CHAPTER I

INTRODUCTION

1.1   Background

The rise of competitive sports forces players and teams to perform as best as they possibly can if they want to stay in the game. Individual athletes may have to plan out a better training schedule and teams may have to plan out their matches and study their opponents in advance. This forces individual players and teams to apply analytics that can assist their decision-making process.

The use of analytics in sports gave birth to a trend known as sports analytics, where existing data is analyzed using statistical analysis methods and predictive modelling in order to help the decision-making process within sports teams and individual athletes [1]. The sports analytics market has exploded in recent years and is currently worth billions of dollars, with projections indicating that the market will keep growing rapidly in the years to come [2]. The rapid development of machine learning and artificial intelligence technologies played a big part in the wide adoption of sports analytics in many teams. By feeding a data to an algorithm, one can identify patterns and even make predictions to aid them in the future. Algorithms such as SVM [3]–[5] and Random Forest [6], [7] has been proven to be able to work effectively in performing sports analytics. Clustering algorithms, such as K-means can also be used to group players into classes based on their performances [8], [9].

Badminton is one example of a sport that can benefits from sports analytics, with potential applications include score prediction, action recognition, and event

detection where the system attempts to identify specific moments in the sport. Indonesia is widely recognized as one of the leading countries in Badminton alongside the likes of China, Denmark, Japan, and others, with two of its players sitting in the Top 10 list of Badminton players according to BWF rankings as of 9th February 2022. As Indonesia continues to nurture more and more world-class talent to compete in the sport, analytics can be proven useful to help this new talent identify possible weaknesses in their playstyles.

The benefits of sports analytics in team play and individual sports have been well documented and researched throughout the years. However, we observed that there is not enough research on how sports analytics can be applied in Badminton and how it can be used to profile players. We feel that the use of historical and video data to gauge badminton players' performances and profiling them based on existing information to be an area work exploring.

Historical match data displays detailed information about prior matches, including the players involved, their overall rankings on the sport/league, how many points they scored in total, match points, game points, points scored by each player in every game/set, and more. This information can be used to identify how a certain player or team generally perform in matches. One example being a player who tend to perform well at the start of the match but also having difficulties in putting out a satisfying performance on the latter half of the match, in which we can conclude that said player has trouble maintaining a consistent performance throughout the entire match. Other factors that can affect a player's chances of winning includes their swinging styles (lob, backhand, etc.), playstyles (prefer long rallies or quick rallies), and mental state.

Historical match data has some limitations, such as the lack of information regarding an athlete's playstyle and preferred tactics. A good way to obtain playstyle information is using video analysis, where an existing video data is analyzed using computer vision and classification algorithms to identify temporal and spatial events in the video [10]. Video footage can give us direct information about how a badminton athlete generally perform in matches. We can also obtain information such as their dominant hand and preferred strokes (backhand, lob, smash, etc.). There are various software models out there that can be used to perform video analysis, with YOLOv5 being one of the more powerful ones. YOLOv5 is a neural network based real-time object detection system that can be used to detect objects in an image or video [11]. The YOLOv5 model can be trained with a custom dataset, which means that we can train a YOLOv5 model to detect badminton players and the shuttlecock from video data. A software to track an object's frame-by-frame can aid profiling even further by allowing one to analyze an athlete or shuttlecock's movement in detail. Tracker is a software that supports video object tracking [12].

Athlete profiling involves assessing an athlete's performance and playstyle for a particular sport, and use them to obtain a basic 'profile' for the athlete. Said profile can refer to their basic playstyles, abilities, strengths, weaknesses, and more. This can be done in several ways, such as score data analysis and video data analysis. Score data can inform us on how players tend to perform during matches and how they may correlate to their chances of winning a match. Video footage can provide information on their playstyle (favorite stroke types, aggressive/defensive plays, etc.). By identifying the patterns of how one player competes in a match, we

can perform profiling on them and obtain information about their general performances. This information can be helpful during athlete coaching.

The correlation between score data analysis and video analysis when assessing and profiling an athlete's performance may be strong when both methods are used effectively. Score data analysis allows us to look at how an athlete generally performs in official matches as score data is the reflection of an athlete's ability in actual matches. Video analysis lets us observe the athlete's movements and playstyle as well identifying their possible strengths and weaknesses. As score data demonstrates the athlete's performances in actual matches, video analysis may allow us to identify what may causes the athlete to perform as they did on the score results – an athlete's strength such as speed and adaptability may allow them to score many points consecutively and win games or matches, while weaknesses such as poor decision making and movement speed may explain their losses shown on the score data.

While coaching is a crucial part in athlete training, coaches may not be able to identify all gaps and flaws within a player's ability. This is where profiling comes in. In general, athlete profiling (or performance profiling) can be described as the act of identifying an athlete's qualities and what is needed for them to be successful in the sport [13]. The information used to profile an athlete's performance can be physical, psychological, technical, and tactical. The results of profiling can be used by athletes and coaches to help them gain better understanding of an athlete's ability [14]. Profiling can help provide information about an athlete that we may not be aware of by analyzing existing data to find patterns and the correlations between their performances and their results in the sport. An example of potential correlation

in the sport is an athlete's endurance levels may determine their consistency in their performances (better stamina might mean they are able to perform well for longer periods of time). Results of profiling can be used to help athletes and coaches to break down the athlete's general performance in matches, what works best for them, what doesn't work, and what can be improved in future matches. For an example, an athlete's profiling process discovers that they typically perform well on the offensive but falters during defensive plays, leading to a loss when faced against opponents with strong offensive skills. The athlete and coach can use the profiling results to plan intensive training to further improve their defensive skills so that they can stand a better chance in future games.

As the competition in Badminton grows more and more fierce, Indonesia needs to be able to keep up with increasing competition, and the use of sports analytics may prove to be important in achieving their goals. Profiling is one way to identify a player's skills in a sport, and we feel that there needs to be more research done on how athlete profiling can be done in badminton. The use of profiling to identify players' skill levels and help them improve may prove to be useful in the country's quest to remain highly competitive in the sport.

1.2 Problem Identification

The problems surrounding the research topic are:

a. How does machine learning and AI performs for Indonesian badminton athlete profiling?

b. Can match data and dynamic movement data be used to profile badminton athletes?

c. Can SVM, Random Forest, K-means, and Hierarchical Clustering perform profiling from score data well?

d. Can AI be used to analyze dynamic movements?

1.3   Problem Boundaries

The boundaries for this research are:

a. Developing a system that can perform athlete profiling of two Indonesian badminton athletes using historical match data from 2018 until 2021

b. Analyze Indonesian badminton players' playstyles and performances using YOLOv5 object detection results

c. Using Tracker to identify athlete and shuttlecock movement patterns

d. Comparing how SVM and Random Forest performs in classification and regression for historical match data analysis

e. Comparing K-means clustering and Hierarchical Agglomerative Clustering in clustering for historical match data analysis

1.4   Problem Formulation

How can SVM, Random Forest, K-means clustering, and Hierarchial Agglomerative Clustering be used to perform clustering, regression, and classification on badminton historical match data to profile athletes by analyzing the data and perform correlations between variables and an athlete's win/lose results, and can video analysis using YOLOv5 and Tracker be used to analyze the athletes' performances and playstyle?

1.5 Research Purpose

The aim of this research is to develop a method that can perform badminton athlete profiling based on historical match data and video data. In addition to developing a profiling system, this research also aims to compare several distinct algorithms (SVM, Random Forest, K-means, and Hierarchical Clustering) in how they perform in historical match data analysis. To get the full picture of how both algorithms can perform in sports analytics, the algorithms will be compared in three different types of data analysis: regression, classification, and clustering. Regression can be used to determine the relationship between several variables, clustering can be used to detect patterns or groupings within team scores, and classification can be used to identify the category of an athlete's performances from the score results. By observing how the two algorithms perform against each other, we can identify which of those two algorithms is better suit to perform badminton athlete profiling based on historical match data. Video data will go through the trained YOLOv5 model for object detection, and the results will be used on Tracker to get a clear picture of the athlete's general performance and plastyle, as well as how it demonstrates our findings from the historical match data analysis.

1.6 Thesis Outline

The writing of this research is divided into at least five chapters, where each chapter discusses their own topic. The writing structure of this research is as follows:

Chapter I Introduction. This chapter discusses a brief overview of this research, including the research background and why this research needs to be done so that research purposes can be formulated.

Chapter II Theory Discussion. This chapter discusses the theories and literature that are related to the problems described in Chapter I. This chapter also serves as a key in determining the methods that will be used in the next chapter.

Chapter III Research Methodology. This chapter discusses the research plan and/or testing plans. This research has two different parts: score data analysis and video analysis. The steps for each part will be explained in detail on this chapter.

Chapter IV Results and Discussion. This chapter explains the results of the research that has been conducted and presents various arguments about what has been done by attaching scientific papers that has been (or will be) published.

Chapter V Conclusion and Suggestions. This chapter discusses the conclusion based on the results obtained from the search as well as constructive suggestions for further research that can improve on this research.

A list of references will be attached at the end of this report, as well as other important attachments and information about the author.