ISSN: 2963-2536 (media online)
Proceedings of Malikussaleh International Conference on Multidisciplinary Studies (MICoMS)
Organized by center of Excellence of Technology Natural Polymer and Recycle Plastics

International Conference Proceedings 00046, (2024) DOI: https://doi.org/10.29103/micoms.v4.2024

Implementation of the Naïve Bayes Method in a Web-Based Fish Species Classification System

Rizki Suwanda[1*], Muhammad Fikry[2] and Said Fadlan Anshari[3]

[1] Universitas Malikussaleh; rizkisuwanda@unimal.ac.id
[2] Universitas Malikussaleh; muh.fikry@unimal.ac.id
[3] Universitas Malikussaleh; saidfadlan@unimal.ac.id
* Correspondence: rizkisuwanda@unimal.ac.id;

Abstract: The current fish resources are abundant, and the discovery of new species has increased the variety of fish in the ocean. These fish are categorized into three groups: demersal, pelagic, and reef fish, each with unique characteristics of their respective groups. The manual classification process for large datasets requires a long time and involves complex procedures. With the advent of data and information technology, it is now possible to recognize and identify several fish species found in the ocean, which can be classified into the three groups. To simplify this classification process, a web-based system has been developed to classify fish into these groups. The data to be processed in this research will be classified using the Naive Bayes method to address this issue. This technique utilizes large datasets to extract information that was previously unknown or inaccessible, and it can provide accurate information for various purposes. The data for this study will be collected from various internet references and direct data obtained from fish landing sites (TPI) in Lhokseumawe and North Aceh. Additionally, a literature review method will be used to complement the data analysis process. The development of the web-based system will be implemented to facilitate the classification of fish species based on the existing data.

Keywords: Fish Resource, Classification, Naïve Bayes, Website

1. Introduction

Advances in information technology require us to continuously innovate and implement the latest policies in utilizing fishery resources. Efforts by the Ministry of Marine Affairs and Fisheries have contributed to increased fish production. Considering Indonesia's vast oceans, which are home to a wide variety of fish species, information technology plays a crucial role in preserving marine sustainability [1].

Natural resources are categorized into biotic and abiotic resources. With the large number of fish species in the ocean, it is essential to classify these fish resources to make them easier to understand. The classification of fish resources facilitates planning and management to prevent rapid depletion and maximize their benefits. As a renewable resource, fisheries have a carrying capacity limit. Therefore, if their utilization does not align with sound management principles, it could lead to extinction.

Manually observing and analyzing the vast number of fish species is impractical, especially considering the continuously growing variety. Therefore, data mining is utilized to classify fish resources using the Naive Bayes method. Data mining is an automated process of discovering useful patterns or knowledge from large datasets. This process is often considered part of Knowledge Discovery in Databases (KDD), a method to uncover valuable knowledge from data[2][3].

KDD involves a series of steps to extract added value in the form of information that cannot be obtained manually from databases. The information generated is obtained through the extraction and identification of significant or interesting patterns within the data. This search process is iterative and interactive, aiming to discover new, beneficial patterns and models.

To address these challenges, a system can be developed that applies the Naive Bayes method, designed to determine the accuracy level of fish species classification [4].

2. Materials and Methods

Research methods are systematic approaches used to manage all aspects of a research activity. Research problems or questions are addressed through specific methodological approaches. Research methods encompass the study of processes and stages involved in conducting research.

Applied research, conducted to solve practical problems or create new products, often leverages the findings of basic research as a foundation for further development[5][6][7].

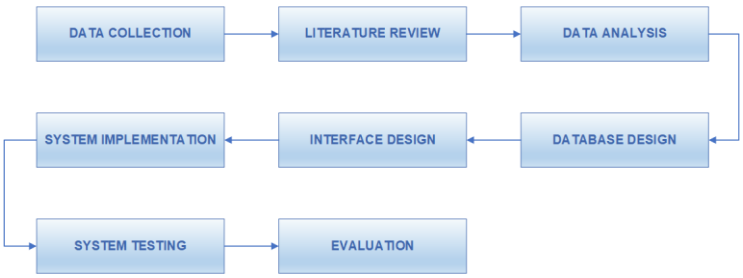

Figure 1. Research Flow Diagram

Based on the requirements to be implemented in this research system, the data is obtained from internet sources related to information about fish and several fish auction sites (TPI) in Lhokseumawe and North Aceh. The data has been collected and organized to serve as the basis for conducting data training, data testing, and accuracy testing for fish species classification.

2.1 System Scheme

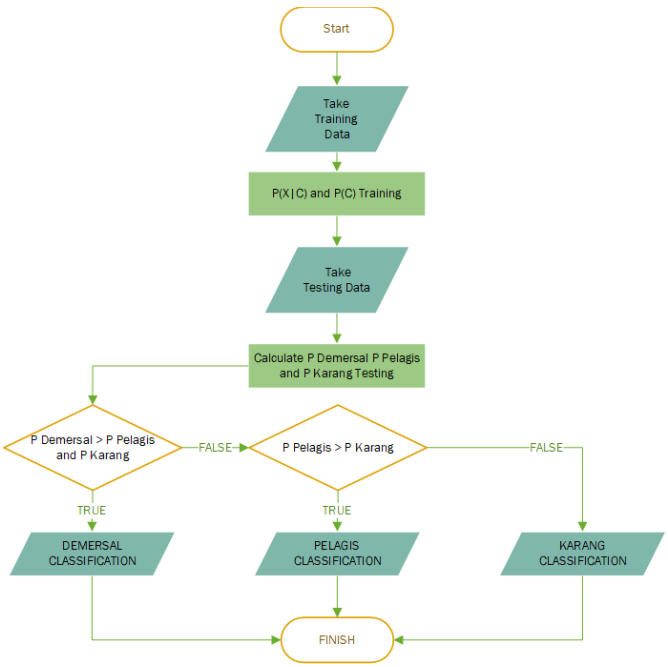

Figure 2. Naive Bayes Method System Scheme

The development of the web-based system will be built based on a system schema, serving as a tool to test the implementation process of classification testing using the available data.

3. Results and Discussion

3.1 Data Training

Training data is a collection of information obtained from field data and used as the foundational material for system training. In this study, the training data was gathered from several fish auction sites (TPI) and various online data sources. The data was manually organized and consists of a total of 200 fish data points, categorized into 66 demersal fish, 73 pelagic fish, and 61 coral fish.

Table 1. Data Training

No	Groups	Body Shape	Habitat	Mouth Position	Classification
1	Big Group	Flatened	High Sea Waters	Terminal	Pelagic
2	Big Group	Flatened	High Sea Waters	Terminal	Pelagic
3	Big Group	Torpedo	High Sea Waters	Sub Terminal	Pelagic
4	Small Group	Flat	Seabed Waters	Sub Terminal	Pelagic
5	Not Group	Flattened	High Sea Waters	Superior	Pelagic
…	…	…	…	…	…
180	Group	Flattened	Seabed Waters	Terminal	Demersal
181	Big Group	Torpedo	Seabed Waters	Terminal	Pelagic
182	Group	Flattened	Seabed Waters	Terminal	Demersal
183	Big Group	Torpedo	Seabed Waters	Terminal	Pelagic
184	Big Group	Flattened	Around Coral Reefs	Superior	Coral
…	…	…	…	…	…
199	Group	Flattened	Waters Near the Coast	Terminal	Pelagic
200	Small Group	Flat	High Sea Waters	Superior	Pelagic

3.2 Calculation of Testing Data Classification

Accuracy calculation is used to evaluate the performance of the algorithm in building the Naive Bayes classifier. In this study, 200 training data points were used to train the model, while 41 testing data points, consisting of 14 demersal fish, 13 pelagic fish, and 14 coral fish, were used to measure the model's accuracy.

a. Calculating P(Ci) → Number of Class/Labels
   P (Y=Label/Class) = Number of Class Labels / Total Number of Data

$$\mathrm{P(Y = Demersal)} = \frac{66}{200} = 0{,}33$$

$$\mathrm{P(Y = Pelagic)} = \frac{73}{200} = 0{,}37$$

$$\mathrm{P(Y = Coral)} = \frac{61}{200} = 0{,}31$$

b. Calculating P(X|Ci) → Number of Cases Matching the Same Class
   - Class Label Y = Demersal

$$\mathrm{P(Groups = Big\ Groups \mid Y = Demersal} = \frac{10}{66} = 0{,}15$$

$$\mathrm{P(Body\ Shape = Torpedo \mid Y = Demersal} = \frac{9}{66} = 0{,}14$$

$$\mathrm{P(Habitat = Around\ Coral\ Reefs \mid Y = Demersal} = \frac{3}{66} = 0{,}05$$

$$P(\text{Mouth Position} = \text{Superior} \mid Y = \text{Demersal} = \frac{14}{66} = 0{,}21$$

- Class Label Y = Pelagic

$$P(\text{Groups} = \text{Big Groups} \mid Y = \text{Pelagic} = \frac{20}{73} = 0{,}27$$
$$P(\text{Body Shape} = \text{Torpedo} \mid Y = \text{Pelagic} = \frac{53}{73} = 0{,}73$$
$$P(\text{Habitat} = \text{Around Coral Reefs} \mid Y = \text{Pelagic} = \frac{1}{73} = 0{,}01$$
$$P(\text{Mouth Position} = \text{Superior} \mid Y = \text{Pelagis} = \frac{14}{73} = 0{,}19$$

- Class Label Y Coral

$$P(\text{Groups} = \text{Big Groups} \mid Y = \text{Coral} = \frac{20}{61} = 0{,}33$$
$$P(\text{Body Shape} = \text{Torpedo} \mid Y = \text{Coral} = \frac{8}{61} = 0{,}13$$
$$P(\text{Habitat} = \text{Around Coral Reefs} \mid Y = \text{Coral} = \frac{61}{61} = 1$$
$$P(\text{Mouth Position} = \text{Superior} \mid Y = \text{Coral} = \frac{32}{61} = 0{,}52$$

c. Calculating P(Ci) * P(X|Ci) → Multiplying All the Class Variable Results
   - Demersal = P(Y = Demersal) * P(Groups ∣ Demersal) * P(Body Shape ∣ Demersal) * P(Habitat ∣ Demersal) * P(Mouth Position ∣ Demersal)
     = 0,33 * 0,15 * 0,13 * 0,04 * 0,21
     = **0,000054054**
   - Pelagic = P(Y = Pelagic) * P(Groups ∣ Pelagic) * P(Body Shape ∣ Pelagic) * P(Habitat ∣ Pelagic) * P(Mouth Position ∣ Pelagic)
     = 0,37 * 0,27 * 0,73 * 0,01 * 0,19
     = **0,0001385613**
   - Coral = P(Y = Coral) * P(Groups ∣ Coral) * P(Body Shape ∣ Coral) * P(Habitat ∣ Coral) * P(Mouth Position ∣ Karang)
     = 0,31 * 0,33 * 0,13 * 1 * 0,50
     = **0,0066495**

d. Calculate the comparison of the highest probability values to indicate the status label/class: Demersal, Pelagic, or Reef. Based on the comparison, the probability value for the Reef class is higher, with a probability value of **0.0066495**, and it is classified into the Coral fish category.

3.3 Calculation of Testing Data Classification

The classification results from the testing with 41 tests are displayed in full in the following Table 2:

Table 2. Testing Data Classification

No	Initial Classification	Test Classification
1	Coral	Coral
2	Demersal	Demersal
3	Pelagic	Pelagic
4	Coral	Demersal
5	Demersal	Demersal
6	Pelagic	Pelagic
7	Coral	Demersal
8	Pelagic	Pelagic

9	Pelagic	Pelagic
10	Pelagic	Demersal
11	Demersal	Demersal
12	Coral	Coral
13	Coral	Coral
14	Demersal	Demersal
15	Coral	Coral
16	Demersal	Demersal
17	Pelagic	Pelagic
18	Pelagic	Pelagic
19	Coral	Coral
20	Demersal	Demersal
21	Demersal	Demersal
22	Demersal	Coral
23	Coral	Coral
24	Coral	Coral
25	Coral	Coral
26	Pelagic	Pelagic
27	Pelagic	Pelagic
28	Coral	Coral
29	Demersal	Demersal
30	Pelagic	Pelagic
31	Coral	Coral
32	Pelagic	Pelagic
33	Pelagic	Pelagic
34	Coral	Coral
35	Demersal	Demersal
36	Coral	Coral
37	Pelagic	Pelagic
38	Demersal	Coral
39	Demersal	Demersal
40	Demersal	Demersal
41	Demersal	Demersal

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Number\ of\ Correct\ Predictions}}{\boldsymbol{Total\ Number\ of\ Predictions}} \boldsymbol{X\ 100\%}$$

$$= \frac{\boldsymbol{12+12+12}}{\boldsymbol{14+13+14}} = \frac{\boldsymbol{36}}{\boldsymbol{41}} \boldsymbol{X100\% = 87,80\%}$$

$$\boldsymbol{Error} = \frac{\boldsymbol{Number\ of\ Incorrect\ Predictions}}{\boldsymbol{Total\ Number\ of\ Predictions}} \boldsymbol{X\ 100\%}$$

$$= \frac{\boldsymbol{2+1+2}}{\boldsymbol{14+13+14}} = \frac{\boldsymbol{5}}{\boldsymbol{41}} \boldsymbol{X100\% = 12,20\%}$$

- Accuracy 87.80% : The test fish consist of 14 Demersal, 13 Pelagic, and 14 Coral

species, with correct predictions being 12 Demersal, 12 Pelagic, and 12 Coral species.

- Error 12,20% : The test fish consist of 14 Demersal, 13 Pelagic, and 14 Coral species, with incorrect predictions being 2 Demersal, 1 Pelagic, and 2 Coral species.

4. Conclusions and Recommendations

4.1 Conclusions

The analysis and system development conducted in this research have been successfully carried out in line with the intended objectives and expected outcomes. The research began with a literature review, followed by data requirement analysis, system requirement analysis, system design, implementation, and system testing.

The conclusion achieved is a system capable of classifying fish into several groups: Demersal, Pelagic, and Coral. The classification process becomes more accurate when a larger amount of training data is used for learning; however, this can increase the time required for classification. The follow-up process involves evaluating the system developed and completed by the research team. This step aims to ensure the system can be utilized as a learning evaluation material, expand knowledge, and serve as a reference in the future.

4.2 Recommendations

It is hoped that the results of this research can continue to be implemented and studied directly and periodically, enabling the system to classify fish with larger datasets and into more than three categories. Additionally, future classification processes may include additional parameters and attributes relevant to fish data, while simultaneously improving accuracy levels.

References

[1] Iswari Wela N. S., & Ranny. Perbandingan Algoritma KNN, C4.5, dan Naive Bayes Dalam Pengklasifikasian Kesegaran Ikan Menggunakan Media Foto. Ultimatics. (2017):117

[2] Rachmat Destriana, Rizki Suwanda, et al, Strategi Sistem Informasi, Penamuda Media. (2024)

[3] Rizki Suwanda, Zulfahmi Syahputra, and Elviawaty Muisa Zamzami, Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K, Journal of Physics: Conference Series (Vol. 1566, No. 1, p. 012058). (2020)

[4] Marlina , L., Muslim, & Siahaan, A. P, Data Mining Classification Comparison (Naive Bayes and C4.5 Algorithms). Internasional Journal Of Engginering Trends and Technology (IJETT), (2016): 383.0020

[5] Aditya, A.N, Jago PHP & MySQL. Jakarta: Dunia Komputer. (2011)

[6] Ramadhani, m., & Murti, D. H, Klasifikasi Ikan Menggunakan Oriented Fast and Rotated Brief (ORB) dan K-Nearest Neighbor (KNN), Jurnal Ilmiah Teknologi Informasi. (2018).:124

[7] Rizki Suwanda, Said Fadlan Anshari, and Wardina Ningsih, Information System for Operational Goods Management at the Career Guidance and Entrepreneurship Center Malikussaleh University. Jurnal Inotera. (2024)