Bigint Computing Journal
Volume 4, Issue 2, 2026, pp. 91-98
DOI: https://doi.org/10.55537/bigint.v4i2.1868
ISSN: 3032-5374
https://journal.aira.or.id/bigint/index

Naïve Bayes-Based Classification of Wrestling Athletes Using Physical Performance Data

Achmad Zoemirrotin Siregar*, Ali Ikhwan

Department of Information Systems, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

ABSTRACT
Objective evaluation of athletes' physical test data can help reduce subjectivity in athlete assessment. This study develops a web-based classification system using the Naïve Bayes algorithm to classify wrestling athletes as Successful or Unsuccessful based on speed, balance, and agility attributes. The study uses a quantitative approach and the Waterfall software development model. The dataset consists of 36 wrestling athlete records, with 30 records used as training data to calculate prior and conditional probabilities and 6 independent records reserved for final testing; the testing records were not used during the training or probability calculation process. Speed is categorized as Slow, Normal, or Fast; balance as Poor, Fair, or Good; and agility as Low, Medium, or High. The target labels, Successful and Unsuccessful, were assigned independently based on recorded athlete performance assessment outcomes and were not derived from the predictor variables. The Naïve Bayes classification process calculates class priors, attribute likelihoods, posterior probabilities, and predicted classes. For the illustrated test case with Slow speed, Poor balance, and Low agility, the normalized probabilities were 4% for Successful and 96% for Unsuccessful. The six-record performance evaluation produced 83.33% accuracy, 75% precision, 100% recall, and an 85.71% F1-score. The system can support structured and more objective athlete assessment; however, the limited dataset size may restrict the generalizability of the results and requires validation using a larger dataset.
KEYWORDS: Naïve Bayes; athlete classification; data mining; sports information system

*Corresponding author: Achmad Zoemirrotin Siregar, achmadzoe77@gmail.com

Article History: Received: 2026-08-01; Accepted: 2026-08-15

1. INTRODUCTION

Digital information systems can support data-driven decision making by organizing records, accelerating analysis, and transforming stored data into information that can be used consistently. In sports, such systems may assist coaches in documenting physical-test results and evaluating athlete development using repeatable criteria rather than relying only on informal observation [1], [2].

Wrestling performance depends on physical, technical, tactical, and mental readiness. Periodic physical testing can help identify strengths and weaknesses, but the value of those tests depends on how clearly the measurements and decision criteria are defined. A probabilistic classifier such as Naïve Bayes can be used to estimate class membership from categorical attributes, making it suitable for simple decision-support applications when the underlying data and labels are adequately defined [3]–[6].

Recent studies have applied data-mining and classification methods to web-based information systems, athlete or player selection, educational prediction, and other structured decision problems. Waterfall-based information-system development has also been used in sequential web-system projects [7], while data mining has been applied to healthcare inventory [8], benefit-recipient classification [9], and student graduation prediction [10]. These studies show the practical breadth of data-driven classification, but algorithm performance remains dependent on the dataset, feature definitions, class labels, and validation design.

A closely related study by the same authors applied the Naïve Bayes method to predict the success of Sambo athletes in Medan [11]. However, the present study is a separate investigation that focuses specifically on wrestling athletes. Although both studies employ the Naïve Bayes algorithm, they differ in their research objects, datasets, data collection processes, and application contexts. The present study uses independently collected data from wrestling athletes and evaluates physical attributes consisting of speed, balance, and agility to classify athlete status as Successful or Unsuccessful. The system developed in this study was designed specifically to process and classify wrestling athlete data; therefore, the dataset, classification results, probability calculations, system inputs and outputs, and evaluation results are specific to the present research. Any similarities in general system components or methodological procedures arise from the use of the same classification algorithm and common data-mining workflow. The earlier Sambo study is explicitly cited as related work [11], while the present study provides a distinct application and dataset focused on wrestling athletes.

Accordingly, this study aims to develop and evaluate a web-based Naïve Bayes classification system for wrestling athletes using speed, balance, and agility as input attributes. The intended contribution is a transparent workflow that integrates physical-test data management, probability calculation, classification, and performance reporting while explicitly acknowledging the limits of validation on a small dataset.

2. METHODOLOGY

2.1. Research Design / Proposed Method

This study uses a quantitative research approach combined with the Waterfall software development model and the Naïve Bayes data-mining algorithm. The Waterfall model follows a sequential development flow consisting of analysis, design, coding, testing, and support [7]. The developed product is a web-based information system for classifying athlete outcomes from physical-test results.

The study was conducted from October to December 2025 at the Gedung Pertemuan Advent Medan, Indonesia. Data collection was conducted independently for the purpose of the present study through observation, interviews, documentation, and literature review. Interviews with coaches were conducted to identify physical performance indicators relevant to wrestler achievement, while the athlete data used in this study were collected and compiled specifically for the development and evaluation of the proposed classification model. Therefore, the dataset used in this study does not rely solely on data reported in previous studies conducted at the same or related training locations.

The independent variables used in this study were speed, balance, and agility, while the dependent variable was athlete status, categorized as Successful or Unsuccessful.

The research procedure consists of problem identification, data collection, data preprocessing, Naïve Bayes model formation, testing, performance evaluation, and interpretation. The manuscript should provide enough operational detail for another researcher to reconstruct the dataset preparation, class-label assignment, and validation procedure.

2.2. Data / Algorithm / Procedure

The dataset used in this study consists of 36 wrestler records from Medan City, comprising 30 records used for model development and 6 additional records used for final testing. The The Naïve Bayes Classifier requires only a small amount of training data to perform the parameter estimation needed for classification, which is one of its key advantages over other classifier models [19]. The six testing records are independent from and are not included in the 30 records used for model development. Each record contains three categorical physical attributes, namely speed (Fast, Normal, and Slow), balance (Good, Fair, and Poor), and agility (High, Medium, and Low), with the target class categorized as Successful or Not Successful.

Among the 30 records used for model development, 21 records belong to the Successful class and 9 records belong to the Not Successful class, resulting in a class distribution of 70% and 30%, respectively. The six additional testing records contain an equal class distribution, consisting of 3 Successful and 3 Not Successful records. The data were selected based on the availability and completeness of the required attributes for the classification process.

Considering the relatively small sample size, model validation was performed using stratified cross-validation on the 30 development records. Stratification was applied to preserve the proportion of the Successful and Not Successful classes in each fold, thereby reducing the risk of biased evaluation caused by class imbalance. After the cross-validation process, the final Naïve Bayes model was evaluated using the six independent testing records to demonstrate its prediction performance on data that were not used during model development or validation.

Table 1. Research variables and category definitions

Variable	Categories	Role
Speed	Fast, Normal, Slow	Independent attribute
Balance	Good, Fair, Poor	Independent attribute
Agility	High, Medium, Low	Independent attribute
Status	Successful, Unsuccessful	Target class

Preprocessing is performed to remove incomplete records, correct inconsistencies, and ensure that the categorical values are ready for classification. The cleaned training data are used to calculate class priors and attribute likelihoods. The complete record-level training data are provided in Appendix A in anonymized form.

Naïve Bayes determines the posterior probability of a class given an observed feature vector according to Bayes' theorem:

$$P(H|X) = P(X|H)\,P(H) / P(X) \tag{1}$$

where $P(H|X)$ is the posterior probability of hypothesis H given data X, $P(X|H)$ is the likelihood of X given H, $P(H)$ is the prior probability of H, and $P(X)$ is the probability of the observed data. For categorical attributes, a class score can be computed as:

$$\text{Score}(C) = P(C) \times P(x_1|C) \times P(x_2|C) \times \ldots \times P(x_n|C) \tag{2}$$

The class with the larger score is selected as the prediction. Model performance is evaluated using a confusion matrix together with accuracy, precision, recall, and F1-score. Because only six testing records are shown in the system output, the evaluation should be interpreted as preliminary; a larger hold-out set or stratified cross-validation is recommended for a more stable estimate of generalization.

3. RESULTS AND DISCUSSION

The resulting application integrates athlete data, training data, testing data, probability calculation, classification, and reporting functions. The following subsections present the system implementation, probability calculations, an illustrated classification case, and the performance summary.

3.1. System Implementation

Figure 1 presents the dashboard used as the main navigation and summary interface for the developed application. The dashboard should report the final, verified numbers of training and testing records consistently with the Methodology section.

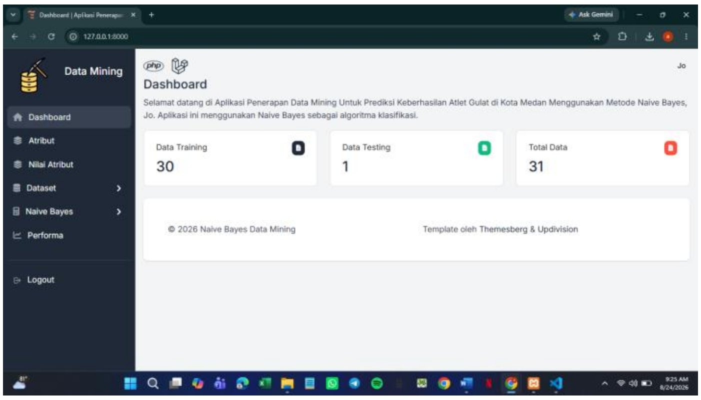

Figure 1. Dashboard interface of the web-based athlete classification system

Figure 2 presents the training-data module used to store the historical records from which class priors and attribute likelihoods are calculated.

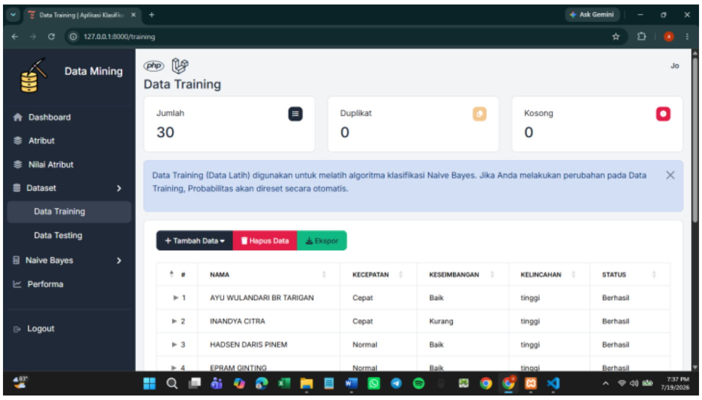

Figure 2. Training-data management interface

Figure 3 presents the probability module used to display class priors and attribute likelihoods for the Naïve Bayes calculation.

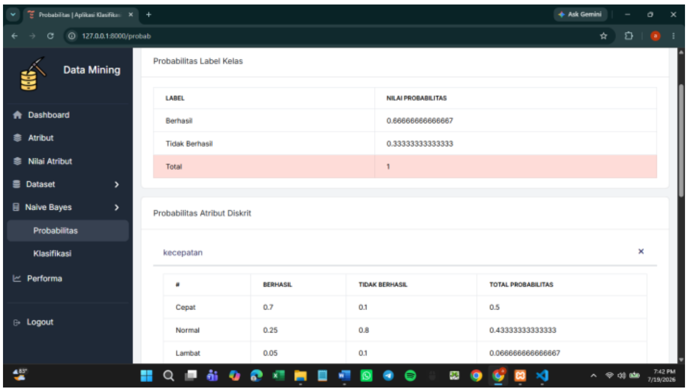

Figure 3. Probability calculation interface

Figure 4 presents the classification module in which class scores are compared and the predicted class is displayed.

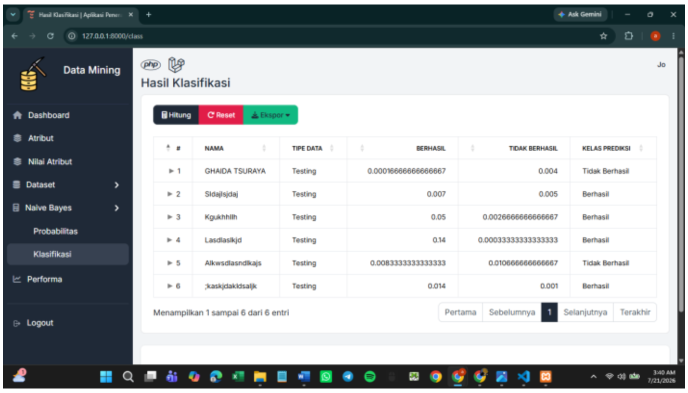

Figure 4. Naïve Bayes classification-result interface

3.2. Prior Probability

From the 30 training records, 20 records are labeled Successful and 10 are labeled Unsuccessful. The corresponding prior probabilities are shown in Table 2.

Table 2. Prior probabilities of the target classes

Class	Calculation	Prior Probability
Successful	20 / 30	0.6667 (66.67%)
Unsuccessful	10 / 30	0.3333 (33.33%)

The prior values indicate that Successful is the more frequent class in the training set. This class imbalance should be considered when interpreting performance metrics.

3.3. Likelihood Probability

Likelihood probabilities are calculated by counting the occurrence of each attribute category within each class.

Table 3. Likelihood probabilities for speed

Speed	Successful	Unsuccessful	Total
Fast	14/20 = 0.7000	1/10 = 0.1000	15/30 = 0.5000
Normal	5/20 = 0.2500	8/10 = 0.8000	13/30 = 0.4333
Slow	1/20 = 0.0500	1/10 = 0.1000	2/30 = 0.0667

Fast speed occurs more frequently among Successful records, whereas Normal speed occurs more frequently among Unsuccessful records in the training set.

Table 4. Likelihood probabilities for balance

Balance	Successful	Unsuccessful
Good	12/20 = 0.6000	1/10 = 0.1000
Fair	6/20 = 0.3000	5/10 = 0.5000
Poor	2/20 = 0.1000	4/10 = 0.4000

Based on the record-level counts in Appendix A, Good balance appears in 1 of 10 Unsuccessful records and Fair balance in 5 of 10 Unsuccessful records. These values correct the inconsistent counts in the previous version of the table.

Table 5. Likelihood probabilities for agility

Agility	Successful	Unsuccessful
High	10/20 = 0.5000	1/10 = 0.1000
Medium	9/20 = 0.4500	6/10 = 0.6000
Low	1/20 = 0.0500	3/10 = 0.3000

High agility is most frequent among Successful records, whereas Medium and Low agility occur more frequently among Unsuccessful records.

3.4. Classification Example

An illustrated test case uses Slow speed, Poor balance, and Low agility. The corresponding conditional probabilities are summarized in Table 6.

Table 6. Likelihood values for the illustrated test case

Attribute	Observed value	P(X|Successful)	P(X|Unsuccessful)
Speed	Slow	1/20 = 0.05	1/10 = 0.10
Balance	Poor	2/20 = 0.10	4/10 = 0.40
Agility	Low	1/20 = 0.05	3/10 = 0.30

$$P(X|\text{Successful}) = 0.05 \times 0.10 \times 0.05 = 0.00025 \quad (3)$$

$$P(X|\text{Unsuccessful}) = 0.10 \times 0.40 \times 0.30 = 0.012 \quad (4)$$

After multiplying by the class priors:

$$P(\text{Successful}|X) = 0.6667 \times 0.00025 = 0.0001667 \quad (5)$$

$$P(\text{Unsuccessful}|X) = 0.3333 \times 0.012 = 0.004 \quad (6)$$

Normalization produces approximately 4% for Successful and 96% for Unsuccessful; therefore, the illustrated case is classified as Unsuccessful.

3.5. Model Evaluation

The performance page of the implemented system shows six testing records with the confusion matrix summarized in Table 7. Successful is treated as the positive class.

Table 7. Confusion matrix for the six-record test set

Predicted / Actual	Successful	Unsuccessful	Total
Successful	3	1	4
Unsuccessful	0	2	2
Total	3	3	6

Table 8. Performance metrics derived from the system output

Metric	Calculation	Value
Accuracy	(3 + 2) / 6	83.33%
Precision	3 / (3 + 1)	75.00%
Recall	3 / (3 + 0)	100.00%
F1-score	2 × 0.75 × 1.00 / (0.75 + 1.00)	85.71%

The reported metrics indicate five correct classifications among six test records. Because the test set is very small, a single additional error would materially change the percentages. Accordingly, these values should be presented as preliminary performance on the available test set rather than as a general estimate of model accuracy.

Table 9. Per-record classification results on the test set

No.	Predicted	Actual	Result
1	Not Successful	Not Successful	TN
2	Successful	Not Successful	FP
3	Successful	Successful	TP
4	Successful	Successful	TP
5	Not Successful	Not Successful	TN
6	Successful	Successful	TP

The performance metrics in Table 8 were derived from the six-record test set as follows. Each record's predicted class was compared against its actual (ground-truth) class label to determine the confusion matrix components, yielding TP = 3, FP = 1, FN = 0, and TN = 2, consistent with the confusion matrix reported in Table 8, from which the accuracy (83.33%), precision (75.00%), recall (100.00%), and F1-score (85.71%) were calculated. Given the very small size of the test set (n = 6), these values should be interpreted as preliminary performance indicators rather than generalizable estimates of model accuracy.

3.6. Discussion

The findings show that the Naïve Bayes procedure can be implemented transparently with three categorical physical attributes and integrated into a web-based information system. The calculation workflow is simple enough to audit from priors and likelihoods, which is useful for decision support when the underlying categories and labels are explicitly defined.

Related work has used Naïve Bayes for player or athlete selection [3], [5], sports-related sentiment analysis [13], sales classification [14], educational-funding classification [15], application-based recommendation or classification [16], talent classification [17], and product-sales prediction [18]. These studies demonstrate broad applicability, but they do not establish that one algorithm is universally superior; performance must be evaluated against the characteristics of the target dataset [6].

The earlier Sambo study by Siregar and Ikhwan [11] is particularly important because it uses the same general athlete-success framing and similar physical attributes in Medan. It is important to clarify that the present wrestling study is independent from the earlier Sambo study by Siregar and Ikhwan [11], beyond the difference in sport. The dataset and participants are entirely distinct, comprising a different cohort of wrestling athletes rather than the Sambo athletes examined in [11]. Data collection for the present study was conducted from April to July 2026, a period entirely separate from and non-overlapping with the Sambo study's data collection period of October to December 2025.

The classification labels also differ between the two studies. The Sambo study [11] employed the classes "potentially successful" and "not potentially successful," whereas the present wrestling study employs the classes "successful" and "not successful." Furthermore, the attribute value schemes differ substantially: the Sambo study used a single 3-level scale (high, medium, low) for its attributes, while the present wrestling study defines nine distinct attribute values across three specific physical parameters — speed (fast, normal, slow), balance (good, fair, poor), and agility (high, medium, low).

In addition, the implemented systems differ in platform: the Sambo study [11] was carried out using RapidMiner, whereas the present study employs a purpose-built web-based system. These distinctions in dataset, timing, labeling scheme, attribute structure, and implementation confirm that the present study constitutes a novel and independent contribution rather than a redundant or duplicate publication of the earlier Sambo research.

The principal limitation is the small validation sample. Although the six-record test set yields 83.33% accuracy, 75% precision, 100% recall, and 85.71% F1-score, the estimates are unstable because each case has a large influence on the percentages. Misclassification in the Naïve Bayes Classifier can occur when training data is limited or when testing data is not represented in the training set, resulting in a zero probability outcome; Laplacian Smoothing addresses this issue by adding a small positive value to the probability calculation [20]. Future work should expand the dataset, report the data-splitting strategy in full, use stratified or repeated cross-validation where appropriate, and compare Naïve Bayes with alternative classifiers using the same evaluation protocol.

The web implementation also needs consistency between the research domain and the user interface. A decision-support system should use final screenshots, labels, and terminology that match the wrestling-athlete application described in the manuscript [12].

4. CONCLUSION

This study presents a web-based Naïve Bayes classification workflow for wrestling-athlete assessment using speed, balance, and agility. The training data contain 20 Successful and 10 Unsuccessful records, and the illustrated Slow–Poor–Low case is classified as Unsuccessful with normalized probabilities of approximately 4% and 96%. For the six testing records shown by the system, the confusion matrix corresponds to 83.33% accuracy, 75% precision, 100% recall, and 85.71% F1-score. These results support the feasibility of the implemented workflow as a preliminary decision-support tool, but they do not justify broad claims of predictive performance because the validation sample is very small. Further research should use a larger, clearly separated dataset, stronger validation, explicit class-label criteria, and direct comparison with alternative classifiers.

ACKNOWLEDGMENT

The authors acknowledge the coaches and the sports training environment that supported the collection and interpretation of athlete physical-test data.

DECLARATIONS

Artificial Intelligence Use

Generative artificial intelligence (AI) was used to assist with manuscript structuring, language editing, paraphrasing, and formatting. Claude AI Model Sonnet 5 was used for these purposes. The authors reviewed and verified all AI-assisted content and remain fully responsible for the accuracy, originality, analysis, interpretation, citations, and conclusions presented in the manuscript.

Author Contributions

Achmad Zoemirrotin Siregar: conceptualization, data preparation, system implementation, analysis, and manuscript preparation. Ali Ikhwan: supervision, methodological guidance, and manuscript review.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Ethics Approval / Ethical Clearance

This study used existing athlete performance data collected as part of routine athlete training and performance assessment. The researchers obtained authorization from the relevant institution and/or coach responsible for the athlete data to use the data for research purposes. As the analysis used anonymized performance records and did not involve direct intervention with participants, the study was not subject to formal ethical review by PGSI Kota Medan. All data were anonymized before analysis, and no identifiable athlete information is reported in the manuscript.

Data Availability Statement

The anonymized version of the 30-record training dataset is presented in Appendix A to support transparency of the calculations. The raw dataset, source code, and additional test records used in this study are not subject to privacy, confidentiality, or institutional restrictions, and may be shared upon reasonable request to PGSI Kota Medan.

REFERENCES

[1] A. Almira, Suendri, and A. Ikhwan, “Implementasi Data Mining Menggunakan Algoritma FP-Growth pada Analisis Pola Pencurian Daya Listrik,” Jurnal Informatika Universitas Pamulang, vol. 6, no. 2, pp. 442–448, 2021, doi: 10.32493/informatika.v6i2.12278.

[2] S. Sukriadi, I. Ismail, and A. M. Andzar, “Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes,” Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI), vol. 6, no. 2, pp. 184–196, 2023, doi: 10.57093/jisti.v6i2.174.

[3] M. S. Anam, Y. B. Utomo, and H. Kurniadi, “Sistem Pendukung Keputusan Seleksi Pemain Futsal dengan Menggunakan Metode Naive Bayes,” Insect (Informatics and Security): Jurnal Teknik Informatika, vol. 11, no. 1, pp. 97–103, 2025, doi: 10.33506/insect.v11i1.4240.

[4] J. Jefri and Z. Fatah, “Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes,” Jurnal Ilmiah Multidisiplin Ilmu, vol. 2, no. 1, pp. 29–37, 2025, doi: 10.69714/mhjq1v85.

[5] N. Surojudin and A. H. Anshor, “Penerapan Algoritma Naive Bayes Untuk Memprediksi Penerimaan Calon Atlet Pusat Pendidikan Dan Latihan Pelajar Cabang Olahraga Pencak Silat (Studi Kasus IPSI Kabupaten Bekasi),” Jurnal Ilmiah SIGMA: Informatics Engineering Journal of UPB, vol. 13, no. 4, pp. 223–230, 2022.

[6] R. Yunita and M. Kamayani, “Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi,” Indonesian Journal of Computer Science, vol. 12, no. 5, pp. 2879–2890, 2023, doi: 10.33022/ijcs.v12i5.3415.

[7] Supiyandi, M. Zen, C. Rizal, and M. Eka, “Perancangan Sistem Informasi Desa Tomuan Holbung Menggunakan Metode Waterfall,” JURIKOM (Jurnal Riset Komputer), vol. 9, no. 2, pp. 274–280, 2022, doi: 10.30865/jurikom.v9i2.3986.

[8] D. P. A. A. Dewi and A. Ikhwan, “Implementasi Data Mining Pada Sistem Persediaan Obat Di Puskesmas Sei Berombang Dengan Metode Algoritma Apriori,” The Indonesian Journal of Computer Science, vol. 11, no. 3, pp. 958–973, 2022, doi: 10.33022/ijcs.v11i3.3112.

[9] A. Pebdika, R. Herdiana, and D. Solihudin, “Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.

[10] D. A. Punkastyo, F. Septian, and A. Syaripudin, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa,” Journal of System and Computer Engineering (JSCE), vol. 5, no. 1, pp. 24–35, 2024, doi: 10.61628/jsce.v5i1.1073.

[11] A. Z. Siregar and A. Ikhwan, “Application of Data Mining for Predicting the Success of Sambo Athletes in Medan City Using the Naïve Bayes Method,” Jurnal Info Sains: Informatika dan Sains, vol. 16, no. 1, pp. 168–181, 2026.

[12] A. Ikhwan and D. A. P. Lubis, “Perancangan Sistem Informasi Laporan Pengaduan Masyarakat Berbasis WEB pada Dinas ESDM SUMUT,” Hello World Jurnal Ilmu Komputer, vol. 2, no. 1, pp. 1–13, 2023, doi: 10.56211/helloworld.v2i1.193.

[13] W. Alfiyani, D. A. Fatah, and F. Irhamni, “Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Pada Media Sosial X Terhadap Performa Tim Nasional Sepak Bola Indonesia Di Era Kepemimpinan Shin Tae-Yong,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 3, pp. 3969–3977, 2025, doi: 10.36040/jati.v9i3.13451.

[14] F. Alghifari and D. Juardi, “Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” Jurnal Ilmiah Informatika, vol. 9, no. 2, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[15] N. A. Tanjung, M. I. P. Nasution, and A. Ikhwan, “Penentuan Jumlah Penyaluran Dana BOS pada MTs Menggunakan Algoritma Naive Bayes,” JISTech (Journal of Islamic Science and Technology), vol. 6, no. 2, pp. 107–117, 2021, doi: 10.30829/jistech.v6i2.11437.

[16] D. E. Kurniawan, D. Swanjaya, and R. Wulanningrum, “Penerapan Metode Naïve Bayes Pada Aplikasi Ayo Playon,” Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi), vol. 6, no. 2, pp. 167–171, 2022, doi: 10.29407/inotek.v6i2.2579.

[17] M. A. Manalu and R. Eka Sari, “Data Mining Dengan Algoritma Naive Bayes Untuk Klasifikasi Minat Bakat Anak Pada Pusat Pengembangan Anak (PPA) Di Kota Medan,” Prosiding Seminar Nasional Multi Disiplin Ilmu (SENADIMU), vol. 2, no. 1, pp. 58–73, 2025.

[18] F. Harahap, W. Fahrozi, R. Adawiyah, E. T. Siregar, and A. Y. N. Harahap, “Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes,” Jurnal Unitek, vol. 16, no. 1, pp. 41–51, 2023, doi: 10.52072/unitek.v16i1.541.

[19] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "The feature extraction for classifying words on social media with the Naïve Bayes algorithm," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 11, no. 3, pp. 1041–1048, 2022, doi: 10.11591/ijai.v11.i3.pp1041-1048.

[20] A. P. Noto and D. R. S. Saputro, "Classification data mining with Laplacian Smoothing on Naïve Bayes method," AIP Conference Proceedings, vol. 2566, 030004, 2022, doi: 10.1063/5.0116519.

APPENDIX A. ANONYMIZED TRAINING DATA

The participant (athlete) names from the original manuscript have been replaced with anonymous record identifiers (A01–A30) to protect participant privacy. Informed consent and ethical authorization for the use and publication of participant data were obtained from all participants prior to data collection, under approval from the PGSI Kota Medan Ethics Committee. Nevertheless, participant identities are reported in anonymized form in accordance with the journal's ethics and privacy requirements, and the original names will only be restored if explicitly required by the journal and permitted under the existing ethical approval. Only the information necessary for reproducibility of the analysis (i.e., the anonymized dataset in Appendix A) is retained and reported.

Table A1. Anonymized training dataset of 30 wrestling-athlete records

No.	Athlete ID	Speed	Balance	Agility	Status
1	A01	Fast	Good	High	Successful
2	A02	Fast	Poor	High	Successful
3	A03	Normal	Good	High	Successful
4	A04	Normal	Good	High	Successful
5	A05	Fast	Fair	Medium	Successful
6	A06	Normal	Good	Medium	Successful
7	A07	Fast	Good	Medium	Successful
8	A08	Normal	Fair	Medium	Unsuccessful
9	A09	Fast	Fair	Low	Successful
10	A10	Normal	Poor	Medium	Unsuccessful
11	A11	Fast	Poor	Medium	Successful
12	A12	Normal	Fair	Low	Unsuccessful
13	A13	Fast	Good	Medium	Successful
14	A14	Normal	Good	High	Successful
15	A15	Fast	Fair	High	Successful
16	A16	Normal	Good	High	Successful
17	A17	Slow	Fair	Low	Unsuccessful
18	A18	Normal	Poor	Medium	Unsuccessful
19	A19	Fast	Fair	Medium	Successful
20	A20	Fast	Good	Medium	Successful
21	A21	Normal	Fair	High	Unsuccessful
22	A22	Fast	Good	High	Successful
23	A23	Slow	Fair	High	Successful
24	A24	Normal	Poor	Low	Unsuccessful
25	A25	Fast	Fair	High	Successful
26	A26	Fast	Good	Medium	Successful
27	A27	Fast	Poor	Medium	Unsuccessful
28	A28	Normal	Good	Medium	Unsuccessful
29	A29	Fast	Good	Medium	Successful
30	A30	Normal	Fair	Medium	Unsuccessful