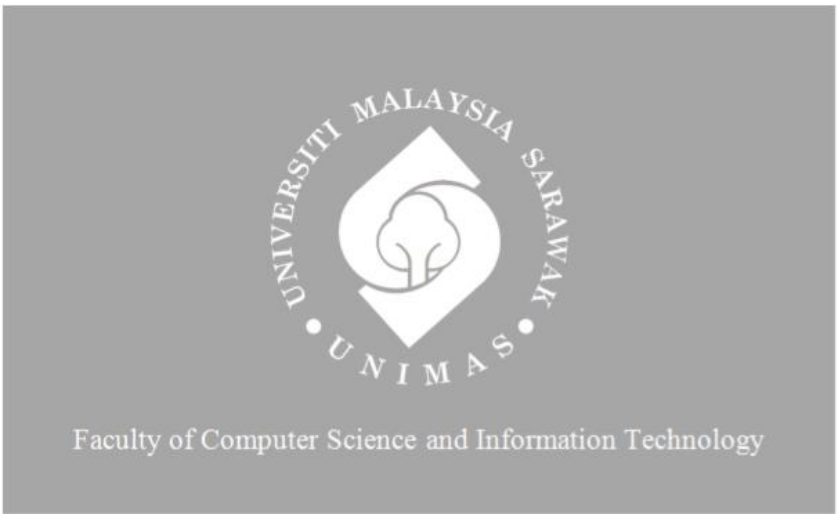


Predicting Cheaters in PlayerUnknown's Battlegrounds (PUBG) using Random Forest Algorithm

NURIN ALYA BINTI HARIS

Bachelor of Computer Science with Honours

(Information Systems)

2023

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE Predicting Cheaters in PlayerUnknown's Battlegrounds (PUBG) using Random Forest Algorithm

ACADEMIC SESSION: 2022/2023

NURIN ALYA BINTI HARIS
(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

☐	CONFIDENTIAL	(Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
☐	RESTRICTED	(Contains restricted information as dictated by the body or organization where the research was conducted)
☑	UNRESTRICTED	

(AUTHOR'S SIGNATURE)

Validated by

DR SHAFRIE ABD RAHMAN
Senior Lecturer (DS52)
Computational Science Programme
Faculty of Computer Science & Information Technology
Universiti Malaysia Sarawak

(SUPERVISOR'S SIGNATURE)

Permanent Address

No. 143, Kampung Segedup,
Jalan Batu Kawa Kem, 93250,
Kuching, Sarawak.

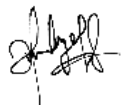
Date: 28/06/2023

Date: 30/6/23

Note * Thesis refers to PhD, Master, and Bachelor Degree
** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

DECLARATION OF ORIGINALITY

I hereby certify that this research altogether with all of its contents is none other than that of my own work, with consideration of the exception of research-based information and relative materials that were adapted and extracted from other resources, which have been evidently quoted or stated respectively.



.....

Signed,

NURIN ALYA BINTI HARIS 28th June 2023

Faculty of Computer Science and Information Technology,

Universiti Malaysia Sarawak

ACKNOWLEDGEMENT

My first and foremost gratitude goes out to my supervisor, Dr. Shapi-ee bin Abd Rahman, as well as my co-supervisor, Encik Ahmad Alif bin Kamal, for their valuable support with this project. Many thanks to both, and my genuine appreciation for the learning opportunities provided by these individuals, who continue to mould each student to surpass their limitations and strive for the best. Their insightful perspective has been, and will continue to be, a driving force behind my progress in the fields of computer science and e-sports. It was such a tremendous pleasure, and the experiences we had from viewing and discussing this artwork with them were priceless.

My accomplishment in completing this project would not have been possible without the assistance of my classmates, friends, and study partners, specially, Syahazwani Nurain, Siti Rubiah, Nur'azra Alia, Nurfazlina, and Dayang Afiqah Liyana. I am indebted to all of your support and encouragement, as well as for pushing me further than I thought I was capable of going.

I would want to express my appreciation to both of my parents and to the rest of my family for the innumerable times that they assisted me throughout this journey in my education. They are the reason I continue to face all of the struggles, pains, and hardships. Their support when times were tough is deeply appreciated and duly noticed, and they are the source of my motivation.

Last but not least, I would like to extend my most heartfelt thanks to my partner, Muhammad Ahnaf Syazwan, for all of his support over the entirety of my final year, both on and off the water. When things were tough, the fact that he was there and willing to provide support and encouragement was a wonderful comfort and a huge relief.

ABSTRACT

PlayerUnknown's Battlegrounds (PUBG) has become a massively popular online video game, attracting a significant number of players of all ages. However, the prevalence of cheating presents an important challenge to maintain a fair gaming environment. This study aims to address this issue by developing a reliable prediction model to identify potential cheaters in PUBG matches. The research methodology involves the collection of a comprehensive data set from Kaggle containing a variety of gameplay features. Patterns and relationships between input variables and cheating behaviours are analysed through the application of supervised learning techniques, specifically a classification model. The primary goal is to use the Random Forest algorithm, an effective machine learning technique, to predict instances of cheating based on the behavioural patterns of participants. Utilising the selected features, the Random Forest model is then trained to produce a robust prediction model. The evaluation of the prediction model demonstrates its accuracy, precision, recall, and F1 score, demonstrating its capacity to identify potential cheaters in PUBG encounters. The developed prediction model obtains an impressive accuracy of 95.7%, demonstrating its reliability in distinguishing cheaters from non-cheaters. The outcomes of this research contribute to the advancement of cheat detection mechanisms in online gaming. The developed prediction model can be integrated into existing systems to improve cheat detection capabilities and promote an enjoyable and fair gaming experience for all PUBG participants.

ABSTRAK

PlayerUnknown's Battlegrounds (PUBG) telah menjadi permainan video dalam talian yang sangat popular, menarik sejumlah besar pemain dari semua peringkat umur. Walau bagaimanapun, kelaziman penipuan memberikan cabaran penting untuk mengekalkan persekitaran permainan yang adil. Kajian ini bertujuan untuk menangani isu ini dengan membangunkan model ramalan yang boleh dipercayai untuk mengenal pasti potensi penipu dalam perlawanan PUBG. Metodologi penyelidikan melibatkan pengumpulan set data komprehensif daripada Kaggle yang mengandungi pelbagai ciri permainan. Corak dan hubungan antara pembolehubah input dan tingkah laku menipu dianalisis melalui aplikasi teknik pembelajaran yang diselia, khususnya model klasifikasi. Matlamat utama ialah menggunakan algoritma Random Forest, teknik pembelajaran mesin yang berkesan, untuk meramalkan kejadian penipuan berdasarkan corak tingkah laku peserta. Menggunakan ciri yang dipilih, model Random Forest kemudiannya dilatih untuk menghasilkan model ramalan yang mantap. Penilaian model ramalan menunjukkan ketepatan, ketepatan, ingatan semula dan skor F1, menunjukkan keupayaannya untuk mengenal pasti penipu berpotensi dalam pertemuan PUBG. Model ramalan yang dibangunkan memperoleh ketepatan yang mengagumkan sebanyak 95.7%, menunjukkan kebolehpercayaannya dalam membezakan penipu daripada bukan penipu. Hasil penyelidikan ini menyumbang kepada kemajuan mekanisme pengesanan cheat dalam permainan dalam talian. Model ramalan yang dibangunkan boleh disepadukan ke dalam sistem sedia ada untuk meningkatkan keupayaan pengesanan cheat dan menggalakkan pengalaman permainan yang menyeronokkan dan adil untuk semua peserta PUBG.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **i**

ABSTRACT ..... **ii**

ABSTRAK ..... **iii**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

LIST OF EQUATIONS ..... Error! Bookmark not defined.

CHAPTER 1: INTRODUCTION **1**

1.1 Project Title 1

1.2 Introduction 1

1.3 Problem Statement 3

1.4 Scope 5

1.5 Aim and Objectives 5

1.5.1 Aim 5

1.5.2 Objectives 6

1.6 Brief Methodology 6

1.6.1 Data Collection 6

1.6.2 Data Pre-Processing 6

1.6.3 Training/Learning/Testing 7

1.6.4 Evaluation 7

1.7 Significance of Project 8

1.8 Project Schedule 9
1.9 Expected Outcome 10
1.10 Thesis Outline 10
1.11 Summary 12
CHAPTER 2: LITERATURE REVIEW 13
2.1 Introduction 13
2.2 Background Study 13
2.2.1 PlayerUnknown's Battlegrounds (PUBG) 13
2.2.2 Cheating in Online Gaming 22
2.2.3 Artificial Intelligence 34
2.2.4 Machine Learning 35
2.2.5 Types of Machine Learning 36
2.2.6 Random Forest Algorithm 38
2.3 Related Works 39
2.4 Conclusion 40
CHAPTER 3: RESEARCH METHODOLOGY 41
3.1 General 41
3.2 Software Requirements 41
3.3 Research Methodology 42
3.3.1 Data Collection 44
3.3.2 Data Pre-Processing 46

3.3.3 Feature Selection 46
3.3.4 Feature Engineering 48
3.3.5 Random Forest Model 48
3.3.6 Performance Evaluation 51
3.4 Conclusion 54
CHAPTER 4: IMPLEMENTATION AND TESTING 55
4.1 Introduction 55
4.2 Data Collection 56
4.3 Data Pre-Processing 57
4.4 Outlier Detection 58
4.4.1 Kills without movement 58
4.4.2 Anomalies in Aim: More than 45 kills 59
4.4.3 Anomalies in Aim: 100% Headshot Rate 61
4.4.4 Anomalies in Aim: Longest Kill 63
4.4.5 Anomalies in Supplies: Weapons 66
4.4.6 Anomalies in Supplies: Heals 68
4.5 Feature Selection and Feature Engineering 70
4.5.1 Feature Selection 71
4.5.2 Feature Engineering 72
4.6 Model Building 73
4.7 Conclusion 77

CHAPTER 5: RESULTS **78**

5.1 Introduction 78

5.2 Results for Feature Selection and Feature Engineering 78

5.3 Performance Evaluation 84

5.2.1 Confusion Matrix 86

5.3 Conclusion 89

CHAPTER 6: CONCLUSION AND FUTURE WORKS **90**

6.1 Introduction 90

6.2 Summary and Conclusion 91

6.3 Limitations 93

6.4 Future Works 94

REFERENCES **96**

APPENDIX **112**

LIST OF TABLES

Table 3. 1 Feature name and description 45

Table 3. 2 Confusion matrix and its description 51

Table 4.1 Potential Cheaters Data from Outlier Detection……...68

Table 5. 1 Feature Sets with Correlation Value 0.6 and above. 80

LIST OF FIGURES

Figure 1. 1 Project Schedule for FYP1 Semester 1 2022/2023 9

Figure 1. 2 Project Schedule for FYP2 Semester 2 2022/2023 9

Figure 2. 1 A player is in the plane waiting to make a drop. 15

Figure 2. 2 Gathering of Arms and Ammunition 15

Figure 2. 3 Survival and Combat Safe Zone 16

Figure 2. 4 PUBG Rewards 17

Figure 2. 5 PUBG Rank Tier-Wise Ban Chart 32

Figure 2. 6 Banned Cheats Chart 33

Figure 2. 7 Process of Machine Learning 36

Figure 2. 8 Types of Machine Learning 36

Figure 3. 1 Research Methodology Flow 43

Figure 3. 2 Example to create correlation matrix in Google Colab 47

Figure 3. 3 Example of coding to generate heatmap of the correlation matrix. 47

Figure 3. 4 Diagram of classification task in Random Forest 48

Figure 3. 5 Example of coding in Google Colab to measure performance evaluation 53

Figure 4. 1 Dataset from Kaggle Files 56

Figure 4. 2 Source Code for Data Pre-Processing 57

Figure 4. 3 Source Code for new feature 'totalDistance' and kills without movement. 59

Figure 4. 4 Source Code for Distribution of Kills 60

Figure 4. 5 Distribution of Kills 60

Figure 4. 6 Source Code to determine players with more than 45 kills. 61

Figure 4. 7 Source Code for Headshot Rate 62
Figure 4. 8 Distribution of Headshot Rate 62
Figure 4. 9 Source Code to determine players with 100% headshot rate. 63
Figure 4. 10 Source Code for Distribution of Longest Kill 64
Figure 4. 11 Distribution of Longest Kill 64
Figure 4. 12 Source Code players with longest kill greater or equal to 100 meters 65
Figure 4. 13 Source Code for Weapons Acquired 66
Figure 4. 14 Distribution of Weapons Acquired 66
Figure 4. 15 Source Code to determine player with 50 or more weapons. 67
Figure 4. 16 Source Code for Distribution of Heals 68
Figure 4. 17 Distribution of Heals 68
Figure 4. 18 Source Code to determine players using 30 or more heals. 69
Figure 4. 19 Data of Potential Cheaters from Outlier Detection 70
Figure 4. 20 Source Code for Feature Selection 71
Figure 4. 21 Source Code for Feature Engineering 72
Figure 4. 22 Source Code for splitting data into training and testing 73
Figure 4. 23 Source Code for Random Forest Model 75

Figure 5. 1 Heat Map of Correlation Matrix 78
Figure 5. 2 Source Code for Performance Evaluation 84
Figure 5. 3 Output of Performance Evaluation 84
Figure 5. 4 Source Code Confusion Matrix 86
Figure 5. 5 Heat map of Confusion Matrix 87
Figure 5. 6 Data for Predicted Cheaters 88
Figure 5. 7 Data for Predicted Non-Cheaters 88

LIST OF EQUATIONS

Equation 3. 1 Formula for Calculating Gini Impurity 49

Equation 3. 2 Formula for Calculating Entropy 50

Equation 3. 3 Formula for Accuracy 52

Equation 3. 4 Formula for Precision 52

Equation 3. 5 Formula for Recall 52

Equation 3. 6 Formula for F-measure 52

CHAPTER 1: INTRODUCTION

1.1 Project Title

Predicting Cheaters in PlayerUnknown ‘s Battlegrounds (PUBG) using Random Forest Algorithm

1.2 Introduction

Esports, which are also referred to as electronic sports, is a type of organised sports competition format that are based on video games (Russ, 2019). Professional gamers compete in esports either individually or as teams with the objective of winning the champion reward in their respective categories. Multiplayer online battle arena (MOBA), first-person shooter (FPS), fighting, card games, battle royales, and real-time strategy (RTS) are the types of video games that are played the most frequently in esports (Goud, 2022). Esports competitions have always been limited to amateur players. On the other hand, its popularity skyrockets in the latter part of the 2000s, coinciding with an increase in the number of professional gamers who take part in it and the number of viewers who watch it via live streaming (Chavers, 2019). According to (Hussein et al., 2017), in 2018, esports competitions attracted a total audience of 380 million people, the vast majority of whom were fans of the industry. The growth that began the year before and extended over into the following year was continued in 2022. The worldwide esports viewership increased by 8.7% year on year in 2022, bringing the total number of viewers to 532 million. This marked another year of tremendous audience growth for esports (Tristao, 2022). In addition, the report also stated that the recently released game, PlayerUnknown’s Battlegrounds (PUBG) was a major contributor to the expanding global growth in esports.

One of the most significant problems in the world of esports, however, is significantly dependent on the answer to the following fundamental question: "How fair is it for all of the players?" Genuine players will become dissatisfied and stop participating in a competitive game if the rules are unfair. The prize pools for esports tournaments can reach as high as $40 million as of November 2022, with the money going to the winners (Gough, 2022). Some players are even willing to adopt illegal strategies in order to gain an advantage in the competition because the prize pools are so high. To get an unfair advantage in competitive battles, some players attempt to modify the game's controls in order to win which is unfair to many in competitive matches. As a corollary, diverse range of strategies have been employed to ensure that their games are as equitable as is humanly possible. One of these steps is to identify any competitors who are cheating or hacking and remove them from the competition.

Cheating occurs in video games for a plethora of reasons. For instance, it can be used to simply win a game when a player otherwise cannot, to have pleasure in circumstances where some players gain satisfaction from the use of cheats, and many other things as well, depending on the game and the context in which it is employed. Nevertheless, cheating is often a very detrimental element that can harm the accomplishment of a game. Consequently, this paves the way for the development of anti-cheat mechanisms within video games, which actively prevent and prohibit players from being able to acquire an unfair edge over other players.

The gaming industry, much like any other industry, collects data about its various players in order to assess the effectiveness of those players. Let's try to grasp this by utilizing PUBG, also known as Player Unknown's Battlegrounds, which is now one of the most well-known and widely played games on the internet. PUBG boasts an average of 30 million players that are actively playing the game each and every day (Curry, 2022). It accomplishes this by gathering information from the gamers and then dividing them up into the different segments, cheaters and non-cheaters. However, manually classifying 30 million players is physically

impossible in this day and age. The concept of machine learning (ML) is developed to solve this problem. An application of artificial intelligence (AI), known as machine learning (ML), has the capability to comprehend a collection of data all by itself and build models based on the data sample in order to come up with a conclusion. The use of technologies such as machine learning, which can provide crucial data that can be used to develop strategies and tactics, has become increasingly advantageous in the process of forecasting the reliable outcome of esports games. Therefore, ML makes it possible to evaluate vast amounts of data, derive conclusive results on all those data, and classify them cheaters into separate categories.

1.3 Problem Statement

The query "Is the player genuine or fake?" is consistently ranked as the most popular inquiry received from all online gamers. The fact that online gamers are frequently unsupervised when they play games makes it simple for those who are eager to cheat to gain an advantage over their competitors in the game. According to the findings of the survey, 60% of online gamers believe that their experience playing multiplayer games has been adversely affected by other players cheating on multiple occasions, 18% of respondents stated that this occurs more frequently, and another 8% stated that it occurs continually. In addition, the findings of the research also revealed that 12% of gamers admitted that they have cheated by using tools that were provided by a third party (Blaukovitsch, 2022). It is highly possible that the actual numbers are likely to be greater. In terms of cheating, PlayerUnknown's Battlegrounds (PUBG) is analogous to other popular online first-person shooter games, in that the use of cheating tools is not only prevalent but also problematic and penalised. In October 2021, PUBG permanently prohibited 2.5 million accounts from Battlegrounds Mobile India (BGMI) and temporarily banned an additional 706,319 accounts.. According to a report by Tencent, the gaming engine behind PUBG, 46% of banned accounts were found to be using auto-aim and x-ray vision hacks,

18% were found to be using speed hacks, 20% were found to be modifying their area damage and the character model, and 16% of banned accounts were banned for some other reason (Goud, 2022). With such a large userbase, it is physically difficult to keep track of each participant, which is especially problematic due to the presence of hackers in the game. The security team behind PUBG constantly monitors massive amounts of data, with an average of over 10 million reports and this is carried out to identify improbable occurrences that may have taken place during the matches (Chavers, 2019). Due to the circumstances, the current research investigates the impact that machine learning algorithms have on the gaming industry's ability to predict players who cheat in online games. This is a binary classification problem as there are two classes in this study: cheaters and non-cheaters. In this research, a substantial dataset consisting of 29 variables requires the application of a supervised learning strategy through the utilisation of a classification model that is capable of being helpful in prediction of patterns that exist between the input variables and the output variables. As a result, the purpose of this research is to apply Random Forest model in an effort to determine whether PUBG players are likely to cheat during a match based on a number of attributes with the player's behavioural pattern. Thus, Random Forest algorithm, which is a classical in the field of machine learning, is going to be the method that we use to solve this issue. A machine learning (ML) system gathers information from players and adapts itself to the needs of the game. If we continue to train it on data from players, then it will always be updated with the most recent data. Additionally, as the game continues, the developers will implement various upgrades and patches. Most significantly, a ML system uses data about the game that the developers of off-the-shelf cheat detection technologies do not have access to (modl.ai, 2022). This is the primary reason why ML cheat detection is more difficult to fool. As a result, this decision was made to investigate the possibility of finding a solution to this problem that may be used by e-sport contest managers, server moderators, or anybody else that would need this to eliminate cheaters.

1.4 Scope

The scope of this study is restricted to specific instances of online gaming, specifically Player Unknown's Battleground (PUBG). The system will not be integrated into the game; rather, it is more analogous to an application developed by a third party that may be used to increase the ability of server moderators to identify possible cheaters. Consequently, this research will make use of the PUBG game statistics that can be found on the Kaggle website. These data have previously been separated into a training set and a testing set. For the entirety of this study, it will only be using the training set because our primary goal is to establish which characteristics are the most significant. The presented dataset has the form (4446966, 29), which implies that the 29 distinctive attributes were extracted from the 4446966 separate instances of different players engaging in a variety of matches. Each set has 29 different attributes, and the dataset of PUBG placement prediction will be utilised in order to arrive at the cheater prediction. It is expected that machine learning will make the system more robust over time and with additional data. Since there is a problem with cheaters in the community, it is viewed as the sensible thing to do to make every effort to stop these cheaters when they are competing in competitive esports games.

1.5 Aim and Objectives

1.5.1 Aim

This study aims to address the issue of cheating behaviours in PUBG by using Random Forest algorithm to precisely identify and predict potential cheaters.

1.5.2 Objectives

1. To develop a machine learning model using Random Forest to detect potential cheaters in PUBG.
2. To identify patterns and behaviours that are highly correlated with cheating behaviour in PUBG.
3. To evaluate the performance of the proposed Random Forest model in identifying and predicting cheating behaviour.

1.6 Brief Methodology

Design and development were selected as the research methodology for this study since a model must be created in order to achieve the overall objective of detecting cheats using machine learning in PUBG. The model was modified and evaluated iteratively until the desired outcome was achieved. This study's methodology can be divided into several distinct parts:

1.6.1 Data Collection

Kaggle was the source from which we obtained the PUBG dataset that was utilised in this research. It includes a vast amount of de-identified PUBG game statistics and is formatted so that each row provides the post-match statistics for a single player. PUBG placement prediction dataset was chosen for our model's development. The given dataset is in the form (4446966, 29), which indicates that the 29 unique features were collected from the 4446966 instances of different players participating in various matches.

1.6.2 Data Pre-Processing

Data pre-processing is essential for preparing a dataset for analysis and modelling. It involves dividing data into training and testing sets in order to enhance the performance of machine learning models. The dataset is imported and saved as a.csv file for convenience of access. In

addition, missing data and outliers are eliminated to reduce biases, improve the precision of statistical analyses, and boost the performance of machine learning models. These pre-processing phases guarantee that the dataset is clean, dependable, and suitable for analysis and modelling, resulting in more reliable research findings.

1.6.3 Training/Learning/Testing

In order to examine the relationships between the features, a cross-correlation matrix will be constructed. By analysing the correlations, we can identify any significant associations or redundancies between features that may affect the performance of the model. Moreover, during the process of feature selection and engineering, any category features that have a minimal effect on the output prediction will be eliminated. This simplifies the dataset and improves the model's performance by highlighting the most informative features. Following the completion of feature selection, the Random Forest algorithm will be selected for model construction. Random Forest is an effective algorithm for machine learning that incorporates the predictions of multiple decision trees. It is renowned for its capacity to manage intricate datasets and generate accurate predictions. Random Forest captures complex relationships between features and the objective variable by leveraging the ensemble of decision trees. Before the final model is constructed, the dataset will be divided into training and testing sets. The training set will be used to train the Random Forest model by exposing it to labelled data and allowing it to discover the underlying relationships and patterns. The testing set, on the other hand, will function as an independent dataset to evaluate the predictive accuracy and performance of the model on unseen data.

1.6.4 Evaluation

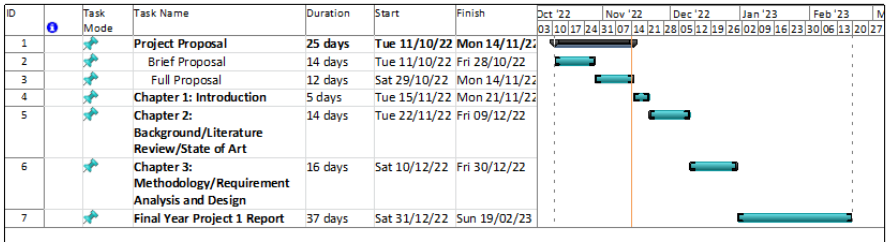
The efficiency of the model will be evaluated. It has been determined that the confusion matrix is the most effective method for representing the accuracy of our classification model,

and it is also the most widely used metric for assessing classification problems. This research endeavour will make use of this methodology.

1.7 Significance of Project

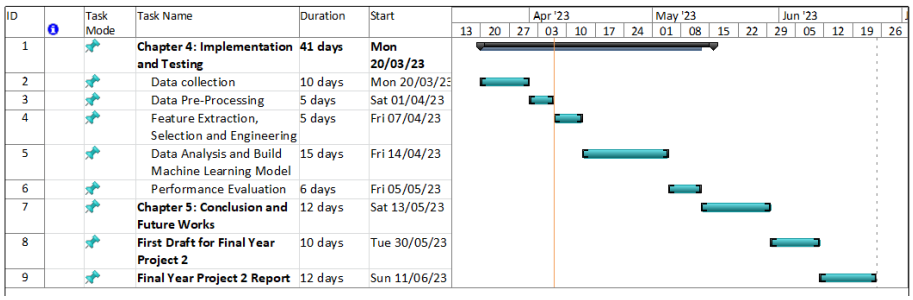
The project's significance extends beyond its immediate effect on game integrity. By effectively identifying and preventing cheating, the initiative contributes to the preservation of a competitive and balanced gaming environment. In addition to diminishing the experience for honest players, cheating also harms the game's reputation. Consequently, the ability of the project to correctly predict cheaters using the Random Forest algorithm contributes to the game's integrity and fairness. In addition, the initiative demonstrates the application of machine learning techniques in the gaming industry. Using the strength of the Random Forest algorithm, the project illustrates how data-driven approaches can be applied to complex problems such as cheat detection. This is a useful case study for researchers, developers, and practitioners interested in utilising machine learning to improve game security. In addition, the project's findings and insights into cheating patterns provide game developers and administrators with valuable information. Understanding the strategies and methods used by cheaters can aid in the development of more effective anti-cheat systems and proactive measures to prevent future instances of cheating. This information can also inform ongoing efforts to enhance game design and player monitoring techniques, resulting in continuous improvements to game safety. The significance of the project lies in its contribution to the gaming industry by promoting fair play, utilising advanced machine learning techniques, and providing valuable insights for the development of efficient cheat detection systems. The project not only benefits the game and its players by addressing the issue of cheating in PUBG, but it also functions as a steppingstone for future research and advancements in the field of game security.

1.8 Project Schedule



ID	Task Mode	Task Name	Duration	Start	Finish
1		**Project Proposal**	**25 days**	**Tue 11/10/22**	**Mon 14/11/22**
2		Brief Proposal	14 days	Tue 11/10/22	Fri 28/10/22
3		Full Proposal	12 days	Sat 29/10/22	Mon 14/11/22
4		**Chapter 1: Introduction**	5 days	Tue 15/11/22	Mon 21/11/22
5		**Chapter 2: Background/Literature Review/State of Art**	14 days	Tue 22/11/22	Fri 09/12/22
6		**Chapter 3: Methodology/Requirement Analysis and Design**	16 days	Sat 10/12/22	Fri 30/12/22
7		**Final Year Project 1 Report**	37 days	Sat 31/12/22	Sun 19/02/23

Figure 1. 1 Project Schedule for FYP1 Semester 1 2022/2023



ID	Task Mode	Task Name	Duration	Start
1		**Chapter 4: Implementation and Testing**	**41 days**	**Mon 20/03/23**
2		Data collection	10 days	Mon 20/03/23
3		Data Pre-Processing	5 days	Sat 01/04/23
4		Feature Extraction, Selection and Engineering	5 days	Fri 07/04/23
5		Data Analysis and Build Machine Learning Model	15 days	Fri 14/04/23
6		Performance Evaluation	6 days	Fri 05/05/23
7		**Chapter 5: Conclusion and Future Works**	12 days	Sat 13/05/23
8		**First Draft for Final Year Project 2**	10 days	Tue 30/05/23
9		**Final Year Project 2 Report**	12 days	Sun 11/06/23

Figure 1. 2 Project Schedule for FYP2 Semester 2 2022/2023

1.9 Expected Outcome

A machine learning system has been developed that can evaluate player performance using data acquired from Kaggle to identify the issue of cheaters in PUBG mobile and discover a solution to it. The machine learning technique will help to examine the usage patterns of PUBG mobile gamers. We have built a system that is able to identify abnormal gaming patterns or actions that disrupt the normal operations of the game and are caused by hack players. Everyone who competes in the e-sports competition will adhere to the rules of fair play, and those who cheat will be eliminated from the competition.

1.10 Thesis Outline

Chapter 1: Introduction

Chapter 1 provides an overview of the proposed system and its components. The problem statement, objectives of the research, a brief methodology carried out in the project, significance of project, the scope of the project and the project outcome are all included in this chapter. The problem statement outlines the challenges that the existing analysis encounters and explains why it is essential for this project to be revised to incorporate a new method of evaluation. The objectives provide information regarding the overall purpose of the project. The scope sheds light on the methods that were implemented in the project in order to reach a high accuracy rate.

Chapter 2: Literature Review

Chapter 2 presents a discussion on the literature review that was carried out on the existing methods and procedures. This review was based on articles, journeys, proceedings, and research papers. This chapter of the review will explain about the methodologies, features, and classifications that were used to discover the feature datasets. These methods have been