion libraries enabled our team to write test cases in a style that best suited our project's needs.

Mocha has proven to be an invaluable asset to our project, significantly enhancing our testing practices. Its adaptability, and vast documentation available online contributed to a higher quality final product.

4.7 Security

Security is a critical aspect of any application, especially when dealing with sensitive data. In the context of this project, security measures were implemented to protect the application from potential threats and vulnerabilities. Among the various security measures available, JSON Web Tokens (JWT) was chosen for securely transmitting information between parties as JSON objects.

4.7.1 JSON Web Tokens - JWT

JWTs play a crucial role in authentication and authorization processes. A key advantage of JWTs lies in their capability to securely and efficiently transmit information between parties as JSON objects. The structure of a JWT comprises three fundamental components: a header, a payload, and a signature. The header provides details about the token type and the algorithm used to sign it. The payload contains claims, which are statements about an entity along with additional data. Lastly, the signature is utilized for verifying the authencity of the JWT sender and ensuring the integrity of the message throughout its transmission. [29].

Utilizing the Token: Once the client possesses the token, it can be included in the header of the HTTP request to the server, which can then verify the token and grant access to the requested resource if the token is valid.

Assessing Token Validity: Receiving a token does not guarantee its validity. The token is verified by the server to make sure that it has not been tampered with and that it is still valid. This process involves checking for token signature and the payload claims.

This is done by checking the signature of the token and verifying the claims in the payload.

Decision on Access: If the token is valid, the server can grant access to the requested resource. If the token is invalid, the server can deny access.

Token Expiration: Expiration time can be set for JWTs, after which they are no longer valid. This helps to prevent the misuse of tokens and enhances security.

4.8 Difficulties

The project faced a few challenges during the development process. The main one that affectted mostly the front end development was the deployment constraints on the Amazon virtual machine. As discussed earlier in this section, the instance configurations used for this project, the t2.micro, was not able to handle the load of the application, which resulted in slow performance and frequent crashes. This led to the decision to switch to a more lightweight technology stack, which included EJS, Express, CORS, and AXIOS. This change improved the performance of the application and resolved the deployment issues. The initial technology stack with Angular, Chart.js, Express JS, and Node JS couldn't be used due to the constraints discussed above.

Chapter 5

System Design

This chapter presents the overall architecture and design of the system. This chapter introduces and explains the various design paradigms used for the development and deployment of various components that interact together to provide the functionalities of the system.

The system architecture calls for a relational database to store aggregated users test and biometric data, a RESTful API to provide secured and customized access to user data in the database, and a web application to provide a user interface for administrative access to the system. Machine learning services were developed and deployed to provide real-time machine learning data analysis and best model selection algorithms.

The various components interact together to provide the system functionalities and were developed and deployed in a loosely coupled manner and interact with each other through well defined RESTful API calls. The Amazon Web Services (AWS) cloud platform was utilized as a Infrastructure as a Service (IaaS) to provide the compute and storage resources required to run the various components with the EC2 Micro Virtual Machine, domain name service was provided using Amazon Route 53 (Domain Naming Service) and Amazon Certificate Manager for TSL/SSL security. Relational Database was provided as a microservice using the Amazon Relational Database Service (MySQL)

5.1 System Architecture

The Overall system was designed for a public cloud deployment while the core functionalities of the system was modelled using the Model-View-Controller (MVC) Software Architectural Pattern. Thus, the system was divided into three major components: the Model, comprising of the databases and the machine learning models; the View, implemented with Dashboard web application; and the Con-

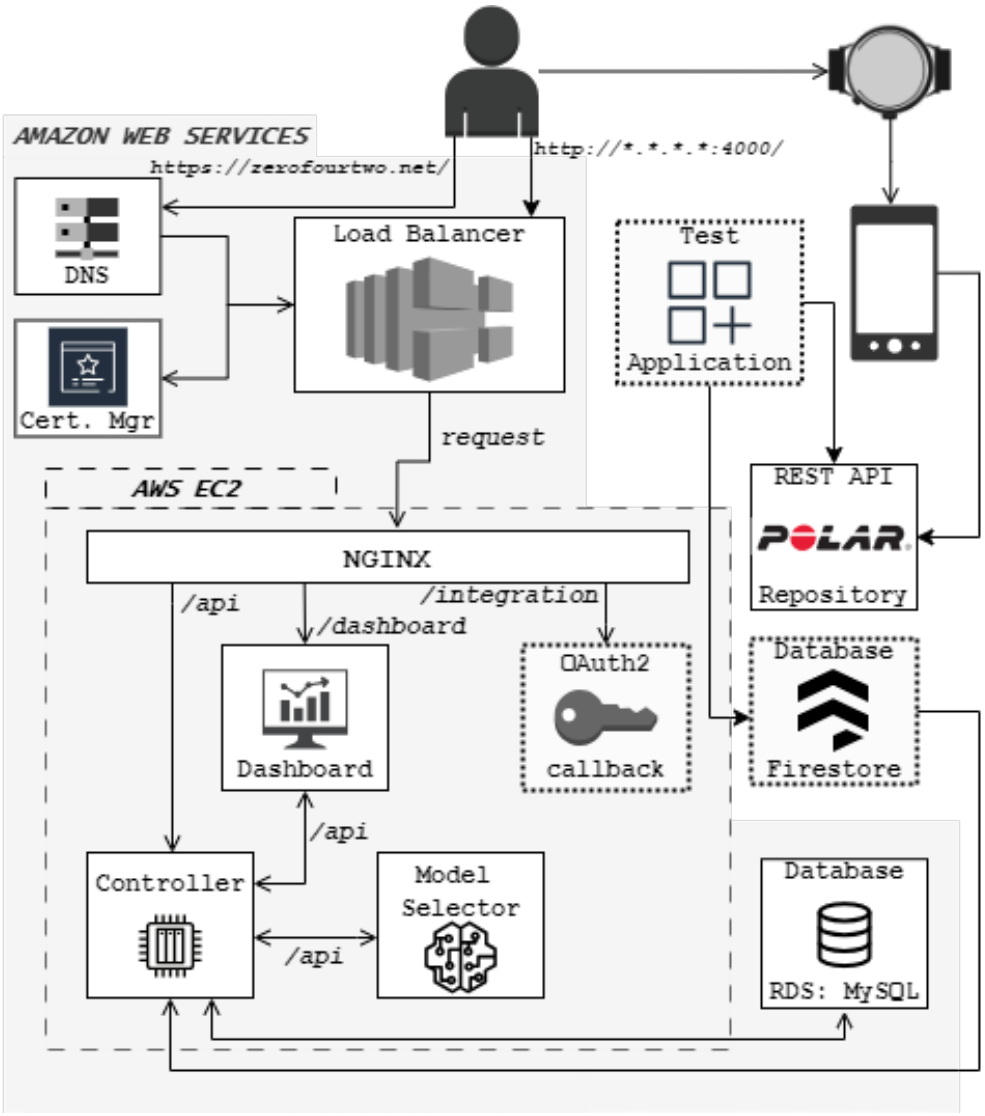


Figure 5.1: System Architecture.

troller, implemented with the Controller web application. Each of these components were developed and deployed as separate services.

5.2 Routing & Security

Communication between various components was modelled with the client-server paradigm utilizing REST API over HTTP. Communication may be initiated by a client and information exchanged with a server using a request-response messaging

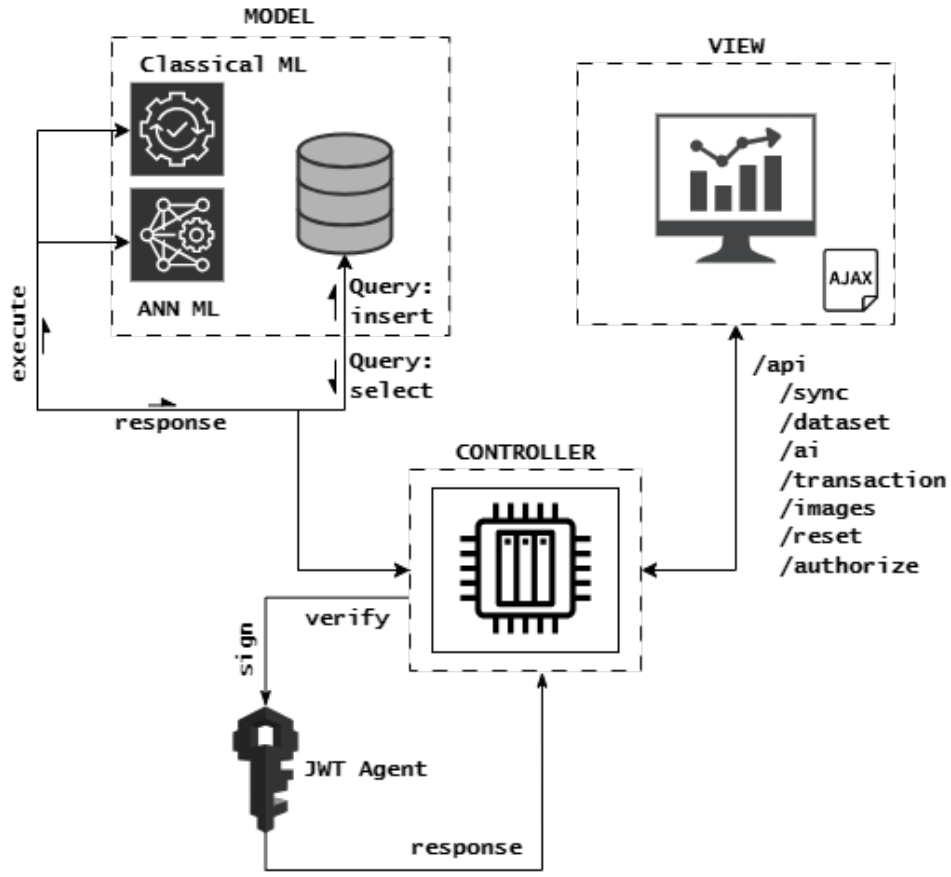


Figure 5.2: MVC Architecture.

pattern. Security was provided to secure access to the system and data. At the network level, the TSL/SSL security layer was added to the domain name to provide end-to-end encryption between the user and the web application. At the application api level, the Bearer Authentication Token was used to secure access to the REST API. Figure 5.3 shows a detailed address resolution, traffic routing, security validation and request-response path between a client and the web application.

Domain Name

A domain name was registered on Route 53 and linked to the Elastic Load Balancer (ELB) to provide a consistent means of accessing the web application. Assigned IP addresses to an instance of the EC2 Micro Virtual Machine are not static and will change each time the instance is restarted, this is guaranteed to break connection to the application. Situations like this are undesirable, and requires

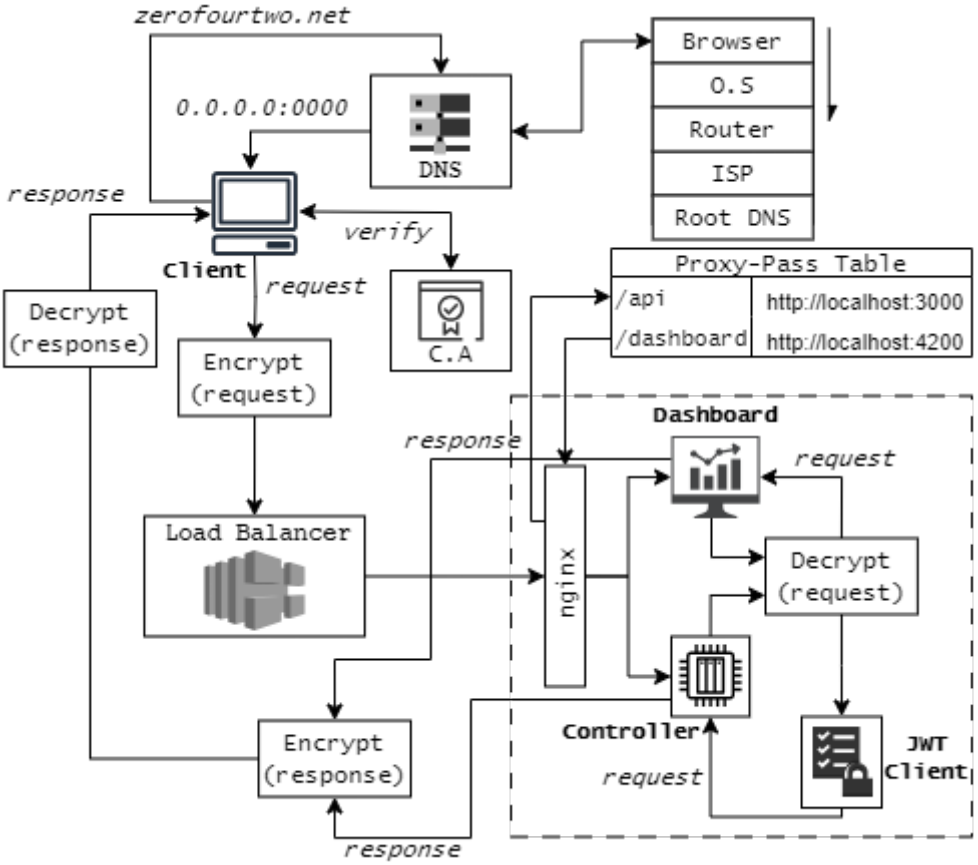


Figure 5.3: Communication & Security.

frequent modification of configuration files every time the IP address changes. Figure 5.4 shows the domain name configuration on Route 53.

Record name	Type	Routin...	Differe...	Alias	Value/Route tra...	TTL (s...	Health ...
zerofourtwo.net	A	Simple	-	Yes	www.zerofourtwo.n...	-	-
zerofourtwo.net	AAAA	Simple	-	Yes	dualstack.defaultro...	-	-
zerofourtwo.net	NS	Simple	-	No	ns-281.awsdns-35.c... ns-1985.awsdns-56... ns-709.awsdns-24.... ns-1040.awsdns-02...	2628000	-
zerofourtwo.net	SOA	Simple	-	No	ns-281.awsdns-35.c...	2628000	-
_ce07b3c4912d1d55...	CNAME	Simple	-	No	_4bb8063b75a548...	2628000	-
www.zerofourtwo.net	A	Simple	-	Yes	dualstack.defaultro...	-	-

Figure 5.4: DNS Records

Endpoint	Auth.	Params.	Query	Body	Desc.	Resp.
GET /api/dataset	Bearer	-	columns	-	Get dataset	200
GET /api/sync	Bearer	-		-	Sync database	200
GET /api/reset	Bearer	-	-	-	Reset database	200
GET /api/ai	Bearer	/:id	-	-	Run M/L with id	200
GET /api/transactions	-	/:id	-	-	Get results for id	203, 200
GET /api/images	-	/:id/path	-	-	Get image	200
POST /api/authorize	Bearer	-	-	json	Authorize user	200

Table 5.1: REST API Endpoints

5.2.1 REST API

The Representational State Transfer (REST) API standard was used to provide a stateless communication between various components of the system. The JSON data format was used to represent the data exchanged between the components. Functionalities of the web application is accessed using a client browser and thus requires a front end application to interact with the user. The front end application referred to as the Dashboard was designed and deployed as a stand alone application and requires data from the Models to perform its functionalities. A REST API was designed and developed to provide customized access to the data in the Models. Table 5.1 shows the list of the various endpoints of the API.

5.2.2 Security

The AWS Certificate Manager was used to provide the TSL/SSL security layer to the domain name, and handles all the complexity involved in key management and certificate provisioning [**?**]. Incoming request is routed to the LoadBalancer which was configured to re-route all incoming traffic to a registered target which is the EC2 Virtual Machine instance. The TSL/SSL security layer was attached to the LoadBalancer as a security policy to provide end-to-end encryption. The value for the domain name A and $AAAA$ records were set to the LoadBalancer DNS name, while the value for the domain name $CNAME$ record was set to that the certificate provided by the AWS Certificate Manager. Figure 5.5 shows the TSL/SSL record configuration on Certificate Manager.

The Bearer Authentication Token was used to further secure access to specific endpoints of the REST API. Consumers of the API are required to provide a valid token as a header in a request to successfully access the API. Generating a token involves digitally signing user information comprising of the header, payload and signature; using a secret key only known to the server. Decoding the token involves

Domain	Status	Renewal status	Type	CNAME name	CNAME value
zerofourtwo.net	Success	-	CNAME	_ce07b3c4912d1d55713b1c186f225e1a.zerofourtwo.net.	_4bb8063b75
*.zerofourtwo.net	Success	-	CNAME	_ce07b3c4912d1d55713b1c186f225e1a.zerofourtwo.net.	_4bb8063b75

Figure 5.5: TSL/SSL Record

verifying the signature using the secret key to ensure the token is valid, then user information is extracted from the payload. Figure 5.6 illustrates the process of signing and verifying a JWT.

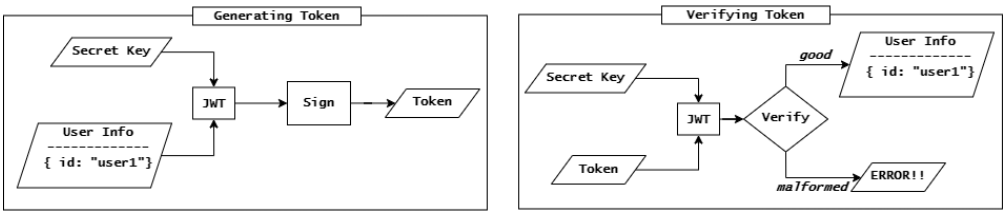


Figure 5.6: JWT Token

5.3 Controller

In a Model-View-Controller (MVC) architectural pattern, the Controller is at the center of the system's process flow and is responsible for handling user requests and producing an appropriate response to requests. The Controller was developed as a standalone web application listening on port 3000 and is capable of servicing multiple requests concurrently and referentially coupled with the Model. The Controller provides an interface for users to interact with the Models using RESTful API calls. Figure 5.7 shows the UML diagram of the Controller implementation.

5.4 View

The View component of the system otherwise known as the Dashboard was implemented as a standalone web application running on port 4200. The Dashboard provides an interface for users to interact with the system and consume the functionalities and data provided by the Controller Models respectively. The

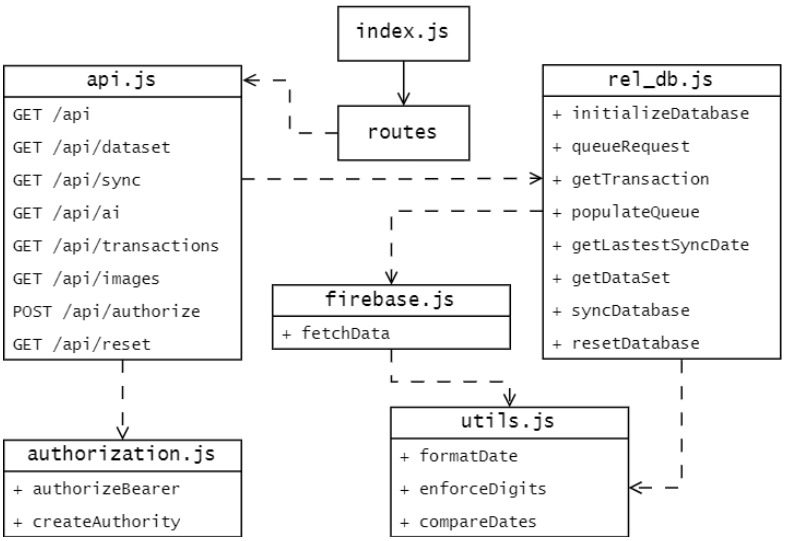


Figure 5.7: Controller UML Diagram.

Dashboard provides both administrative and data analysis functionalities. The administrative functionalities include pulling user data from the Firestore database, synchronizing the relational database with the Firestore database, resetting the relational database, setting up user authentication. The data analysis functionalities include filtering data for analysis and running machine learning algorithms for performance prediction. The Dashboard also allow users to visualize the results from the machine learning algorithms.

To provide real-time information to the user, the Dashboard utilizes Asynchronous Javascript and XML (AJAX) to make asynchronous requests to the controller and lazily updates the user interface with eventual response from the controller.

5.5 Model

The Model component of the system comprises the Machine Learning Models and the Databases. The controller queries the Models and gets a response which is passed on to the View on request.

5.5.1 M/L Models

For the purpose of finding correlation and predicting user performance given their biometric data, two different Machine Learning models were designed and devel-

oped. The models were designed to predict user's performance based on their biometric data. Algorithm and hyperparameter tuning was performed to select the best model for the Automatic Neural Network and Classical M/L Algorithms. An adhoc messaging protocol was developed to interact with the models and provide a comprehensive feedback about model performance. The M/L Algorithms were deployed as standalone services that can be activated by the Controller on request. On the low level, the Controller fires up the M/L Algorithm service and passing in a session identification parameter which is a unique identifier to a path where output from the M/L Algorithm is stored. Response from the M/L Algorithm is then read by the Controller and stored in the database. Further down the pipeline any enquiry for the stored session will be used to identify and retrieve associated resource from the database. Fig. 5.8 shows the sequence of interactions between the Controller and the M/L Algorithm.

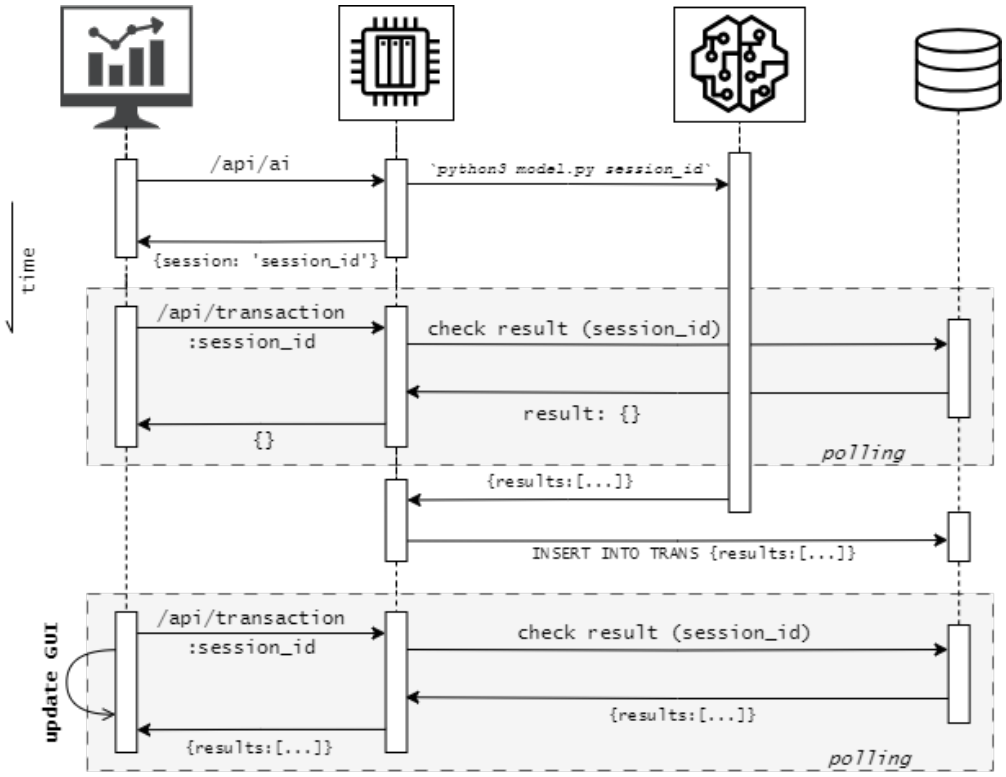


Figure 5.8: M/L Model Communication.

Classical M/L Algorithm

The model selection algorithm was developed to compare various classical machine learning algorithms and select the best performing model. For every independent

variable, the algorithm compares the performance of the various models using a 5-fold cross validation technique and selects the best model. Machine learning algorithms used in the model selection include: Linear Regression, Support Vector Machine, Random Forest and K-Nearest Neighbors.

Automatic Neural Network

A fully-connected Automatic Neural Network was developed to solve the same problem and provide a different perspective to the problem. The rationale for developing the Automatic Neural Network was to provide a more complex model capable of learning more complex pattern in the data.

5.5.2 Databases

The database stores user's information and provide a convenient and efficient way of accessing user's data. The relational database provides all data required to run the application and the firebase database provides an unstructured data bank for user's information from where data is sourced for further regularization.

Relational Database

The relational database was designed to store user's information and other information required to effectively manipulate, update, store and view the application state. Figure 5.9 shows the database schema.

Firestore Database

The Firebase database is part of the inherited infrastructure for this project. Data collected from the Polar Vantage V2 smartwatch and the Test Application are stored as documents in separate collections. These unstructured data requires further processing collation before use, thus they are aggregated and synchronized with the relational database. Figure 5.10 shows the process of synchronizing the Firestore database with the relational database.

The machine learning services were developed using the sklearn library, Keras TensorFlow library and Flask web framework. The services were deployed on a cloud server to provide real-time best model selection for users' performance prediction.

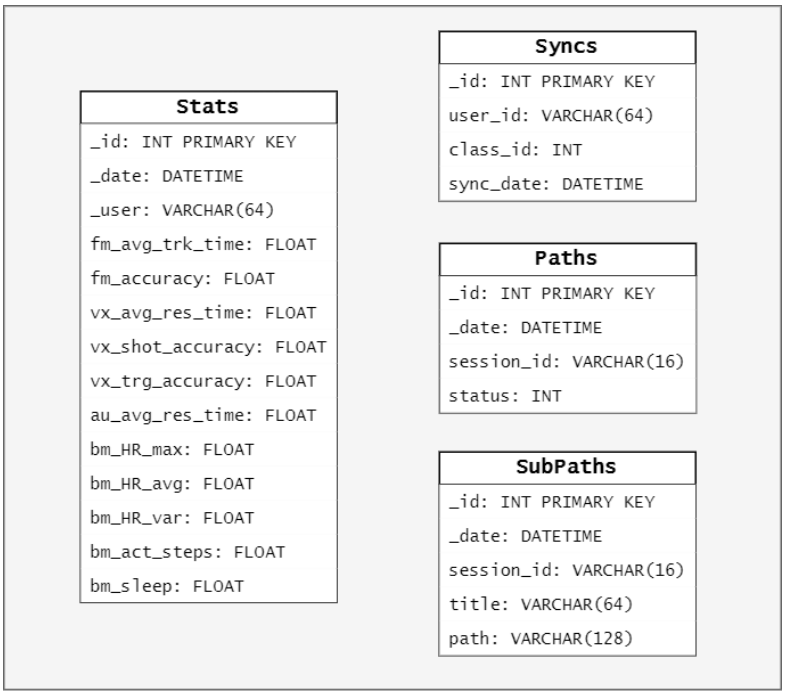


Figure 5.9: Database Schema

5.6 Inherited Infrastructure

A desktop application Test Platform was part of the inherited infrastructure for this project and was used to generate test data. The Test Platform was developed using Unity3D Game Engine and C# programming language. Results from the Test Platform were stored in a Firestore database. The Firestore database which is a document based database, and provides a convenient and efficient storage for unstructured data as obtainable in the Test Platform. The database also provides a very secured and scalable storage solution with real-time access.

For the purpose of collecting biometric data, the Polar Vantage V2 smartwatch was used. The smartwatch manufacturer provides a RESTful API repository for storing and accessing user's data collected by the smartwatch. The TestPlatform application provides functionalities to access the API and retrieve relevant biometric data from the smartwatch. The data is then stored in the Firestore database.

User authentication endpoint for Auth2.0 flow required for enrolling volunteers into the system was also inherited from the previous project. The endpoint was developed using Express.js and Embedded Javascript (EJS) technologies. The endpoint provides a secure way to authenticate users on the Polar Flow API to grant access to the users' biometric data.

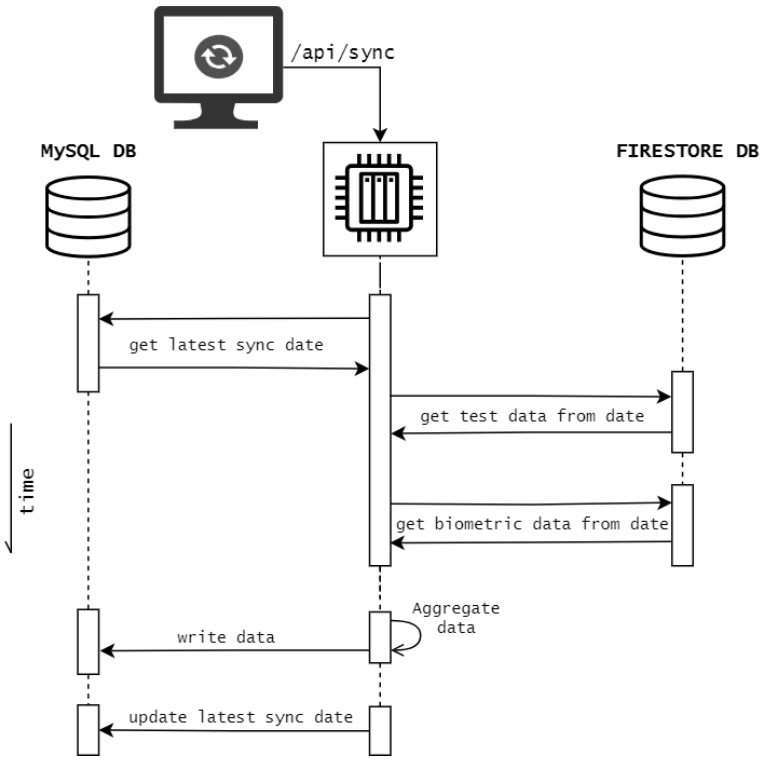


Figure 5.10: Database Synchronizing

Chapter 6

System Evaluation

This chapter presents the evaluation of the system infrastructure, architecture and the verification and validation techniques used. Also results from the machine learning models were presented and analyzed.

6.1 System Verification

Several components of the system interact together to perform the functionalities of the system and thus it is important to verify that each component works as expected. Each component was tested as a separate unit employing a test-driven development approach. An automated test suites were written for every functional component of the system to test its correctness and reliability.

6.1.1 Unit Testing

An automated unit testing strategy was employed to verify continuously evolving components of the system. This strategy was also utilized for Regression Testing to detect any breaking changes that may be introduced by new codes. User's data were derived from different collections (Biometric data & Performance data) of the Firestore database and were combined together as a single tuple in the relational database. Testing effort was focused on boiler-plate CRUD (Create, Read, Update, Delete) required to synchronize data between the databases and service the various API endpoints. Test cases for Asynchronous Javascript and XML (AJAX) calls were written verify integrity of data being fetched from the API endpoints. Unit tests suites were developed using Mocha for both the Dashboard and Controller components. Fig 6.1 shows the test cases developed for the system.

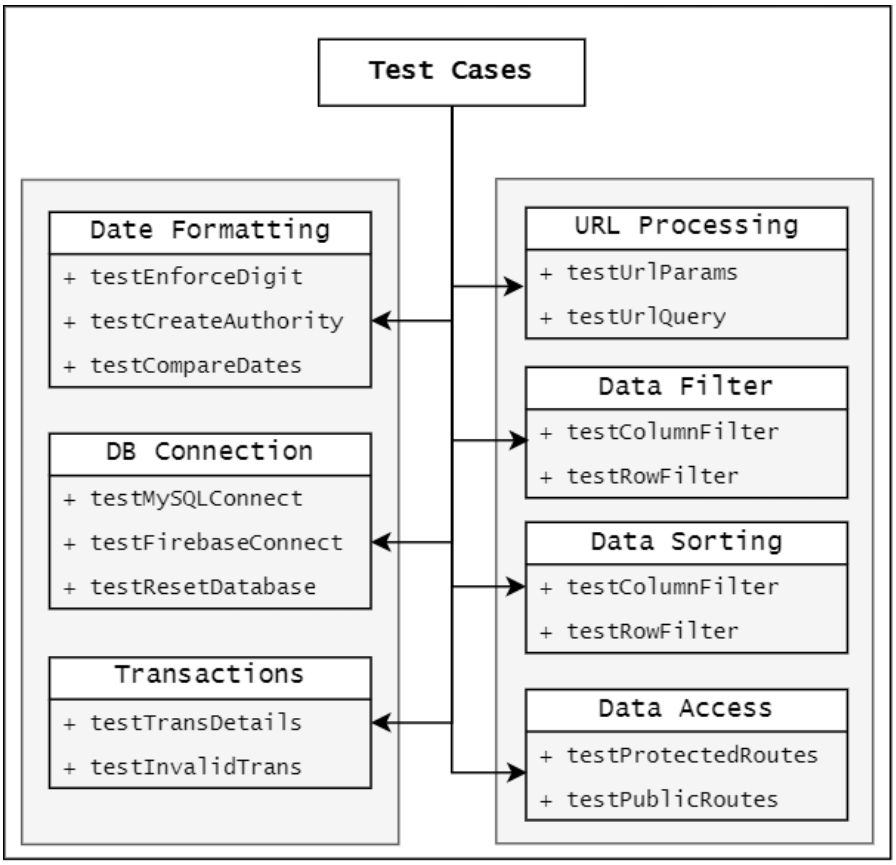


Figure 6.1: Test Cases

6.2 System Validation

Validating the system involved ensuring that the system meets the stated requirements and objectives. Acceptance testing was carried out to verify that the proposed functionalities of the system were implemented satisfactorily. The Acceptance testing was carried out with real data with the help of the supervising lecturer and volunteers. Functional and non-functional testing were carried out to validate the system requirements, performance and security. The following metrics were used to validate the system:

- **Performance** - The system was tested for handling multiple requests by simulating multiple users using Postman to send requests. The system satisfactorily serviced all requests without any noticeable delay.

- **Security** - The system was tested for security by attempting to access the

API endpoints without the required authentication. Such requests were rejected and appropriate error messages and code were returned.

- **Accessibility** - The system was tested for accessibility using different devices and browsers to ensure a consistent user experience across board and validate that AJAX calls were handled correctly across different browser implementations.
- **Usability** - The system was tested for usability, as the system was designed for non-technical users with the help of volunteers. The User Interface were supposed to be minimalistic and provide all functionalities in a single view. The users were able to perform the required tasks without any assistance.

6.3 Machine Learning Model Evaluation

6.4 Project Objectives

The initial project objective was to build an enduring infrastructure that can access users data and do real-time analysis. The system was shown to be able to reliably perform the stated tasks effectively. The achievement recorded in this project is summarized below:

- **Infrastructure** - A cloud-based infrastructure was successfully implemented using Amazon Web Services (AWS) as the cloud provider,

Chapter 7

Machine Learning Models

7.1 Introduction

As explained in dept in the methodology chapter, the biometric data collected through the watches generated a total of 6 independent variables:

- Heart Rate Maximum bm_HR_max
- Heart Rate Average bm_HR_avg
- Heart Rate Variability bm_HR_var
- Activity Steps bm_act_steps
- Sleep bm_sleep

The dependent variables that were collected through the gaming tests are:

- Fine Motor Average Tracking Time $fm\_avg\_trk\_time$
- Fine Motor Accuracy $fm\_accuracy$
- Visual Average Response Time $vx\_avg\_res\_time$
- Visual Shot Accuracy $vx\_shot\_accuracy$
- Visual Target Accuracy $vx\_trg\_accuracy$
- Audio Average Response Time $au\_avg\_res\_time$

The goal based on the research question was to find the correlation between the independent and dependent variables. When approaching a machine learning problem, one of the fundamental considerations is whether the problem is a regression or classification problem.

Classification vs Regression Problem

The main difference between a classification and regression problem is the type of dependent variable. The dependent variable, can be one or more, and they are the variables that are being predicted. In classification, the dependent variable is categorical, meaning it can take one of a limited number of values. Examples includes predicting whether an email is spam or not, predicting whether a patient has a disease or not. In regression, the dependent variable is continuous and numerical, meaning it can take any value within a range. Examples includes predicting house prices, stock prices, temperature. In this project, the dependent variables are continuous and numerical, making it a regression problem. The goal was to predict the dependent variables based on the biometric data collected from the watches.

7.2 Data Exploration

The first step in the machine learning process was to explore the dataset. The dataset was loaded into a *pandas* dataframe and the first 5 rows were displayed to get an overview of the data. The shape of the dataset was checked to see the number of rows and columns. The data types of the columns were checked to ensure that the data types were correct. The summary statistics of the dataset were checked to analyze the min and max, mean, standard deviation, and median values of the dataset. The heatmap 7.1 was then used to visualize the correlation between the variables.

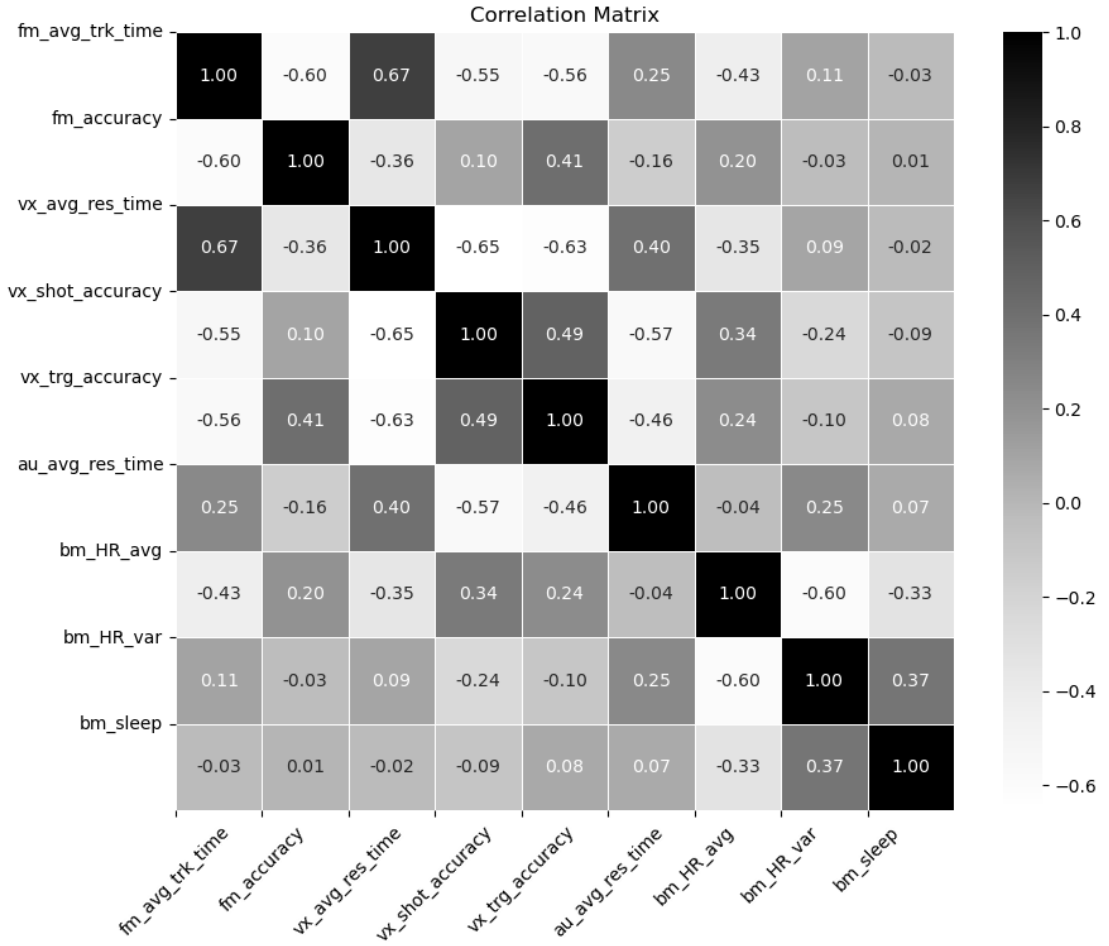


Figure 7.1: Correlation Heatmap

The heatmap employs the Pearson correlation coefficient to measure both the strength and direction of the linear relationship between two variables, in this case between the features (independent variables) and the target variables (dependent variables). It is a color-coded map, where the intensity of the color represents the strength of the correlation. With values that range from -1 to 1, where the negative values indicates a negative correlation, positive values indicates a positive correlation and 0 indicates no correlation. A value closer to 1 implies a strong correlation positive correlation, meaning as one variable increases, so does the other. A value closer to -1 implies a strong negative correlation, where one variable increases as the other decreases. The correlation heatmap provides a valuable insight into which features have the most significant influence on the target variables, helping understand the relationships within the data. [30]

Heart Rate Average (*bm_HR_avg*)

- Fine Motor Tracking Time (*fm_avg_trk_time*): A negative correlation of -0.43 suggests that as the average heart rate increases, the fine motor tracking time tasks decreases, indicating that individuals with faster heart rates tend to complete fine motor tasks a bit quicker.
- Fine Motor Accuracy (*fm_accuracy*): A positive correlation of 0.20 indicates a weak relationship suggesting that a higher heart rate average might be very slightly associated with higher fine motor accuracy. It shows that there is a small tendency for individuals with a higher heart rate to be slightly more precise in tasks that need fine motor skills.
- Visual Average Response Time (*vx_avg_res_time*): A negative correlation of -0.35 suggests a moderate relationship where a higher average heart rate is associated with faster response times. It indicates that individuals with higher heart rates also tend to react faster to visual things.
- Visual Shot Accuracy (*vx_shot_accuracy*): A positive correlation of 0.34 indicates a moderate relationship, suggesting that individuals with higher heart rate might have a better accuracy in shooting tasks.
- Visual Target Accuracy (*vx_trg_accuracy*): A positive correlation of 0.24 indicates a weak relationship, suggesting that higher heart rates averages might be associated with slightly better visual target accuracy.
- Audio Average Response Time (*au_avg_res_time*): A very weak negative correlation of -0.04 suggest almost no relationship between average heart rate and audio response time. It indicates there is no real connection between heart rate and how quickly the individuals responds to sounds.

Hear Rate Variability (*bm_HR_var*)

- Fine Motor Tracking Time (*fm_avg_trk_time*): A positive correlation of 0.11 suggests a weak association where greater heart rate variability might be associated with slightly longer fine motor tracking time tasks.
- Fine Motor Accuracy (*fm_accuracy*): A negative correlation of -0.03 suggest a very weak inverse relationship, where higher heart rate variability could be slightly associated with a decrease in fine motor accuracy.
- Visual Average Response Time (*vx_avg_res_time*): A positive correlation of 0.09 indicates very weak relationship, with slightly tendency for higher heart rate variability to be associated with longer visual response times.

- Visual Shot Accuracy (*vx_ shot_ accuracy*): A negative correlation of -0.24 indicates a weak to moderate inverse relationship, suggesting that higher rate variability might be associated with a decrease in visual shot accuracy. It suggests that individuals with higher heart rate variability might have a lower accuracy in shooting tasks.

- Visual Target Accuracy (*vx_ trg_ accuracy*): A negative correlation of -0.10 indicates a weak inverse relationship, suggesting that higher heart variability could slightly correlate with lower visual target accuracy. It suggests that those with more variation in their heart rate might not be quite as good at tasks that involve quickly identifying and selecting targets.

- Audio Average Response Time (*au_ avg_ res_ time*): A positive correlation of 0.25 suggests a weak relationship, indicating that higher heart rate variability might be associated with slightly longer audio response times.

Sleep (*bm_ sleep*)

- Fine Motor Tracking Time (*fm_ avg_ trk_ time*): With a value of -0.03, this negative correlation is very weak, indicating minimal inverse relationship between sleep and fine motor tracking time.

- Fine Motor Accuracy (*fm_ accuracy*): A positive correlation of 0.01 suggests a very weak relationship, indicating that longer sleep duration may be very slightly correlated with longer fine motor accuracy.

- Visual Average Response Time (*vx_ avg_ res_ time*): The value of -0.02 indicates a weak negative correlation, suggesting that increased sleep could be potentially be associated with a slight improvements in visual average response.

- Visual Shot Accuracy (*vx_ shot_ accuracy*): At -0.09, the correlation is weak, indicating there is almost no relationship between sleep and visual shot accuracy.

- Visual Target Accuracy (*vx_ trg_ accuracy*): At 0.08, the correlation is weak, suggesting a minimal tendency for more sleep to be associated with slightly better visual target accuracy.

- Audio Average Response Time (*au_ avg_ res_ time*): The correlation value of 0.07, which is very weak and suggests there is little to no meaningful correlation between sleep and response time to audio stimulus.

7.3 Data Pre-processing

The data underwent pre-processing to prepare it for machine learning models. This involved handling missing values by removing rows with missing entries as imputation wasn't possible. Features like *Activity Steps* and *Heart Rate Maximum* were found to have outliers that were excluded from the data. To ensure all features contributed equally, the data was scaled using *StandardScaler* from the *scikit-learn* library. Irrelevant columns like *_id*, *_date* and *_user* were removed. Finally, the preprocessed data was split into training (80%) and testing (20%) sets. The independent variables (features) were separated into X (independent), and the dependent variable (target) into a y. After pre-processing, the data was ready for model selection and evaluation.

7.4 Model Selection and Evaluation

Various machine learning models were explored to address the research question and predict the dependent variables based on the collected biometric and gaming test. Each team member focused on developing and evaluating a model to achieve the best predictive performance.

7.4.1 Classical ML Regression Models

The classical machine learning regression models were explored to predict the dependent variables based on the independent variables. The training process involved fitting the models on the training dataset and then assessing their performance on the unseen testing dataset and later on the full dataset. Mean squared error (MSE) served as the evaluation metric, with the model achieving the lowest MSE being chosen as the best performing model for further analysis. This approach ensures the chosen model generalizes well to unseen data, avoiding overfitting to the training set. The following algorithm were explored:

- Linear Regression: Chose as a baseline for its simplicity and interpretability, assuming a linear relationship between the dependent and independent variables. [31]
- Random Forest Regression: This algorithm is capable to capture complex non-linear relationships in the data and its robustness to overfitting. [32]
- Support Vector Regression: The capabilities of this algorithm goes beyond linear regression, capturing more complex relationships in the data and its robustness to outliers. [33]

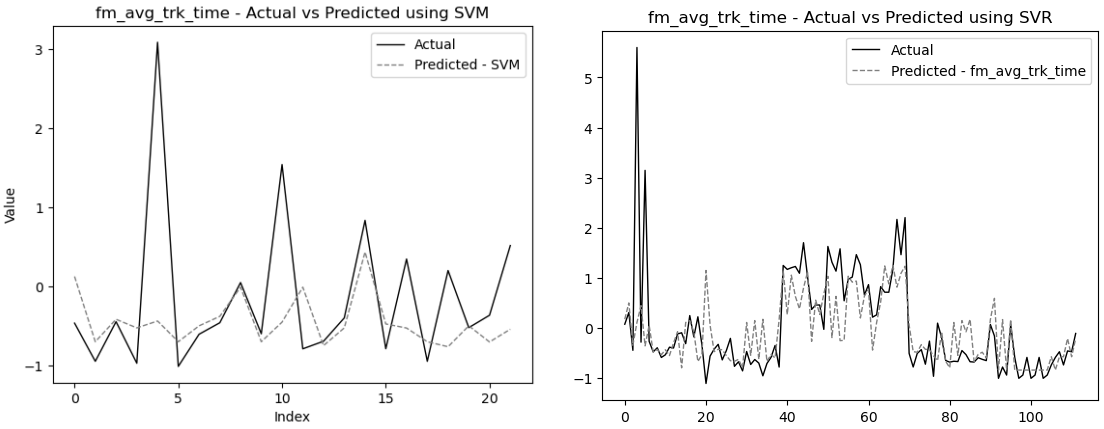
- K-Nearest Neighbors Regression: Chose for its simplicity and effectiveness in regression tasks, making predictions based on the average of the k-nearest neighbors in the feature space. [34]

Using cross validation, the models were trained on the training dataset and evaluated using the testing dataset. The model with the lowest mean squared error was selected as the best performing model for each dependent variable. The actual versus predicted values were plotted to visualize how well the model performed.

Actual vs Predicted Analysis

The actual values were plotted against the predicted values to see how well the model predicted the dependent variables. The plot was used to see if the model was underfitting or overfitting the data. If the points were close to the line, it indicated that the model was performing well. If the points were scattered, it indicated that the model was not performing well. It was also used to see if the model was capturing the underlying patterns in the data. Below are the actual versus predicted plots for the best performing model for each dependent variable.

Fine Motor Tracking Time



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.2: Fine Motor Tracking Time Actual vs Predicted

Test Data Evaluation

- **Trend**: Figure 7.2a shows the actual versus predicted values for the Fine Motor Average Tracking Time using the test data. The model that performed

the best for this dependent variable was the Support Vector Regressor. The predictions generally matched the trajectory of the actual test data. The peaks and troughs of the actual data were also found in the predicted data, with some deviations.

- **Variance**: There is a noticeable variance between the actual and predicted values, specially at the extremes. For instance, the model appears to underestimate the highest values and overestimate the lowest values.

- **Consistency**: The model shows decent consistency when the actual values are around the mean but is less consistent at capturing sudden changes in the actual data, such as sharp spikes ot dips.

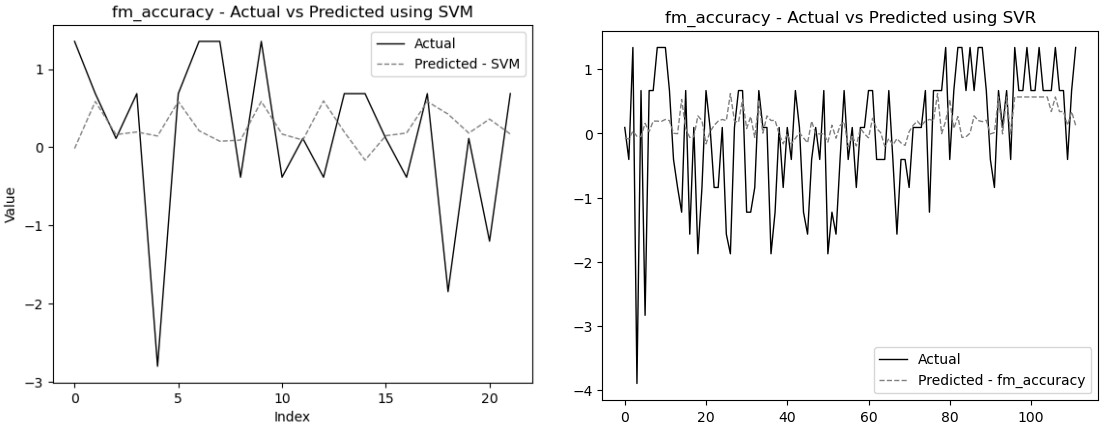
All Data Evaluation

- **Trend**: Figure 7.2b shows the actual versus predicted values for the Fine Motor Average Tracking Time using all the data available. When evaluating all the data, the model predictions closely follow the actual data's trend. This indicates that the model has learned the overall behavior of the dataset quite well. Similar to the test data, the model captures the general pattern of movement in the actual values, but might not always match the amplitude of changes.

- **Variance**: The variance between the actual and predicted values over all the data seems to be lower compared to the test data. This suggests that the model has been effectively trained to understand the data as a whole. Some exceptions occur where the actual values show significant deviation from the mean. In these areas, the predictions do not fully capture the extent of the actual values and show some deviation.

- **Consistency**: The model shows good consistency in predicting the values when the actual values are around the mean. The predictions often match the actual values quite closely.

Overall the Support Vector Regressor Model performed well in both scenarios, capturing the general trends of the Fine Motor Tracking Time and showing good consistency in prediction. The model, however, seems to struggle with accurately predicting the more extreme values in the dataset. It highlights the need for improving the model's performance on the more complex or extreme segments of the data.



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.3: Fine Motor Accuracy Actual vs Predicted

Fine Motor Accuracy

Test Data Evaluation

- **Trend**: Figure 7.3a shows the actual versus predicted values for the Fine Motor Accuracy using the test data. The model that performed the best for this dependent variable was the Support Vector Regressor. The model appears to capture the general trend of the actual data. It follows the directional changes, going up and down as the actual values do, but not with perfect alignment.

- **Variance**: A noticeable discrepancy between the actual and predicted values is noted, particularly evident at the extremes. While the model tends to underpredict some of the sharper declines in actual values, it still remains relatively closely to the true data points.

- **Consistency**: There is a moderate level of consistency in the predictions. The model seems to perform well when the actual values are not showing extreme behavior. However, in instances where the actual values show sharp changes, the model's consistency is reduced.

All Data Evaluation

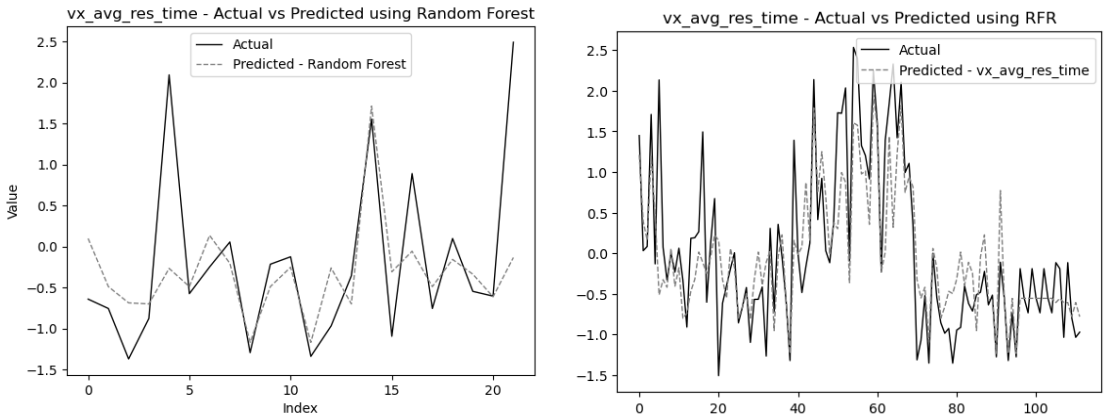
- **Trend**: Figure 7.3b shows the actual versus predicted values for the Fine Motor Accuracy using all the data available. Across the full dataset, the Support Vector Regressor model generally traces the movements of the actual values well. This suggests a solid understanding of the underlying patterns

in the full range of data. Despite some misalignment, the predicted values consistently mirror the ups and downs in the actual data, indicating the model's capability to track the overall trend.

- **Variance**: Compared to the test data, the variance in the full data set appears to be slightly more visible. The predictions occasionally deviate significantly from the actual values, particularly where there are sharp changes in the actual data or outliers in the actual data.
- **Consistency**: The model shows good consistency in predicting the values when the actual values are around the mean. It does, however, display some inconsistencies, again mainly where there are larger deviations in the actual data.

Overall, the Support Vector Regressor model demonstrated an ability to find the overall patterns in the Fine Motor Accuracy, both in the test set and the full dataset. While the model tends to stay close to the actual values, it does show some variance and inconsistency, particularly in areas where the actual data shows sharp changes or outliers. This suggests that the model may need further refinement to improve its performance in these more complex or extreme segments of the data. Also, the model may benefit from more data to help it better understand the full range of patterns in the dataset.

Visual Average Response Time



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.4: Visual Average Response Time Actual vs Predicted

Test Data Evaluation

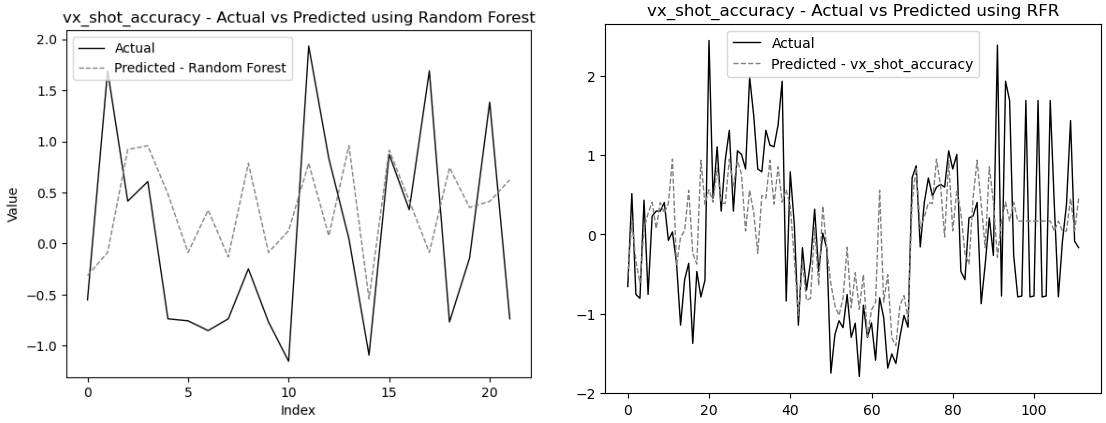
- **Trend**: Figure 7.4a shows the actual versus predicted values for the Visual Average Response Time using the test data. The model that performed the best for this dependent variable was the Random Forest Regressor. The model successfully captured the underlying trend of the actual data. It demonstrates an ability to learn the fundamental patterns in the data, following the general trajectory of the actual values.
- **Variance**: While the model generally aligns with the actual values, there are areas where the deviation from the actual data is noticeable. The model doesn't always capture the sharpness and valleys which can be seen in several areas of the actual data.
- **Consistency**: In sections where the data does not fluctuate significantly, the predictions are consistent with the actual values. This suggests a level of reliability of the model, under the condition that the actual data does not show extreme behavior.

All Data Evaluation

- **Trend**: Figure 7.4b shows the actual versus predicted values for the Visual Average Response Time using all the data available. When applied to the entire dataset, the model displays an ability to mimic the overall trend line of the actual data. The predictions fluctuate in line with the actual values, showing an understanding of the larger patterns in the data.
- **Variance**: As with the test data, there are discrepancies between the actual and predicted values; these are most apparent where is a sharp change in the actual data. The model sometimes smooths out these abrupt changes, leading to a slight deviation from the actual values.
- **Consistency**: The model shows a decent level of consistency throughout the entire data range. While there are mismatches, particularly in areas of high variability, the model maintains a close following with the actual values for the most part.

In both the test data and the full dataset, the Random Forest Regressor model for the Visual Average Response Time demonstrates effectiveness in capturing the main trends and movements. It shows a good degree of consistency, with understandable variance in places where the data is more complex or extreme. The model appears to perform better when dealing with average levels of fluctuation but might require further tuning to more accurately predict the more extreme variations observed in the actual data.

Visual Shot Accuracy



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.5: Visual Shot Accuracy Actual vs Predicted

Test Data Evaluation

- **Trend**: Figure 7.5a shows the actual versus predicted values for the Visual Shot Accuracy using the test data. The model that performed the best for this dependent variable was the Random Forest Regressor. The model seems to reasonably follow the general of the actual values. It responds to the directional shifts in the data, rising and falling in correlation with the actual data. However, there are noticeable disparities at certain points, indicating that while the model detects the overall pattern, it does not always align perfectly with the actual data's trajectory.

- **Variance**: It is clear that the model's predictions do not perfectly match the actual values, especially at the peaks. The model performs more reliably at data points that are closer to the central trend and shows more deviation at the extremes.

- **Consistency**: The model demonstrates moderate consistency throughout the test dataset, capturing the general trend of the actual data with some fidelity. However,
its ability to mirror the actual data precisely at every point is limited, especially where there is significant fluctuation in the actual data.

All Data Evaluation

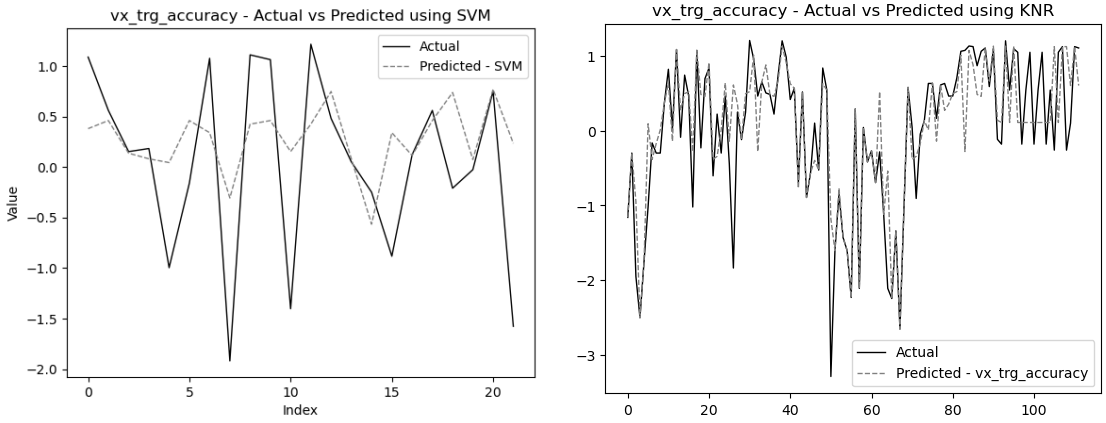
- **Trend**: Figure 7.5b shows the actual versus predicted values for the Visual Shot Accuracy using all the data available. With the entire dataset, the model's prediction still follow the actual values' general trend, suggesting an understanding of the global behavior of the data. The model struggles with sharp spikes and drops, smoothing over some of the more extreme changes in the actual data.
- **Variance**: The variance across the full dataset seems to be slightly more controlled than in the test data, although discrepancies remain. The model does not perfectly capture the amplitude of changes, particularly the higher peaks and lower troughs in the actual data.
- **Consistency**: Across the full scope of data, the model's predictions display an adequate level of consistency, reflecting a steady predictive performance that often aligns well with the real values. However, the model's prediction can diverge from the actual data at points, indicating room for improvement in model accuracy.

The Random Forest Regressor model shows that it captured the central tendencies and fluctuations in Visual Shot Accuracy both in the testing phase and across the complete dataset. The model's capability of following the trends is a strong point, but its prediction variance and the consistency of its performance present areas for improvement.

Visual Target Accuracy

Test Data Evaluation

- **Trend**: Figure 7.6a shows the actual versus predicted values for the Visual Target Accuracy using the test data. The model that performed best for this dependent variable was the Support Vector Regressor Model. The model's predicted outcomes show a basic alignment with the actual data's trend. The general ups and downs in the actual data are mirrored in the model's predictions, indicating a broad understanding of the data's patterns. However, there are instances where the model does not perfectly trace the actual data's peaks and troughs, suggesting that while the trend is recognized, the precision of the model could be improved.
- **Variance**: The variance between the actual and predicted values seems moderate. The model seems to smooth out some of the actual data fluctuations, not capturing the extremes as it should.



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.6: Visual Target Accuracy Actual vs Predicted

- **Consistency**: There is a moderate level of consistency in the model's predictions across the test dataset. The predicted values are regularly close to the actual data, but there are instances where the model's predictions deviate from the actual data.

All Data Evaluation

- **Trend**: Figure 7.6b shows the actual versus predicted values for the Visual Target Accuracy using all the data available. Over the entire dataset, the Support Vector Regressor model's predictions generally follow the trend of the actual values, indicating a stable understanding of the data's patterns. Similar to the test data, the predictions are in line with the overall movements but occasionally miss the mark in capturing the specific oscillations in the actual data.

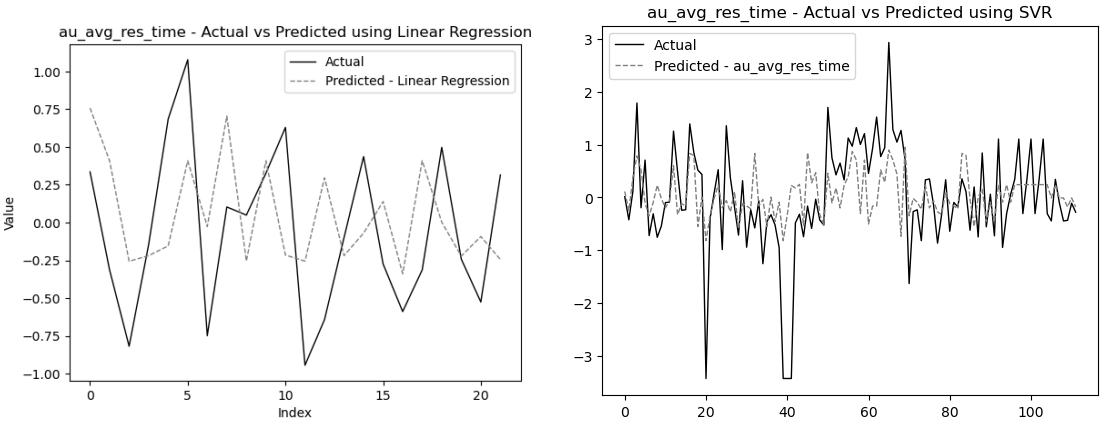
- **Variance**: The prediction variance across the full dataset appears somewhat greater compared to the test data. This could be due to the inclusion of more diverse data points, revealing the model's limitations in adapting to broader patterns. The model especially tends to underpredict some of the lower values, leading to a noticeable deviation in the more extreme data points.

- **Consistency**: Throughout the full dataset, the consistency of the predictions are apparent but not without error. It seems to stay true to the mean trajectory but lacks in tracking the finer details of the actual data.

The Support Vector Regressor Model for the Visual Target Accuracy demonstrates a basic understanding of the data's patterns, capturing the general trends

in the test data and the full dataset. However, the model exhibits a level of variance that indicates room for improvement in its predictive accuracy. Specifically, the model could benefit from refinement to better capture the more extreme values in the actual data.

Audio Average Response Time



(a) Actual vs Predicted (Test Data) (b) Actual vs Predicted (All Data)

Figure 7.7: Audio Average Response Time Actual vs Predicted

Test Data Evaluation

- **Trend**: Figure 7.7a shows the actual versus predicted values for the Audio Average Response Time using the test data. The model that performed best for this dependent variable was the Linear Regression. The model's predictions tends to reflect the actual data's trend, indicating a general grasp of the data's patterns. The prediction line shows a consistent rise and fall pattern that corresponds with the actual data line, illustrating the model's ability to track changes in response time.

- **Variance**: There is a significant difference between the predicted and actual values at multiple points. This difference highlights areas where the model's predictions are not as close to the actual values, especially at sharp downward and upward turns in the actual data.

- **Consistency**: Despite some variances, the Linear Regression model remains fairly consistent in its predictions across throughout the test dataset. While

it does not precisely match the actual values, it provides a reasonably consistent estimation, particularly in the middle range of the response time values.

All Data Evaluation

- **Trend**: Figure 7.7b shows the actual versus predicted values for the Audio Average Response Time using all the data available. Looking at the entire dataset, the Linear Regression model again mirrors the general trend of the actual data, showing an ability to comprehend the ups and downs in the response time. It captures the global trend across a broader range of data points.
- **Variance**: The overall variance in predictions across the entire dataset appears to be somewhat controlled, but deviations are still present. The model occasionally smoothens the extremities in the actual data, leading under or overestimation in some segments.
- **Consistency**: In terms of consistency, the model show a reasonable degree of reliability across the entire dataset. Its predictions are in alignment with the direction of the actual data, even though it does not match every peak and valley with precision.

The Linear Regression model for the Audio Average Response Time demonstrates a competent understanding of the Audio Average Response Time patterns, both in the test and the full dataset. It demonstrates a level of consistency that suggests it captures the variable's fundamental characteristics. However, the model's prediction exhibit some variance, indicating room for improvement in its ability to capture more subtle fluctuations in the data.

7.4.2 Neural Network Regression Model

In addition to the classical machine learning models, a neural network regression model was explored to predict the dependent variables based on the independent variables. The neural network model was constructed using the *MPLRegressor* from the *scikit-learn* library.

Hyperparameters

The neural network was initialized as an *MLPRegressor* with the following configurations:

- **Hidden Layers**: Two hidden layers were used with the first containing 15 neurons and second 10 neurons. This configuration is complex enough to

capture the non-linear patterns in the data but not too complex to overfit the data.

- **Activation Function**: The activation function used was the *tanh* function, allowing the model to account for non-linear relationships between the features and the target variables.

- **Solver**: The *adam* solver was chosen due to its effective performance in deep learning applications, particularly in handling noisy data and large datasets.

- **Max Iterations**: The training process was set to a maximum of 1000 iterations to ensure sufficient learning without overfitting.

Training Process

The training process was defined to run for a maximum of 1000 iterations ('epochs'). Each epoch represents a complete pass through the entire training dataset.

Loss Curve Analysis

The *loss curve* attribute of the model was plotted against the number of epochs to visualize the model's training progress as shown in Figure 7.8. This curve represents the model's performance improvement over time as it learns from the training data. The plotted loss curve shows a typical pattern of sharp decline followed by a gradual leveling off, indicating initial rapid learning followed by a slower convergence towards the optimal set of weights.

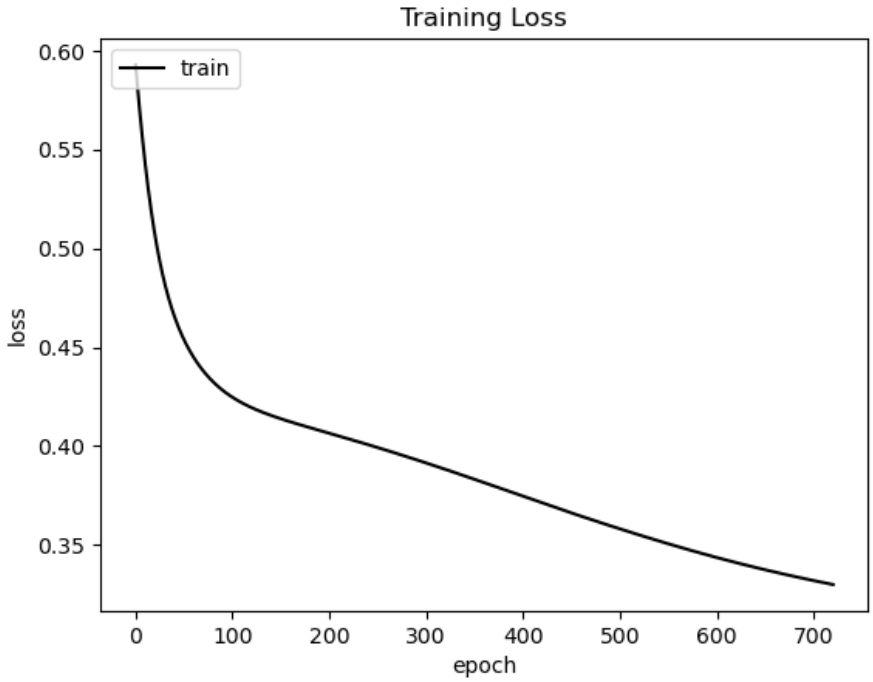


Figure 7.8: Neural Network Loss Curve

The plotted training loss curve confirms the effective learning capability of the model with given configurations. The gradual decrease in loss over the epochs indicates the optimization of weights and biases to minimize the error between the predicted and actual values.

Model Evaluation

The model was evaluated using the test dataset and all the data available. The predictions were compared to the actual values to assess its performance. The actual versus predicted values were plotted to visualize how well the model predicted the dependent variables. The plots were used to compare the model's performance with the classical machine learning models.

When contrasted with the classical machine learning models, the neural network model demonstrated a ability to capture complex, nonlinear interactions within the data, fitting closely to the variances and trends in the actual data. This characteristic can lead to highly accurate predictions on the training dataset but may also cause the model to overfit, as evidenced by prediction lines that follow the actual data with high fidelity, including noise and outliers.

- **Trend**: The neural network model are dynamic, capable of adapting to data fluctuations, which can be both an asset and a liability. They may

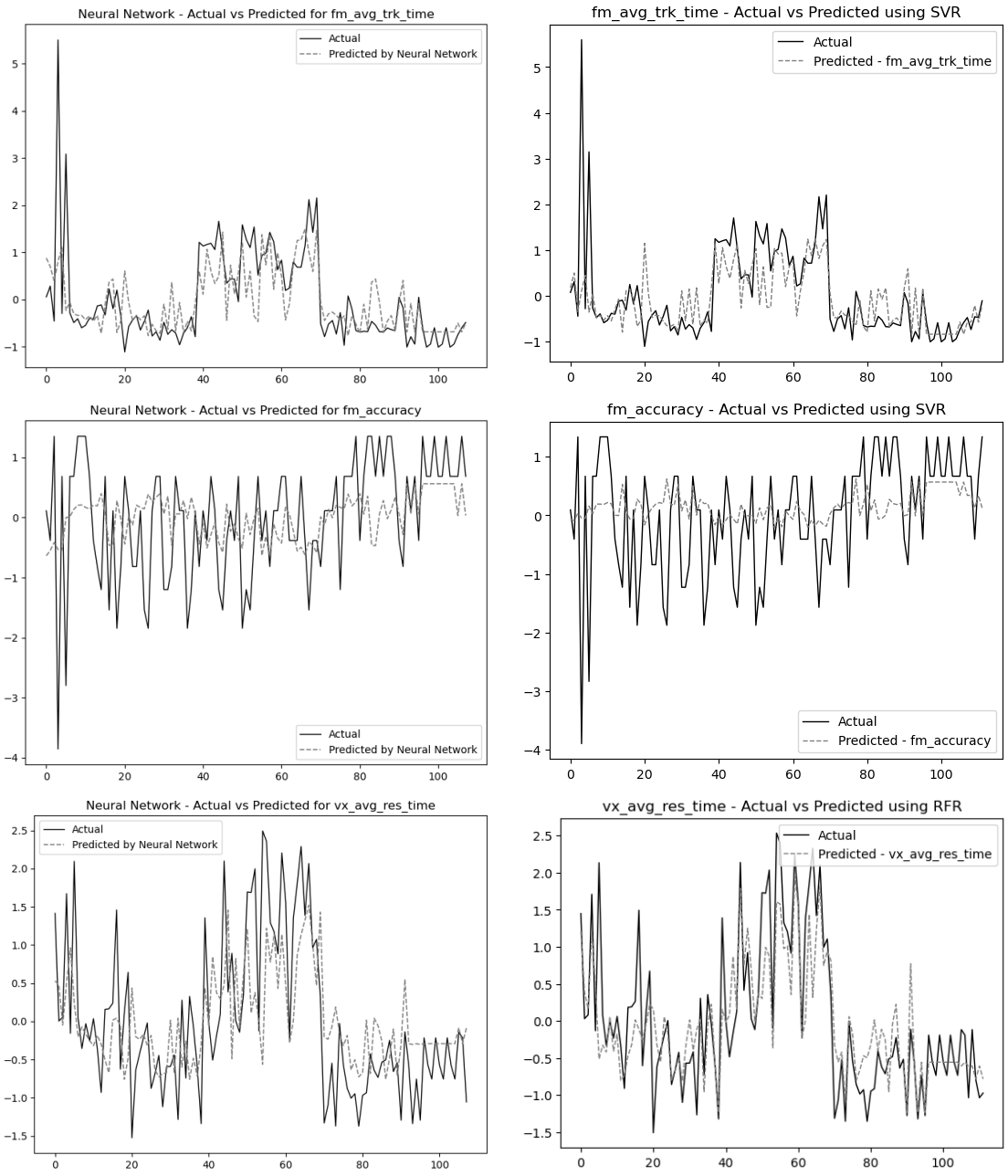


Figure 7.9: Neural Network Model vs Classical ML Models

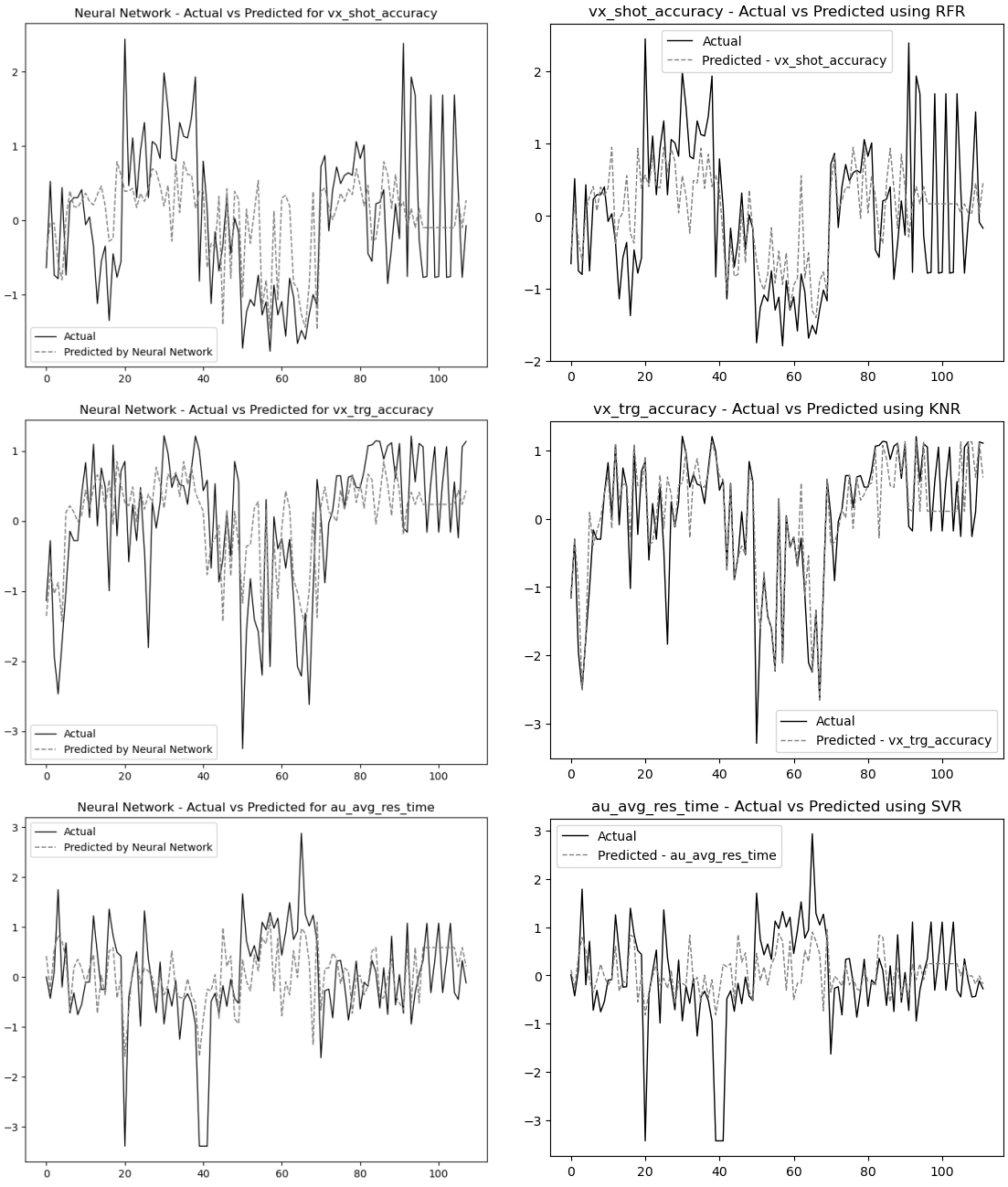


Figure 7.10: Neural Network Model vs Classical ML Models

reflect underlying patterns with precision but also amplify anomalies. In contrast, classical models such as Support Vector Regressors and Random Forest Regressors favor a moderate approach, often smoothing over data to project a more stable and generalized trend line.

- **Variance**: The neural network model exhibit a higher variance in predictions compared to classical models. This can be beneficial in capturing the full range of data patterns but also poses the risk of overfitting to noise and outliers. On the other hand, classical models typically show less variance, indicating a tempered response to data variability, which might overlook complex patterns but also ensures robustness against noise.

- **Consistency**: The adaptability of neural networks, which enables them to capture a wide range of data behaviors, may also affect their consistency across different conditions. In contrast, the classical models can be specially valuable when the objective emphasizes stability over detailed accuracy.

Overall the NN model outperforms the classical models in terms of trend following, variance capturing and overall consistency, particularly when evaluating the full dataset. This suggests that the neural network model is well-suited for capturing the complex relationships between the independent and dependent variables in the dataset, while classical models provide a more generalized approach that may be more suitable for stable, less complex datasets. For this project, due to a very close performance between the neural network and classical models, we conclude that as the dataset grows, the neural network model may be more suitable for capturing the complex relationships between the independent and dependent variables in the dataset, but the classical models should not be overlooked as they provide a more generalized approach that may be more suitable for stable, less complex datasets.

The balance between sensitivity to data changes and maintaining a level of generalization seems to be the area where the neural network's performance could be further optimized.

Chapter 8

Conclusion

Appendix

Ethics Application

Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

Taught Programme Research Ethics Approval Application Form

Research undertaken by taught students must receive ethical approval unless deemed exempt. This application form may be completed by an individual student or by a Programme Board/Lecturer for a group of similar research projects.

This application is completed by:

Student: ✓ OR Lecturer on behalf of Programme Board: ☐

PART A

Applicant Details	
Name:	Otito Mbelu, Rodrigo Almeida
Student ID:	G00397738, G00377123
Programme Title:	Computing in Software Development
Programme Stage:	Year 4
Research Supervisor's Name:	Dr. Damien Costello

Project Details	
Research Study Title:	Biometric Data Analysis in Digital Game Scenario
Research Study Summary (max 100 words): The main hypothesis of this research is to find if there is any correlation between a user's biometric data and their performance in a first-person shooter game scenario. This research study will try to quantify how biometric data such as Heart Rate Variation, Heart Rate, active steps taken, quality and quantity of sleep, etc. affect select gaming skills like eye-to-hand coordination, fine-motor skills and reaction time.	

Risk Checklist Please answer ALL the questions in each of the sections below – Tick YES or NO			
	Will the research study….?	YES	NO
1	Involve direct and/or indirect contact with human participants?	✓	
2	Involve analysis of pre-existing data which contains personal or sensitive information not in the public domain?		✓
3	Require permission or consent to conduct?	✓	
4	Require permission or consent to publish?		✓
5	Have a risk of compromising confidentiality?		✓
6	Have a risk of compromising anonymity?		✓
7	Collect/contain personal data i.e. any information that relates to an identified or identifiable individual?		✓
8	Collect/contain sensitive personal data e.g. health data, sexual orientation, race religion?	✓	
9	Contain elements which you OR your supervisor are NOT trained to conduct?		✓
10	Use any information OTHER than that which is freely available in the public domain?		✓
11	Involve respondents to the internet or other visual/vocal methods where participants may be identified?		✓
12	Include a financial incentive to participate in the research?		✓
13	Involve our own students or staff?	✓	
14	Take place outside Ireland?		✓
15	Involve participants who are vulnerable or at risk?		✓
16	Involve any participants who are unable to give informed consent?		✓
17	Involve data collection taking place BEFORE informed consent is given?		✓
18	Involve any deliberate deception or covert data collection?		✓
19	Involve a risk to the researcher or participants beyond that experienced in everyday life?		✓
20	Cause (or could cause) physical or psychological harm or negative consequences?		✓
21	Use intrusive or invasive procedures?		✓
22	Involve a clinical trial?		✓

Risk Checklist Please answer ALL the questions in each of the sections below – Tick YES or NO			
	Will the research study….?	YES	NO
23	Involve the possibility of incidental findings related to participant health status?		✓
24	Involve the remuneration of research participants?		✓

If, as a student, you answered **NO** to all the above questions your research supervisor will review, and if in agreement sign below to indicate that this form does not have to be submitted to the Taught Programme Research Ethics Committee.

Name		*Signed*		*Date*	

If you answered **YES** to any of the above questions, you need to complete part B below.

PART B

1	**Project Overview**

Please give a brief overview of the study, including a summary of the aims and objectives.

Help: Describe the purpose of the research and what question(s) the project should answer.

Introduction to PUBG: PUBG, short for Player Unknown's Battlegrounds, is a popular online multiplayer battle royale game developed and published by PUBG Corporation, a subsidiary of Bluehole Studio. In PUBG, up to 100 players parachute onto an island and scavenge for weapons and equipment to kill others while avoiding getting killed themselves. The game features a shrinking safe area to force players into close encounters, promoting tactical gameplay and intense firefights. The last player or team standing wins the game. PUBG became widely popular upon its release in 2017 and is available on various platforms, including PCs, consoles, and mobile devices1. Since its launch in 2017, PUBG has become incredibly popular, with more than 280M active players. People of all ages and from all over the world enjoy playing this game. In PUBG, players need good hand-eye coordination and quick reflexes to compete well against others. The game offers a variety of weapons, and players can customize their controls to fit their preferences. This project builds upon the research done by university students, focusing on collecting biometric data to improve player performance in digital games.

The original project, titled "Biometric Data Collection for Performance Optimization in a Digital Game Scenario", aimed to find out if a player's body data could help them play a video game better. The researchers set up a practice game similar to PUBG: Battlegrounds, with the same weapons and controls in a Unity Desktop Application. Biometric data was supplied by a Smart Watch. In this instance, the only data used to test the application was produced by the developers themselves and from the stakeholders, for this reason, it wasn't necessary an Ethics application as there wasn't a recruitment of volunteers at the time. The goal of the original project was to see if there was a connection between how players performed in the game and their body data. In our third-year project, our team was given the challenge of enhancing the visualization of data from both the test games and the Polar API. To achieve this, we developed a chart API, enabling users to conveniently view all the results on a single page. The chart API was specifically created to present the data collected during the initial research in a coherent and user-friendly chart format, integrating various user data into one unified visual representation. Project Objective The aim of this research project is to overcome the limitations mentioned in previous projects and complete the future development goals of both projects. These include:

• Chart API Integration: Integration of developed Chart API into the test application.

• Offline Data Storage: Provision will be made for offline temporary file storage to improve the overall reliability of the whole system

• PUBG API: Further research on new developments in the PUBG API for better user experience.

And eventually, seek to answer the following research questions:

- Can the user's current physical condition as indicated by their Biometric data, have any direct relationship with their performance in such a gaming scenario?

- Can Biometric and test data help suggest the most suitable settings for different game scenarios?

2	**Methodology**

Please give a description of the methodology, including any data collection and analysis methods.

Help: Give an outline of the study here. If the project is complex, you can also submit the research proposal/protocol (no more than 2-3 A4 sides) if this would help the reviewer's understanding of the project. Include details of your (or the Research Supervisor's) appropriate skills and qualifications to carry out this research. Consideration of how, and for what duration are stored should be provided under Section 7 below.

This research is geared towards finding a correlation between users' biometric data to their performance in a first-person shooter gaming scenario. The quality of the user data collected will be critical in ensuring that the research aims are achieved. For Quality Assurance purposes, a systematic step-by-step approach has been designed for the data capturing process.

An activity monitoring device will be distributed to volunteers in the form of a Polar Watch. Individuals are expected to use these devices at a pre-designated times prior to undertaking a test in the form of a First-Person Shooter Game. Results from the test will be paired with their biometric data for further analysis.

Data of Interest

For the purpose of the research, different categories of data will be collected from the volunteers through the monitoring device to help achieve the stated goals. These data relate to the volunteer's physical information, sleep data, daily activities, and nightly recharge.

For the Test Game, volunteers are expected to play a total of 3 different categories of test to gauge their performance in these tests. The tests are Audio Test, Visual Test, and Fine Motor Test. The tests are designed to be played sequentially and will typically take a total of 15 minutes to conclude.

Procedure

The data collection process is expected to follow the below listed steps to maximize throughput and minimize the average time spent processing each participant.

Location

The data collection effort will be carried out at the Gym on the Atlantic Technological University (ATU) Galway campus.

Time

The following weekly schedule will be available for volunteers at various times that best suit their personal schedule.

· Tuesdays: 12:00 - 13:00, 15:00 - 16:00

· Wednesdays: 12:00 - 13:00, 15:00 - 16:00

· Thursdays: 12:00 - 13:00, 15:00 - 16:00

· Fridays: 13:00 - 15:00

Note: The timing is open for adjustment to suite volunteers

Biometric Data Procedure:

For the very first time processing a volunteer, some user information will be needed to have them registered with the activity monitoring device and their data is saved to the manufacturer's repository.

The Activity Monitoring Device to be used in this research is the Polar Vantage Smart Watch. The watch is capable of monitoring and capturing user's biometric data and saving the data to a repository where it can be accessed online via an API. The device is commercially available and accessible to the public. The rationale for using the Polar Vantage Watch is because it has been widely used in both academic and industrial research projects and validation. Most importantly, these devices are available to us in such quantity that can satisfy our research needs.

Physical Information

These are the first group of data to be collected about the individual to register the Monitoring device manufacturer. The data is stored with the manufacturer (Polar) which is GDPR compliant.

Sleep Data

According to the manufacturer's manual, the watch is capable of recording the quality and quantity of sleep and also showing how long spent on each stage of sleep. Light Sleep, Deep Sleep, and REM Sleep with the duration of each type of sleep. For the purpose of the research, an aggregated total of the various categories of sleep will be recorded and used.

Volunteers are expected to wear the watch to sleep the previous night before their scheduled test

· Sleep (minutes)

Daily Activities

According to the Activity monitor manufacturers' manual, "Polar device uses an internal 3D accelerometer to record your wrist movements. It analyses the frequency, intensity and regularity of your movements together with your physical information." calories, active calories, active steps and their respective durations are collected from through the device. For the purpose of the research, the active steps will be used.

Volunteers are expected to wear the previous day for this data to be available.

· Active Steps (count)

Nightly Recharge

From the Activity monitor manufacturers' manual, the Nightly Recharge is recorded as follows: " is an overnight recovery measurement that shows how well your body has coped with overall stress you have experienced lately." The parameters, measured during roughly the first four hours of your sleep are heart rate, heart rate variability and breathing rate. For the purpose of the research, the following parameters and derived parameters will be used:

· Heart Rate Average (bpm)

· Heart Rate Maximum (bpm)

· Heart Rate Variability (HRV)

Once again, volunteers are expected to wear the watch to sleep the previous night before their scheduled test.

Test Data Collection

The test data is generated at the completion of the test session in the Test Application. A test session consists of three categories of tests, which will be undertaken sequentially in different stages. For quality assurance, the test will be undertaken in a controlled environment using the same hardware in similar conditions over the course of the trials.
For the purpose of the trial, a special designated room has been reserved for the data collection effort. The following metrics for the different tests will be captured and will be subsequently used for the research effort.
Audio Test: designed to test users' auditory precision. This test is expected to generate.
`Average Response Time`: this is a measure of how quickly a user is able to react and identify the source of a sound with varying audibility (in decibels) in a 3-Dimensional space.
Visual Test: designed to test users' visual perception. At the end of this category of test, the following metrics were designed to measure users' performance.
`Average Response Time`: a measure of how quickly a user can identify and engage targets.
`Shot Accuracy`: a measure of the percentage of successful shots to the number of targets spawned.
`Target Accuracy`: a measure of the percentage of successful shots taken to a number of targets hit.
Fine Motor Test: designed to test users' perception of depth and eye to hand coordination.
`Average Tracking Time`: measure of average time it took for a player to successfully track, engage, and eliminate a target.
`Accuracy`: measure of the percentage of shots fired to the number of targets hit.

3	**Main Ethical Considerations**

Please give a brief description of the main ethical considerations involved in the study.

Help: Highlight here the main ethical considerations for the study (which may concern, e.g., the type of participants, the sensitive nature of the study, the data collection process, and security-sensitive research) and advise how the main issues will be addressed. If the project is funded, give details here, and whether there are any potential conflicts of interest involved in the study. NB: Section 5 below addresses: recruitment; voluntary participation; consent; and, the right to withdraw. Those details need not also be entered here.

There are no conflicts of interest, and the project is not funded.

4	Human Participants

If the study includes Human Participants (or their data), please give a description of who will be included.

Help:

- Please note this should include sample size/number of participants, whether the project will focus on any particular groups/individuals, if it will include any at risk or vulnerable participants, participants aged 16 years or under, etc. Please also specify the rationale for including/excluding groups of participants.
- If the research involves secondary data not in the public domain, give details in this section.

For the research, physically active individuals that participate in work out activities at a minimum of 3 days a week are desired. Furthermore, the participants are expected to be regular console game players.

We expect a minimum of 4 participants to make the research viable and a maximum of 20 volunteers which will be separated into a control group and intervention groups.

5	Recruitment, Voluntary Participation, Consent and Right to Withdraw

If the study includes Human participants, please give a brief description of the recruitment process, how voluntary participation will be ensured, if (and how) informed consent will be obtained prior to participants taking part in the study, and the right of withdrawal from the research process.

Help:

- This should include clear information on how participants will be identified, approached and recruited; whether the study will include any covert research or deliberate deception; whether help is required from a third party/ gatekeeper to access participants; what information will be given to participants, etc.
- If expenses or any incentives are to be offered to participants, give full details.
- If research involves students, colleagues and/or other employees then specify the rationale for this and how issues of coercion or feelings of obligation will be addressed.
- If data is held on participants, research using that data may require permission from the participant.
- Regarding withdrawal from the study, discuss the different stages/dates a participant could withdraw or withdraw their data, and how they could do this.

The poster below will be printed and placed around the ATU Campus. It will also be sent through email to all students registered on Galway Campus. The poster contains an electronic form that volunteers will be prompted to fill out when scanning the QR code from the poster or accessing the link https://forms.office.com/e/EmhfampDUT.

Once volunteers register interest the email address will be kept as stated in the form and then we will contact them through their registered email.

6	**Risks and Benefits**

Please give a brief description of how, when and where the research will take place and whether there are any risks and/or benefits involved.

Help:

- This should include information on what participants will be required to do, the rationale for this and the level of risk involved.
- When considering risks, please refer to risks to the participants (e.g., for research in sensitive areas, where there is a balance of power), the researcher, any other parties to the research; and also any health and safety issues for anyone involved (e.g., for lone researchers carrying out fieldwork).

The research is a software project that analyses data and does not involve the volunteers performing any physical activity other than does in their normal routine.

Volunteers are expected to wear a biometric data monitoring device that records physical activities.

The volunteers are expected to play a game that will be installed on their computer. On average the game is expected to last for 15 minutes.

Data from the game and device monitor are collected during gameplay and stored in a secured database.

Regression and Correlation analysis are done using these data.

7	Personal Data, Anonymity and Confidentiality

Please specify what type of information/data will be collected/analysed and the source(s). In addition, specify if and how the anonymity of participants will be ensured, and information be kept confidential.

Help: This should include information on whether new information/data are being collected or uses data that are already in the public domain; whether the data includes personal data; whether the data includes sensitive personal data e.g. health data, sexual orientation, race, religion; how the data will be processed and stored; who will have access to it; who it will be shared with; how long data will be retained; how it will be destroyed; the Data Protection requirements for any sensitive personal data, etc. In addition, include whether there may be any requirements for disclosure of information to other parties due to professional practice or legal reasons. If there are limits to confidentiality, explain clearly how the participants would be advised about these limits and possible outcomes.

There are basically two categories of data to be collected which are:

- Game test data that are generated during gameplay.
- Biometric data is obtained from a biometric data monitor (watch).

Specifically, the data to be collected from the Biometric data monitor are exhaustively listed below:

- Maximum Heart Rate
- Average Heart Rate
- Active Steps
- HRV
- Sleep (hours)

To get a volunteer setup for data collection, the volunteer will have to register with the activity monitor manufacturer (i.e. Polar Flow). Such information like height, weight, etc will be taken to ensure the accuracy of data being collected by the activity monitor as advised in their documentation. These data are not of interest to the research and will be stored with the Activity monitor manufacturers.

Volunteer's data collected will be stored in a Firestore database through Firebase which is a secured Backend-as-a-Service Cloud service that provides encryption, security, and availability. Individual user-specific data are stored with the alias/username they choose to register with.

** It is important to note that no data will not be traceable to any individual as their data will be annotated with their chosen alias/username.

8	Reporting and Dissemination

Please give details of the planned dissemination and specify if the findings from the research will be published and whether any permission is required for this.

Help: This should include information on the methods of dissemination (e.g., dissertation/thesis) and/or what will be published and where (research papers, conference presentations). Specify if any permission is needed (e.g., from participants, clients, gatekeepers, etc.) prior to publication, and whether there are any potential issues relating to Intellectual Property Rights when creating or using materials.

As indicated in section A. There are no plans at the moment to publish the research.

9	Location of research

Will the research take place outside of Ireland?

YES ☐ **NO** ☐ **If yes, give details below.**

Help: If yes, please specify where the research will take place. Research must comply with the laws of the country where it is taking place and also comply with local Data Protection and Intellectual Property legislation: you must confirm that your research is compliant with local requirements and how you have ascertained this. Advise if the project requires ethical approval in-country and how this has been ascertained. If approval is required, a copy of this should be included in the application or details of the process of how it will be obtained. Please make reference to insurance and indemnity cover for the project where relevant.

Note: If data is to be processed or stored outside the EEA contact dpo@atu.ie

NO

10	Collaborative Projects

Is the research a collaborative project (i.e., it involves more than one institution)?

YES ☐ **NO** ☐ **If yes, give details below.**

Help: If yes, please specify the other institutions involved and if ethical approval needs to be / has been given by them. Please also specify what procedures have been put in place to ensure ethical compliance from all partners.

Note: If personal data is being shared between institutions then a data sharing agreement must be in place. Contact dpo@atu.ie

NO

11	Any other permission or external ethical approval required to undertake the project

Please specify if the project requires any other ethical approval or permissions not mentioned previously in this application and how and when these will be obtained.

Help:

- Other permissions: ethical approval does not give the right of access to the Institute's students, staff or the use of Institute premises to carry out research, and you may need to contact an appropriate Institute gatekeeper for agreement to approach potential participants or for the use of premises, so please give details.
- Gatekeepers: permission of a gatekeeper for initial access to participants may be required or to carry out data collection on their premises.
- If the project requires approval from an external ethics committee, this should normally be obtained prior to submitting this application.
- If a Disclosure and Barring Service check is required due to the specific participant group, give details.
- Regarding insurance and indemnity cover, some projects will require individual confirmation of cover. See the Research Ethics Procedures document for more details.

NO

SUPPORTING DOCUMENTATION: what to submit with the application

For projects involving human participants, you must submit, where appropriate, the Participant Information Sheet/s and consent form/s. You must also submit every communication a participant will see or receive. Failure to do so will cause delays to the application.

DECLARATIONS AND SIGNATURES

STUDENT			
I confirm that I will undertake this project as detailed in Part A and Part B of the application. I understand that I must abide by the terms of this approval and that I may not make any substantial amendments to the project without further approval. I understand that research with human participants or their data must not commence without ethical approval.			
Signed	Rodrigo Almeida Otito Mbelu	*Date* *31/10/2023*	

RESEARCH SUPERVISOR RECOMMENDATION FOR STUDENT PROJECT					
I confirm that the committee has considered part A and part B of the application. The project is viable and the student has the appropriate skills to undertake the project. Where applicable, the Participant Information Sheet and recruitment procedures for obtaining informed consent are appropriate and the ethical issues arising from the project have been addressed in the application. I understand that research with human participants must not commence without ethical approval. I recommend this project for approval.					
Name		*Signed*		*Date*	
Comment(s): E.g. if similar research projects have been previously approved.					

LECTURER ON BEHALF OF PROGRAMME BOARD			
I confirm that the project will be undertaken as detailed in stage one and stage two of the application. I understand that I must abide by the terms of this approval and that I may not make any substantial amendments to the project without further approval. I understand that research with human participants or their data must not commence without ethical approval.			
Signed		*Date*	

PROJECTS APPROVED BY THE RESEARCH ETHICS SUB-COMMITTEE					
I confirm that this project was considered by the Taught Programme Research Ethics Committee and has received ethical approval.					
Chair		*Signed*		*Date*	

This form will be retained for the purposes of quality assurance of compliance and audit for THREE years

Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

PARTICIPANT INFORMATION SHEET

Biometric Data Analysis in Digital Game Scenario

We are Rodrigo Almeida and Otito Mbelu and we are students in Bachelor of Science (Honours) in Computing in Software Development at ATU Galway City. We are recruiting volunteers to take part in a research study. The aim of this study is to answer the following research questions:

- Can the user's current physical condition as indicated by their Biometric data, have any direct relationship with their performance in such a gaming scenario?

- Can Biometric and test data help suggest the most suitable settings for different game scenarios?

This participant information sheet outlines what the study involves and what will be required of you if you choose to volunteer to participate.

What is the purpose of this study?

The main hypothesis of this research is to find if there is any correlation between a user's biometric data and their performance in a first-person shooter game scenario.

This research study will try to quantify how biometric data such as Heart Rate Variation, Heart Rate, active steps taken, quality and quantity of sleep, etc. affect select gaming skills like eye-to-hand coordination, fine motor skills and reaction time.

What will be required of you?

For this study, you will be required to wear the Polar watch which is referred sometimes to "Activity Monitoring Device" in this text, a day prior to undertaking a test.

Location
The data collection effort will be carried out at the Gym on the Atlantic Technological University (ATU) Galway campus.
Time:
The following weekly schedule will be available for volunteers at various times that best suit your personal schedule.

- Tuesdays: 12:00 - 13:00, 15:00 - 16:00
- Wednesdays: 12:00 - 13:00, 15:00 - 16:00
- Thursdays: 12:00 - 13:00, 15:00 - 16:00
- Fridays: 13:00 - 15:00

Note: The timing is open for adjustment to suit you

Results from the test will be paired with their biometric data for further analysis.

What will happen to the information that is collected about me?

The information gathered from the study will be handled in complete confidence and cannot be traceable back to you. These data are stored in an encrypted database system with everything relating to a user linked with the alias (username) they choose to register with. When the study is finished, information will be kept on the researcher Luke Smyth.

What are the benefits?

The benefit of taking part in this research is that if such a correlation is found between your biometric data and your performance in the game you can use the results to improve your own performance when setting up in any First-shooter game. You may benefit by taking part in seeing how a final year project is run you may take some valuable knowledge and apply it to your own studies if applicable.

What are the risks?

There will be no risks as this study is based on the volunteer's normal activities and playing a game in a controlled environment.

What if you change your mind during the study?

Please be aware that individuals who volunteer to take part in the study have the right to withdraw from the project at any time without the need to provide a reason or notice period. Should you feel at any stage that you want to stop taking part in the study, then this is dealt with in a sensitive and confidential manner.

What happens at the end of the study?

At the end of the study, the data gathered will be disseminated for the purposes of a written academic report and submitted to ATU for assessment purposes. A summary of the results can be provided to you, upon request. All data will be held securely, in line with GDPR regulations, for a period of ten years, as per the ATU records retention policy.

What if you have more questions or do not understand something?

It is important that you feel completely at ease during the research. If you do not understand any aspect of the research, please contact us or my research supervisor to discuss any questions that you might have. Alternatively, you may contact the Head of Department, if you wish to speak to someone independent from the study.

Research Supervisor: Dr. Damien Costello

Head of Department: Dr. Gareth Roe

Thank you for taking the time to read this. I would be grateful if you would consider participating in this study.

Yours sincerely,

Student Name: Rodrigo Almeida, Otito Mbelu

Student Signature:

Date: [*Insert date of circulation of this document here*]

Contact Details: G00377123@atu.ie or G00397738@atu.ie

Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

PARTICIPANT CONSENT FORM

Effects of Post-Action potentiation on sprint performance

Please tick the appropriate boxes

	YES	NO
1. I have read the Participant Information Sheet for this study and understand what's involved.	☐	☐
2. I have been given the opportunity to ask questions about the study.	☐	☐
3. I agree to take part in the study under the conditions set out in the Participant Information Sheet.	☐	☐
4. I understand that my taking part is voluntary and that I can withdraw my consent at any time before my data is de-identified or amalgamated with other data.	☐	☐
5. I understand that, in any report on the results of this study, my identity shall remain anonymous.	☐	☐

Signature of Participant: **Date:**

Signature of Student: **Date:**

Appendix D: Participant Withdrawal Form

PARTICIPANT WITHDRAWAL FORM

Reference Number:

Participant name or Study ID Number:

Title of Project: Biometric Data Analysis in Digital Game Scenario

Name of Principal Investigator: Rodrigo Almeida and Otito Mbelu

Name of the person to whom this form should be submitted: Gary Flynn

Participant to complete this section. Please initial one of the following boxes:

1. I confirm that I wish to withdraw from the study before data collection has been completed and that none of my data will be included in the study.	
2. I confirm that I wish to withdraw all of my data from the study before data analysis has been completed and that none of my data will be included in the study.	
3. I confirm that although the results of the study have already been produced and cannot change, I wish to be forgotten and that all of my personal data is deleted from verification records maintained by the university about the study. I understand that this means that only those data identifying me will be deleted.	

Your name is required to verify that you have withdrawn your data from the study as specified above. In the case of (3), above, we will need to retain this form until…….

It may be necessary to share this information with internal examiners, external examiners, and / or journal editors for the purposes of verification of findings and tracing results of studies to the raw data used.

This form will be stored securely until ……, when it will be destroyed, and will not be shared with anyone else.

Signature of participant:	Date:
Signature of person who will ensure that the stated data have been deleted:	Date:

Microsoft Form for Volunteer Recruitment

Research Participant Needed

Join our study exploring the link between biometric data and gaming performance. Help us understand how your body's responses impact your gaming abilities. Your input is crucial for our research. Thank you for participating!

* Required

* This form will record your name, please fill your name.

1. How many times a week do you exercise? (Select one option) *

- ◯ 1-3 times/week
- ◯ 3-5 times/week
- ◯ 5+ times/week
- ◯ I don't exercise

2. What is the average duration of your exercise sessions? *

- ◯ 30 minutes/session
- ◯ 45 minutes/session
- ◯ 60 minutes/session
- ◯ 75 minutes/session
- ◯ 90 minutes/session
- ◯ 90+ minutes/session

3. Do you own an activity tracker (E.g. fitness watch)? (Select one option) *

- ◯ Yes
- ◯ No

4. Do you play video games? (Select one option) *

- ◯ Yes
- ◯ No

5. What gaming platform do you use most frequently? (Select one or more options) *

- ☐ PC
- ☐ Xbox
- ☐ PlayStation
- ☐ Other

6. What kind of games do you enjoy the most? (Select one or more options)

- ☐ First Person Shooter
- ☐ Third Person Shooter
- ☐ Soccer
- ☐ Car Racing
- ☐ Other

7. What brand is your activity tracker? *

- ◯ Polar
- ◯ Garmin
- ◯ Apple
- ◯ Fitbit
- ◯ Other

8. I consent to be contacted by the researchers about my participation in this project and my data might be kept for the purpose of this research and will not be shared with any third parties.
(You have the right of withdrawal at anytime, by emailing G00377123@atu.ie) *

- ◯ Yes
- ◯ No

Bibliography

[1] Mihai Liviu Despa. Comparative study on software development methodologies. *Database Systems Journal*, 5(3), 2014.

[2] Emily Ortega and CJK Wang. Pre-performance physiological state: Heart rate variability as a predictor of shooting performance. *Applied psychophysiology and biofeedback*, 43:75–85, 2018.

[3] Marek Malik, J Thomas Bigger, A John Camm, Robert E Kleiger, Alberto Malliani, Arthur J Moss, and Peter J Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3):354–381, 1996.

[4] James McNames and Mateo Aboy. Reliability and accuracy of heart rate variability metrics versus ecg segment duration. *Medical and Biological Engineering and Computing*, 44:747–756, 2006.

[5] Daniel Bonnar, Benjamin Castine, Naomi Kakoschke, and Gemma Sharp. Sleep and performance in eathletes: for the win! *Sleep Health*, 5(6):647–650, 2019.

[6] Polar. Polar vantage v user manual, 2324.

[7] AWS. Aws, 2024.

[8] Angular Components.

[9] Angular Modules.

[10] Angular Services.

[11] EJS. Ejs, 2024.

[12] Helder Da Rocha. *Learn Chart. js: Create interactive visualizations for the web with chart. js 2*. Packt Publishing Ltd, 2019.

[13] News. *Significance*, 9(4):2–3, 08 2012.

[14] Scikit-learn. Scikit-learn, 2024.

[15] Bo Pang, Erik Nijkamp, and Ying Nian Wu. Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics*, 45(2):227–248, 2020.

[16] Keras. Keras, 2024.

[17] `https://nodejs.org/en/`. Node.js, Accessed: 23 January 2024.

[18] Stefan Tilkov and Steve Vinoski. Node.js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83, 2010.

[19] Inc. (2021) npm. npm public registry, 2021.

[20] Mike Cantelon, Marc Harter, TJ Holowaychuk, and Nathan Rajlich. *Node. js in Action*. Manning Greenwich, 2014.

[21] David Gonzalez. *Developing Microservices with node. js*. Packt Publishing Birmingham, UK, 2016.

[22] Olivier Aumage, Gabriel Antoniu, Luc Bougé, Vincent Danjean, and Raymond Namyst. Getting started with pm2. *LIP, ENS-Lyon*, 2001.

[23] Dirk Merkel et al. Docker: lightweight linux containers for consistent development and deployment. *Linux j*, 239(2):2, 2014.

[24] Inc. Docker. Docker hub, 2021.

[25] Will Reese. Nginx: the high-performance web server and reverse proxy. *Linux Journal*, 2008(173):2, 2008.

[26] Paul DuBois. *MySQL*. Addison-Wesley, 2013.

[27] Evan Hahn. *Express in Action: Writing, building, and testing Node. js applications*. Simon and Schuster, 2016.

[28] Mocha. Mocha, 2024.

[29] JWT. Jwt, 2024.

[30] Alan Inglis, Andrew Parnell, and Catherine B Hurley. Visualizing variable importance and variable interaction effects in machine learning models. *Journal of Computational and Graphical Statistics*, 31(3):766–778, 2022.

[31] Michael A Poole and Patrick N O'Farrell. The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, pages 145–158, 1971.

[32] Yanli Liu, Yourong Wang, and Jian Zhang. New machine learning algorithm: Random forest. In *Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3*, pages 246–252. Springer, 2012.

[33] Olvi L Mangasarian and David R. Musicant. Robust linear and support vector regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):950–955, 2000.

[34] Yunsheng Song, Jiye Liang, Jing Lu, and Xingwang Zhao. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, 251:26–34, 2017.