ISSN 2797-7889

Application of Data Mining for Predicting the Success of Sambo Athletes in Medan City Using the Naïve Bayes Method

Achmad Zoemirrotin Siregar[1], Ali Ikhwan[2]
Universitas Islam Negeri Sumatera Utara, Medan, North Sumatera . Indonesia
Email: achmadzoe77@gmail.com, ali_ikhwan@uinsu.ac.id

Sambo is one of the martial arts sports that has been developing in Indonesia, including in the city of Medan. In the process of athlete development, analytical methods are needed to assist coaches and sports organizations in predicting athlete success more objectively and based on data. This study aims to apply data mining techniques to predict the success of Sambo athletes in Medan using the Naïve Bayes method. This method was chosen because it has good classification capabilities and can process data with a relatively simple and efficient computational process. The data used in this study were obtained from training results and athlete performance records, including several variables such as age, training duration, training intensity, physical condition, technical skills, and competition achievement history. The research process begins with data collection, data preprocessing, model development using the Naïve Bayes algorithm, and evaluation of prediction results. The resulting classification model is then used to determine the success category of athletes, such as having high potential for success or lower potential. The results of the study indicate that the application of the Naïve Bayes method is able to provide predictions of Sambo athletes' success with a fairly good level of accuracy. This system can assist coaches and sports administrators in the selection and development process of athletes in a more effective and targeted manner. Therefore, the implementation of data mining using the Naïve Bayes method can be an alternative solution to support decision-making in the development of Sambo athletes in Medan.
Keywords: Data Mining, Naïve Bayes, Prediction, Sambo Athletes, and Classification.

Corresponding Author:
Achmad Zoemirrotin Siregar
Universitas Islam Negeri Sumatera Utara, Medan, North Sumatera . Indonesia
achmadzoe77@gmail.com

1. Introduction

The development of information technology has brought about significant changes in various areas of life, including the world of sports. Technology is a system that allows manual processes to be replaced by computerized ones, allowing data and information processing to be faster, more accurate, and more efficient. According to Han Jiawei (2012), modern data processing technology enables organizations to analyze large amounts of data, generating valuable information for decision-making. In the context of sports, the use of information technology is becoming increasingly important to support the systematic and data-driven development of athletes.

With the rapid development of digital technology, various sports have begun integrating information systems and data analysis techniques to improve the quality of training and athlete performance. The use of technology in sports extends beyond recording training results to include performance analysis, physical condition monitoring, and predicting an athlete's potential success. According to Tom M. Mitchell (1997), machine learning and data analysis techniques can be used to identify patterns in data that can then be used to make more accurate predictions or decisions. Therefore, the use of analytical technology in sports is an effective approach to improving the quality of athlete development.

Sambo is a rapidly growing martial art in Indonesia, including in Medan. Developing Sambo athletes requires a systematic approach to ensure peak performance in various competitions. During this development process, coaches regularly evaluate the athletes' physical condition, technique, and abilities. These evaluations are crucial because they provide an overview of the athletes' readiness for upcoming matches or competitions.

In Medan, Sambo athletes utilize various training facilities designed to support optimal performance. One of the venues used for physical training is the Medan State University Stadium. This stadium provides various sports facilities to support training programs such as track running, endurance training, and strength training. Through these facilities, athletes can improve their physical abilities, including muscle strength, endurance, and speed.

In addition to physical training, developing technical skills is also a crucial part of developing Sambo athletes. One of the locations used for technical training is SMA Negeri 7 Medan, which provides a spacious hall for technical training. Here, athletes can practice basic techniques, practice match simulations, and spar with the aim of improving their fighting skills and strategies for facing opponents. With adequate training facilities, the athlete development process can be carried out in a more optimal and structured manner.

However, one of the common problems in athlete development is the subjective nature of the evaluation process. In many cases, decisions regarding an athlete's eligibility for competition are often based solely on the coach's direct observations without systematic data analysis. This can lead to less objective decisions and potentially hinder athlete development. Therefore, a data-driven approach is needed that can help coaches more accurately evaluate and predict an athlete's potential success.

One approach to addressing these issues is data mining. Data mining is the process of extracting hidden information or patterns from large amounts of data using statistical, mathematical, and machine learning methods. According to Ian H. Witten (2016), data mining can help identify patterns that are not directly visible, thus supporting data-driven decision-making. The Naïve Bayes algorithm is a frequently used method in data mining. This method is a classification technique based on Bayesian probability theory, assuming that each attribute in the data is independent of the other attributes. According to Zhang (2020), the Naïve Bayes method offers advantages in terms of algorithmic simplicity, computational speed, and the ability to produce a high level of accuracy even when using a relatively small dataset. Therefore, this method is widely used in various studies related to data prediction and classification.

In the sports context, the Naïve Bayes method has been used in various studies to predict athlete performance based on training data and physical condition. Research conducted by Li et al. (2021) showed that the Naïve Bayes method can provide an accuracy rate of up to 87% in predicting athlete success based on training history data, competition results, and athlete health conditions. These results demonstrate that data mining-based approaches have significant potential to improve the effectiveness of athlete development systems.

In addition, research conducted by Tedy Setiadi and Jamaludin (2018) on predicting the type of training for PSHT pencak silat students using the Naïve Bayes method showed test results with an accuracy value of 95%, a precision of 95%, and a recall value of 98%. The results of this study indicate that the Naïve Bayes method can be used effectively in the data classification process in the field of sports. Another study conducted by Nurhadi Surojudin and Abdul Halim Anshor (2022) on predicting the acceptance of prospective athletes to the pencak silat sports education and training center using the Naïve Bayes method with the help of RapidMiner software showed an accuracy level of 89.24%. These results indicate

that the Naïve Bayes method has a high level of accuracy and can be used in the decision-making process related to athlete selection and development.

Based on the problem description and the results of previous research, it can be concluded that the application of data mining using the Naïve Bayes method has great potential in assisting the evaluation process and predicting athlete success. Therefore, this study aims to apply data mining techniques using the Naïve Bayes method in predicting the success of Sambo athletes in Medan City based on athlete balance, speed, and agility data. This research is expected to contribute to the development of a more objective, measurable, and technology-based sports coaching system in Medan City.

2. Literature Review

Information Systems and Data Mining

An information system is a combination of technology, people, and procedures designed to collect, process, store, and distribute information to support decision-making within an organization. According to Kenneth C. Laudon and Jane P. Laudon (2018), an information system is a set of interconnected components that function to collect, process, store, and disseminate information to support coordination, control, analysis, and decision-making within an organization. Modern information systems are generally computer-based and utilize internet networks, enabling faster and more accurate data processing. In the context of sports, information systems can be used to store athlete data, record training results, and analyze athlete performance to support a more effective coaching process. With an integrated information system, coaches and sports organization managers can obtain the information they need in real time, allowing for a more objective and data-driven decision-making process.

Data mining is a technique used to extract hidden information or patterns from large data sets. According to Jiawei Han, Micheline Kamber, and Jian Pei (2012), data mining is the process of discovering interesting patterns or knowledge from large amounts of data using statistical, mathematical, and machine learning methods.

Data mining has become crucial in the big data era because it enables organizations to process large amounts of data into valuable information. In sports, data mining can be used to analyze athlete performance, predict match outcomes, and determine more effective training strategies.
In general, the data mining process consists of several stages known as Knowledge Discovery in Databases (KDD), namely:

1. Data Selection.This is the stage of selecting relevant data from various available data sources.
2. Data Preprocessing.This is the data cleaning stage to remove incomplete, duplicate, or inconsistent data.
3. Data Transformation.This is the data transformation stage to suit the format required in the analysis process.
4. Data Mining.This is the stage of applying a particular algorithm to find patterns or relationships in data.
5. Evaluation and Interpretation.This is the stage of evaluating the results of the analysis to ensure that the patterns found have value and can be used in decision making.

Through these stages, data mining is able to produce information that can be used for various analysis and prediction needs.

Classification

Classification is a data mining technique used to group data into specific categories or classes based on their characteristics. According to Tom M. Mitchell (1997), classification is a machine learning process that aims to map data into predetermined classes based on patterns learned from training data. In the classification method, the system will learn patterns from data whose classes are already known (training data) and then use these patterns to predict the class of new data (testing data). Classification techniques are widely used in various fields, such as health, finance, marketing, and sports. In this study, classification techniques were used to determine whether an athlete has the potential for success in a match based on several physical ability indicators such as balance, speed, and agility.

Naïve Bayes and RapidMiner algorithms.

The Naïve Bayes algorithm is a classification method in data mining based on Bayesian probability theory. This method works by calculating the probability of an event based on previously known information. According to Zhang (2020), Naïve Bayes is a simple classification algorithm but has a fairly high level of accuracy in various prediction applications. The basic concept of the Naïve Bayes algorithm comes from Bayes' Theorem, first introduced by Thomas Bayes. Bayes' Theorem is used to calculate the probability of a hypothesis based on available evidence.
Mathematically, the Naïve Bayes formula can be written as follows:

$$P(H|X)=\frac{P(X|H)\times P(H)}{P(X)}$$

Information:
P(H|X)= probability of hypothesis H based on condition X
P(X|H)= probability of data X if hypothesis H is true
P(H)= initial probability of the hypothesis
P(X)= probability of observed data
The advantages of the Naïve Bayes method include:

1. Simple and fast calculation process
2. Able to work with large datasets
3. Has a fairly good level of accuracy
4. Doesn't require a lot of training data

Due to these advantages, the Naïve Bayes algorithm is often used in various applications such as document classification, email spam detection, sentiment analysis, and predictions in the sports field. Meanwhile, RapidMiner is software used for data analysis, machine learning, and data mining. This software provides a variety of data analysis algorithms that can be used without the need for complex programming.

According to Hofmann and Klinkenberg (2016), RapidMiner provides various features for data processing, such as preprocessing, classification, clustering, and model evaluation. With its visual workflow-based interface, users can easily build data analysis models systematically.
In this study, RapidMiner was used to assist the data analysis process and implementation of the Naïve Bayes algorithm in predicting the success of Sambo athletes based on the physical condition data of the athletes.

Sambo Sport

Sambo is a martial art originating from Russia and combining various wrestling techniques and traditional self-defense techniques. It emphasizes locks, throws, and controlling opponents in the arena. The international organization that governs this sport is the International Sambo Federation.

 ISSN 2797-7889

Sambo is growing rapidly in various countries, including Indonesia. This sport not only trains technical fighting skills but also improves physical strength, agility, balance, and speed. Therefore, the development of Sambo athletes requires a structured training system and regular performance evaluations.
In training Sambo athletes, several very important physical condition components include:

1. Balance
   The body's ability to maintain a stable position while moving or when receiving pressure from an opponent.
2. Speed
   The ability of an athlete to perform movements in a short time.
3. Agility
   The ability to change direction of movement quickly without losing balance.

These three components are important factors that can influence an athlete's performance in a match.

Previous Research

This research is based on previous research that has been carried out, as in the following table:

Table 1. Collection of Previous Research.

No	Name	Title	Year	Contents
1	Muhammad Ikhsan, Rakhmat Kurniawan	Application of Text Mining in the Student Thesis Advisor Recommendation System Using the Naïve Bayes Classifier Algorithm	2023	This research aims to make it easier for study program managers to determine the appropriate Thesis Supervisor between the topic and the lecturer's scientific knowledge.
2	Jefri, Zaehol Fatah	Data Mining Classification to Predict Student Graduation Using the Naïve Bayes Method	2025	This study aims to determine whether students will be able to graduate on time or not using the classification method to predict student graduation using the Naïve Bayes algorithm.
3	Desti Astari Umbu Zaza, Elfira Umar, Felysitas Ema Ose Sanga	Application of Text Mining in Classifying Student Thesis Titles at Stella Maris University, Sumba Using the Naïve Bayes Method	2024	This research aims to guide students to choose thesis titles that are relevant to their field of study using the text mining method.
4	Rani Yunita, Mia Kamayani	Comparison of SVM and Naive Bayes Algorithms in Sentiment Analysis of Thesis Obligation Elimination Policy	2023	This study aims to compare the evaluation results of the Support Vector Machine (SVM) algorithm with Naïve Bayes using 80% training data and 20% test data.
5	Salmin Dania, Rezqiwati Ishak M.Kom, Hastuti Dalai M.Kom	Application of the Naive Bayes Classifier Algorithm for Classifying Thesis Titles Based on Concentration	2024	This study aims to develop a thesis title classification system based on concentration using the Naive Bayes Classifier algorithm.
6	Satrio Hadi Wijoyo, Satrio Agung Wicaksono,	Attribute Selection Analysis in Predicting Student Thesis Failure Using the Naïve Bayes Algorithm	2023	This study aims to analyze attribute selection for predicting student failure in completing their

 ISSN 2797-7889

No	Name	Title	Year	Contents
	Admaja Dwi Herlambang	in the Information Systems Department, Brawijaya University		thesis using the naïve Bayes algorithm.
7	Sukriadi, Ismail, A. M Andzar	Application of Text Mining in Classifying Student Proposed Thesis Titles Using the Naive Bayes Method	2023	This research aims to make it easier for study programs to monitor the suitability of research roadmaps, based on the concentration of study programs with research conducted by students.

So the difference in this research proposal is the Application of Data Mining with the Naive Bayes Classifier Method to Support the Promotion Strategy of the Faculty of Science and Technology, UINSU.

3. Research Methods

This study uses a quantitative approach with data mining methods to analyze and predict the success of Sambo athletes in Medan City. The quantitative approach was used because this study utilizes numerical data obtained from the results of measurements of athletes' physical conditions such as balance, speed, and agility. The data is then analyzed using a classification algorithm to produce a predictive model. The data mining method was chosen because it is able to find patterns or relationships from a large amount of available data so that it can be used to produce more objective predictions. In this study, the algorithm used is Naïve Bayes, a classification method based on Bayesian probability theory and is often used in predictive analysis because it has a fairly good level of accuracy and a relatively simple computational process. This study was conducted in Medan City with the research subjects being Sambo athletes who are members of the Medan City Sambo training community. Data collection activities were carried out at several athlete training locations, including the Medan State University Stadium, which is used as a physical training venue for athletes, and Medan State High School 7, which is used as a technical training and match simulation venue. This study was conducted from April 2025 to July 2025 with several stages including data collection, data processing, data analysis using the Naïve Bayes method, and evaluation of the prediction results.

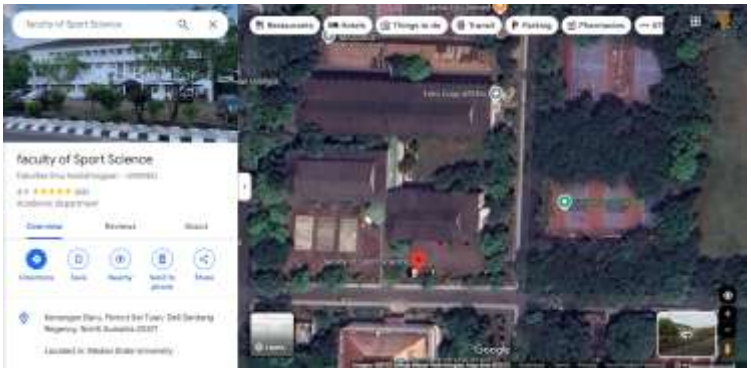

Figure 1. Location Map of the Faculty of Sports, State University of Medan

The population in this study was all Sambo athletes actively participating in training programs in Medan City. This population consisted of athletes who participated in regular training and had recorded physical condition test results. The research sample was a subset of the population used as a data source. The

sample used in this study consisted of Sambo athletes who had recorded physical condition test results, including:

1. Balance
2. Speed
3. Agility

The sample data was obtained from the results of physical tests conducted periodically by Medan City Sambo coaches on athletes. The data sources for this study consisted of two types:

1. Primary Data
   Primary data is data obtained directly from the research subjects. This data was obtained through interviews and data collection on the athletes' physical condition tests conducted by Medan City Sambo coaches, Mr. Filli Azandi and Mr. Anggi Muda Siregar.
2. Secondary Data
   Secondary data is data obtained from other sources such as books, scientific journals, and previous research related to data mining, the Naïve Bayes algorithm, and athlete performance analysis.

The data collection techniques used in this study include several methods as follows:

1. Observation
   Observations were carried out by directly observing the Sambo athlete training process and physical condition testing activities carried out by the trainer.
2. Interview
   Interviews were conducted with Sambo athlete coaches in Medan City to obtain information regarding the athlete development system, training methods used, and indicators used to assess athlete readiness to participate in competitions.
3. Literature Study
   Literature studies were conducted by studying various library sources such as books, scientific journals, and previous research related to data mining and the Naïve Bayes method.
4. Documentation
   Documentation is carried out by collecting data from the results of physical condition tests on athletes, including balance, speed and agility.

The research stages carried out in this study include the following steps:

1. Identification of problems
   Determining the problem to be researched, namely how to predict the success of Sambo athletes based on physical condition data.
2. Data collection
   Collecting data from the results of physical condition tests on athletes, including balance, speed and agility.
3. Data Preprocessing
   Perform data cleansing to remove incomplete or invalid data.
4. Application of the Naïve Bayes Method
   Using the Naïve Bayes algorithm to build a classification model that can predict athlete success.
5. Model Testing
   Testing the classification model using evaluation methods such as confusion matrix to determine the level of system accuracy.
6. Results Analysis
   Analyze the prediction results generated by the system to determine the success rate of the method used.

 ISSN 2797-7889

Figure 2. Research flow

In this study, the data analysis process was conducted using RapidMiner software. RapidMiner is a frequently used application in data mining processes because it provides various data analysis algorithms that can be easily used through a visual workflow-based interface.
RapidMiner is used to perform several processes, including:

1. Data processing (data preprocessing)
2. Application of the Naïve Bayes algorithm
3. Classification model testing
4. Evaluation of the system accuracy level

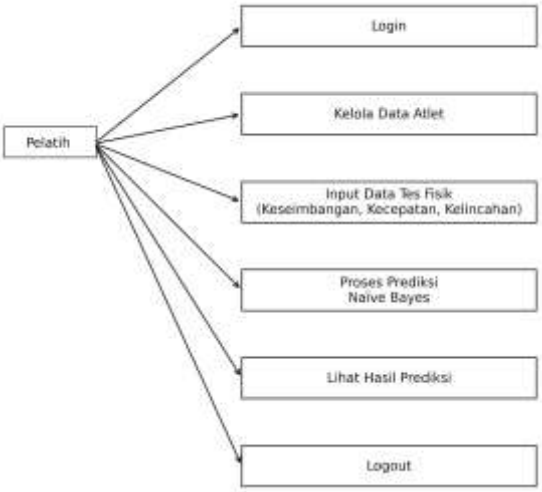

Figure 3. Prediction Usecase Diagram

Use Case Diagrams are used to illustrate the interaction between users and the system. In this system, there is one main actor, the Coach, who has several access rights to manage athlete data and view prediction results.
Actor

Coach
Use Case

1. Login to the system
2. Managing athlete data
3. Inputting athlete physical test data
4. Performing the prediction process
5. View the prediction results
6. Logout

The data used in the athlete prediction system consists of several attributes as follows:

Table 2. Data Attributes

No	Attribute	Information
1	Athlete Name	Name of sambo athlete
2	Balance	Balance test results
3	Speed	Speed test result value
4	Agility	Agility test results
5	Results	Athlete success categories

The data will be processed using the Naïve Bayes method to produce predictions of athlete success.

4. Results and Analysis

General Description of Research Object

This research was conducted on Sambo athletes in Medan. Sambo is a martial art that combines wrestling and locking techniques to defeat an opponent. This sport is developing in various regions in Indonesia, including Medan, which has several active athletes who regularly participate in training and competitions. In the process of coaching Sambo athletes, coaches implement various training programs aimed at improving the athletes' physical and technical abilities. Some physical aspects frequently tested in the athlete training process include balance, speed, and agility. These three aspects are important components in determining athlete performance during competitions. The athletes' physical training was conducted at the Medan State University Stadium, while technical training and match simulations were conducted at Medan State High School 7. The data from the athletes' physical tests were then used as a dataset in this study to predict athlete success using the Naïve Bayes method.

Research Dataset

The dataset used in this study comes from the results of physical tests of Sambo athletes which include three main attributes, namely:

1. Balance
2. Speed
3. Agility

Each attribute is then classified into several value categories, such as Good, Average, and Poor. This data is then used as training data to build a classification model using the Naïve Bayes algorithm.

Table 3. Dataset used in the research

No	Athlete Name	Balance	Speed	Agility	Durability	Strength	Results
1	Athlete 1	Tall	Tall	Tall	Tall	Tall	Succeed
2	Athlete 2	Tall	Currently	Tall	Currently	Tall	Succeed
3	Athlete 3	Currently	Tall	Tall	Currently	Tall	Succeed

 ISSN 2797-7889

No	Athlete Name	Balance	Speed	Agility	Durability	Strength	Results
4	Athlete 4	Currently	Currently	Tall	Currently	Currently	Succeed
5	Athlete 5	Tall	Currently	Currently	Tall	Currently	Succeed
6	Athlete 6	Currently	Tall	Currently	Currently	Tall	Succeed
7	Athlete 7	Tall	Tall	Currently	Currently	Tall	Succeed
8	Athlete 8	Currently	Currently	Currently	Currently	Currently	Not successful
9	Athlete 9	Low	Currently	Currently	Low	Currently	Not successful
10	Athlete 10	Currently	Low	Currently	Currently	Low	Not successful
11	Athlete 11	Low	Low	Currently	Low	Low	Not successful
12	Athlete 12	Currently	Currently	Low	Currently	Currently	Not successful
13	Athlete 13	Tall	Tall	Currently	Tall	Tall	Succeed
14	Athlete 14	Currently	Tall	Tall	Currently	Currently	Succeed
15	Athlete 15	Tall	Currently	Tall	Tall	Currently	Succeed
16	Athlete 16	Currently	Currently	Currently	Currently	Currently	Not successful
17	Athlete 17	Low	Currently	Low	Low	Currently	Not successful
18	Athlete 18	Currently	Low	Currently	Low	Currently	Not successful
19	Athlete 19	Tall	Currently	Currently	Tall	Tall	Succeed
20	Athlete 20	Currently	Tall	Currently	Currently	Currently	Succeed
21	Athlete 21	Currently	Currently	Tall	Currently	Currently	Succeed
22	Athlete 22	Tall	Tall	Tall	Tall	Currently	Succeed
23	Athlete 23	Low	Currently	Currently	Currently	Low	Not successful
24	Athlete 24	Currently	Low	Low	Currently	Low	Not successful
25	Athlete 25	Tall	Currently	Tall	Currently	Tall	Succeed
26	Athlete 26	Currently	Currently	Currently	Low	Currently	Not successful
27	Athlete 27	Low	Low	Currently	Low	Low	Not successful
28	Athlete 28	Currently	Tall	Tall	Tall	Currently	Succeed
29	Athlete 29	Tall	Tall	Currently	Currently	Tall	Succeed
30	Athlete 30	Currently	Currently	Currently	Currently	Currently	Not successful

This dataset is then used to build a classification model using the Naïve Bayes method. Total Data: 30 Athletes

Class distribution:

1. Success: 17 Athletes
2. Failed: 13 Athletes

Predictor variables:

1. Balance
2. Speed
3. Agility
4. Durability
5. Strength

Classification targets:

1. Succeed
2. No

 ISSN 2797-7889

Naïve Bayes Calculation Process

The Naive Bayes method is used to calculate the probability of a class based on the attributes of the data. This method uses the concept of conditional probability based on Bayes' Theorem proposed by Thomas Bayes.

The basic formula for the Naïve Bayes method is:

P(H|X)=————

Information:
P(H|X)= probability of hypothesis based on data conditions
P(X|H)= probability of data based on hypothesis
P(H)= initial probability of the hypothesis
P(X)= data probability
In this study, the hypothesis used is:

1. Potentially Successful
2. Not Potential

Calculation Example

For example, there is athlete data with the following attributes:

1. Balance = Good
2. Speed = Medium
3. Agility = Good

Based on the available training data, the probability of each class is calculated.
Calculation example:
P(Potential) = 6 / 10 = 0.6.
P(No) = 4 / 10 = 0.4

The probability of each attribute relative to that class is then calculated. The results are then compared to determine the class with the highest probability. If the probability value for the Potential class is greater than the No class, then the athlete is predicted to have the potential to succeed in the competition.

Table 4. Naïve Bayes Probability Calculation

Variables	Probability
High Balance | Success	0.47
Medium Balance | Successful	0.47
Low Balance | Success	0.06
High Balance | No	0.08
Medium Balance | No	0.46
Low Balance | No	0.46
High Speed | Success	0.53
Medium Speed | Success	0.41
Low Speed | Success	0.06
High Speed | No	0.08
Medium Speed | No	0.46
Low Speed | No	0.46
High Agility | Success	0.59
Moderate Agility | Successful	0.35

 ISSN 2797-7889

Variables	Probability
Low Agility | Successful	0.06
High Agility | No	0.08
Moderate Agility | No	0.54
Low Agility | No	0.38

Based on the test results using the Naïve Bayes method, the classification model was able to predict athlete success with an accuracy rate of 83%. This value indicates that the Naïve Bayes method is quite effective in classifying athlete success based on balance, speed, and agility variables. Evaluation results using a confusion matrix showed that most of the data was correctly classified. The prediction error was relatively small, so the model can still be used as a tool in the athlete selection and development process.

Prior Probability Calculation

Prior probability is the initial probability of each class before considering other attributes. In this study, there are two classes: Successful and Unsuccessful.
Number of datasets = 30 athlete data
Class distribution:

Category	Amount
Succeed	17
Not successful	13
Total	30

Prior probability calculation:

P(Success)=1730=0.57P(Success)=3017=0.57P(Unsuccess)=1330=0.43P(Unsuccess)=3013=0.43
So the initial probability of the class is:

Class	Probability
Succeed	0.57
Not successful	0.43

Calculation of Likelihood of Each Variable

Likelihood is the probability of a particular attribute appearing in each class.

Equilibrium Variables

Balance	Succeed	Not successful
Tall	8	1
Currently	8	6
Low	1	6

Speed Variable

Speed	Succeed	Not successful
Tall	9	1
Currently	7	6
Low	1	6

Agility Variable

Agility	Succeed	Not successful

 ISSN 2797-7889

Agility	Succeed	Not successful
Tall	10	1
Currently	6	7
Low	1	5

Endurance Variable

Durability	Succeed	Not successful
Tall	8	1
Currently	7	7
Low	2	5

Strength Variable

Strength	Succeed	Not successful
Tall	9	1
Currently	7	8
Low	1	4

New Athlete Predictions

Suppose there is new athlete data with the following criteria:

Variables	Mark
Balance	Tall
Speed	Tall
Agility	Currently
Durability	Currently
Strength	Tall

Based on the results of testing using the Naïve Bayes method, an accuracy rate of 83% was obtained in predicting the success of sambo athletes in Medan City. This indicates that the Naïve Bayes method is quite effective in classifying based on the variables of balance, speed, agility, endurance, and strength. This model can be used as a tool for coaches in identifying athletes with higher potential for success in sambo.

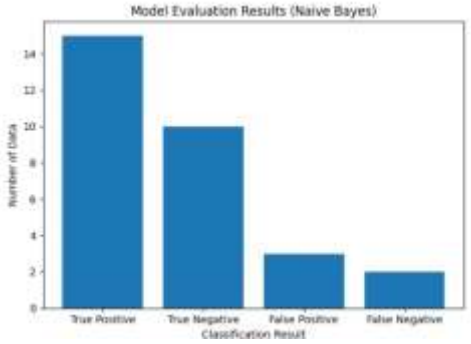

Figure 4. Evaluation Results

The interpretation for the above image is:

1. True Positive (TP) = 15→ Successful athletes are predicted to be successful
2. True Negative (TN) = 10→ The athlete did not succeed, it was predicted that he would not succeed.
3. False Positive (FP) = 3→ Athlete not predicted to succeed

4. False Negative (FN) = 2→ Successful athletes are predicted to fail

So that the model accuracy is obtained at 83%.

5. Conclusion

Based on the results of the study on the application of the Naïve Bayes method to predict the success of sambo athletes in Medan City, it can be concluded that this study has successfully designed and built a web-based information system capable of processing and analyzing athletes' physical data more systematically. The developed system can store athlete data and display the results of success predictions based on several important parameters, namely balance, speed, agility, endurance, and strength. The implementation of the Naïve Bayes method in this system has been proven to be used well in the process of classifying and predicting athlete success through calculating the prior probability and likelihood of each athlete's attribute. Based on the results of testing using a dataset of 30 athlete data, an accuracy value of 83% was obtained, which indicates that this method is quite effective in predicting the potential success of athletes based on the analyzed physical variables. In addition, the developed prediction system can assist coaches in conducting evaluations and making decisions regarding the selection of potential athletes to participate in competitions more objectively and measurably, so that the assessment process does not only depend on the coach's subjectivity alone. Thus, the results of this study indicate that the use of data mining techniques in the field of sports can make a positive contribution in increasing the effectiveness of the athlete development system, especially in the sport of sambo in Medan City.

6. References

Jiawei Han, Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

Tom M. Mitchell, Mitchell, T.M. (1997). Machine learning. McGraw-Hill.

Eibe Frank, Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

Trevor Hastie, Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Christopher M. Bishop, Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Zhang, H. (2020). The optimality of Naïve Bayes. AAAI Conference on Artificial Intelligence, 1(2), 3–9.

Li, X., Wang, Y., & Chen, Z. (2021). Predicting athlete performance using Naïve Bayes classification method. Journal of Sports Analytics, 7(2), 101–110.

Setiadi, T., & Jamaludin. (2018). Application of the Naïve Bayes method to predict the type of training of PSHT pencak silat students. Journal of Information and Computer Technology, 5(2), 45–52.

Surojudin, N., & Anshor, AH (2022). Predicting the acceptance of prospective pencak silat athletes using the RapidMiner-based Naïve Bayes method. Journal of Information Systems and Technology, 4(1), 15–22.

Ian H. Witten, Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques (2nd ed.). Morgan Kaufmann.