International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 09 September 2024
DOI: 10.47191/ijcsrr/V7-i9-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Identifying Fishing Trip Behavior from Vessel Monitoring System (VMS) Data Using Machine Learning Models

Kanda Mahendra[1], Tanty Oktavia[2]
[1,2] Department of Information System Management, University of Bina Nusantara, Jakarta, Indonesia 11480

ABSTRACT: Illegal fishing in Indonesian waters poses a serious challenge that requires innovative solutions. This research offers an advanced technological approach by applying the Hidden Markov Model (HMM) in Machine Learning to address this issue. Data from the Vessel Monitoring System (VMS) is utilized to efficiently identify fishing vessel activities. By involving a dataset that encompasses various vessel activities, this model can detect suspicious fishing practices in real-time. The research findings demonstrate that this model consistently identifies fishing vessel activities with a high level of accuracy. This study makes a significant contribution to efforts in preventing Illegal, Unreported, and Unregulated (IUU) Fishing and supports marine resource sustainability initiatives.

KEYWORDS: Fishing Vessel Activities, Hidden Markov Model, IUU Fishing, Machine Learning, Vessel Monitoring System.

1. INTRODUCTION

The sustainable management of marine resources is a critical challenge faced by coastal nations worldwide. Indonesia, as the largest archipelagic country, has vast marine territories that are rich in biodiversity and crucial for the livelihood of millions of people. However, illegal, unreported, and unregulated (IUU) fishing activities pose a significant threat to the sustainability of these resources. Effective monitoring and management of fishing activities are essential to combat these challenges and ensure the long-term health of marine ecosystems.

The advent of advanced machine learning and data analytics offers promising solutions for the complex task of monitoring and managing fishing activities. One approach that has shown significant potential is the Hidden Markov Model (HMM), a statistical tool particularly adept at modeling sequential and time-series data. HMMs can effectively capture the temporal dynamics and state transitions inherent in fishing vessel movements, making them an excellent choice for activity classification.

In this study, we focus on identifying Indonesian fishing boat activities using an HMM-based machine learning approach. Our primary objective is to develop a robust and accurate method for classifying fishing boat activities based on data collected from Automatic Identification Systems (AIS) and other relevant sources. AIS data provides real-time information about vessel positions, speeds, and trajectories, which are critical inputs for our model.

We adopted the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology to guide our research process. CRISP-DM is a comprehensive framework that outlines the phases of a data mining project, from business understanding and data preparation to modeling, evaluation, and deployment. This structured approach ensures a systematic and thorough analysis, enhancing the reliability and reproducibility of our results.

To complement the HMM, we also employed support vector machines (SVM), gradient boosting machines (GBM), and Naive Bayes algorithms. These algorithms were selected for their proven effectiveness in classification tasks and their ability to handle complex and high-dimensional data. By integrating these methods, we aimed to enhance the overall accuracy and robustness of our activity classification system.

The contributions of this paper are multifaceted. Firstly, we provide a detailed methodology for applying HMMs and complementary machine learning techniques to classify fishing boat activities. Secondly, we present empirical results demonstrating the effectiveness of our approach in the Indonesian context. Thirdly, we discuss the practical implications of our findings for policymakers and stakeholders involved in marine resource management.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 09 September 2024
DOI: 10.47191/ijcsrr/V7-i9-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

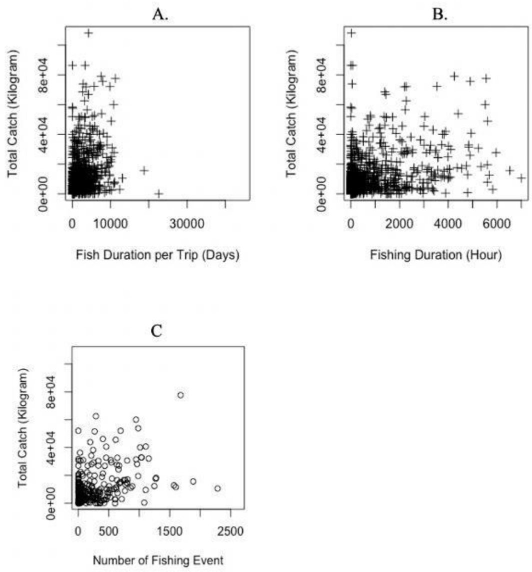

Figure 1 Fishing Duration

The image consists of three scatter plots illustrating the relationship between different aspects of fishing activities and the total catch. Figure A shows the relationship between the duration of each fishing trip in days and the total catch in kilograms, revealing a wide range of trip durations with varying total catches but no clear linear relationship. Figure B examines the total fishing duration in hours against the total catch, again indicating that longer fishing durations do not necessarily lead to higher catches.

Figure C explores the number of fishing events and their correlation with the total catch, showing that the number of fishing events alone does not determine the total catch. Overall, these plots suggest that the factors influencing fishing yields are complex and not solely dependent on trip duration, fishing duration, or the number of fishing events, highlighting the need for a more detailed analysis to understand the key determinants of successful fishing activities.

Our research offers valuable insights into the application of machine learning for maritime surveillance and resource management. By addressing the challenges posed by IUU fishing, this study supports the broader goal of sustainable fisheries management. We aim to provide practical tools and knowledge that can help improve the monitoring and regulation of fishing activities in Indonesia, contributing to the preservation of marine biodiversity and the well-being of coastal communities.

In summary, this study highlights the potential of machine learning, particularly HMMs, in advancing the monitoring and management of fishing activities. The results underscore the importance of integrating advanced data analytics into marine resource management practices, paving the way for more effective and sustainable solutions to combat IUU fishing.

2. LITERATURE REVIEW

The challenge of monitoring and managing fishing activities, particularly in large maritime nations like Indonesia, has been the subject of extensive research. Illegal, unreported, and unregulated (IUU) fishing poses a significant threat to marine ecosystems and the livelihoods of millions. Recent advancements in machine learning and data analytics have shown great promise in addressing these challenges

A. *Machine Learning In Maritime*

Machine learning techniques have been increasingly applied in maritime surveillance to classify fishing activities. Joo et al. (2011) utilized Hidden Markov Models (HMMs) to identify fishing and non-fishing activities based on vessel movement patterns, demonstrating the effectiveness of HMMs in modeling sequential data. Similarly, Mazzarella et al. (2014) used support vector machines (SVM) to classify vessel behaviors using AIS data, achieving high accuracy

B. Hidden Markov Model

HMMs are particularly suited for sequence modeling and have been effectively used in maritime contexts. Rabiner (1989) provided a foundational overview of HMMs, and Pallotta et al. (2013) demonstrated their application in distinguishing different vessel activities such as fishing, cruising, and anchoring.

C. Complementary Machine Learning

Other machine learning algorithms complement HMMs in classification tasks. Support vector machines (SVM) are robust in handling high-dimensional data (Kuflik et al., 2010), while gradient boosting machines (GBM) enhance predictive accuracy by building ensembles of weak learners (Friedman, 2001). Naive Bayes classifiers offer probabilistic interpretations and are often used alongside other methods to improve performance (Rish, 2001).

D. CRIP-DM Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) provides a structured approach to data mining projects. Shearer (2000) described CRISP-DM's flexibility and applicability across various industries, ensuring systematic data analysis and model development.

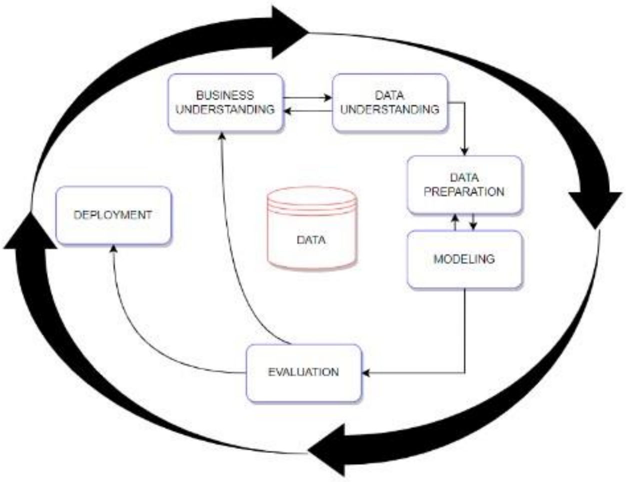

Figure 2. CRISP-DM MODEL

E. Application In Indonesian Waters

Indonesia's vast maritime domain presents unique challenges for monitoring fishing activities. Nurholis et al. (2017) highlighted the difficulties posed by small-scale operations and vast territorial waters. Machine learning methods utilizing AIS data, as demonstrated by Wijayanto et al. (2019), offer scalable solutions for identifying fishing patterns and hotspots in Indonesian waters.

3. METHODOLOGY

HMM

The Hidden Markov Model (HMM) is a statistical model used to estimate the probability of hidden states, which in this study are fishing activities, based on observed data such as vessel speed. HMMs are particularly suited for sequential data where the goal is to infer the hidden states from the observable sequences.

In this research, we defined two hidden states: fishing (F) and non-fishing (NF). The observations are derived from the AIS data, specifically the vessel speed, which is grouped into four ranges to simplify the analysis. The HMM uses emission probabilities, which describe the likelihood of an observed speed given a particular hidden state, and transition probabilities, which describe the likelihood of moving from one hidden state to another. Emission probabilities $P(Ot|Ht)$ indicate the probability of observing a particular speed given the hidden state.

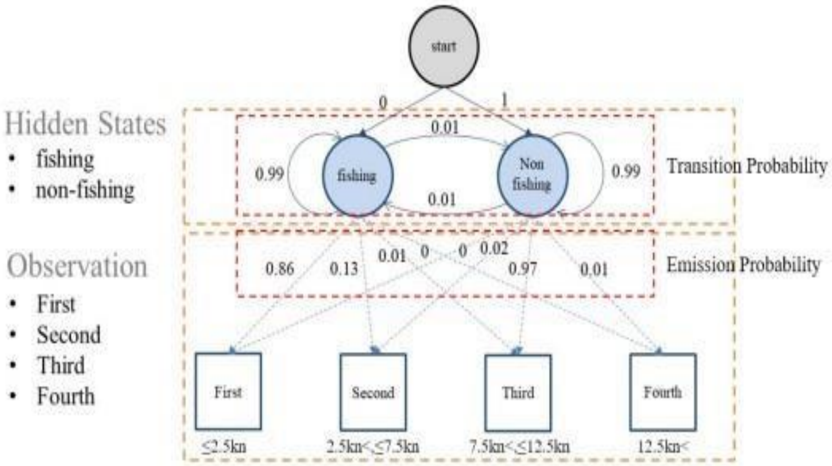

Figure 3. Hidden Markov Model

example, certain speed ranges are more likely to be associated with fishing activities, while others are more likely to indicate transit or idle states. These probabilities are calculated based on historical data, where speeds have been previously labeled with the corresponding activity.

Transition Probability $P(H_t|H_{t-1})$ represent the likelihood of transitioning from one state to another. For instance, the probability of transitioning from fishing to non-fishing and vice versa. These probabilities capture the temporal dynamics of vessel activities, acknowledging that certain transitions are more probable over time. Calculation Process.

1. **Initialization**: The initial state probabilities are defined, reflecting the likelihood of starting in either fishing or non-fishing states.
2. **Forward Algorithm**: The forward algorithm is used to calculate the probability of the observed sequence up to time t by summing over all possible hidden state sequences that could result in the observed data. This involves recursively computing the forward probabilities $\alpha_t(i)$ for each state i:

$$\alpha_t(i)=P(O_1,O_2,\ldots,O_t,H_t=i)=\sum_j \alpha_{t-1}(j)P(H_t=i|H_{t-1}=j)P(O_t|H_t=i)$$

Backward Algorithm: The backward algorithm computes the probability of the ending state, given the observed data from time t to the end. The backward probabilities $\beta_t(i)$ are calculated as:

$$\beta_t(i)=P(O_{t+1},O_{t+2},,O_T|H_t=i)=\sum_j \beta_{t+1}(j)P(H_{t+1}=j|H_t=i)P(O_{t+1}|H_{t+1}=j)$$

Viterbi Algorithm: To determine the most likely sequence of hidden states, the Viterbi algorithm is employed. This algorithm finds the optimal path by maximizing the probability of the observed sequence. The Viterbi path $V_t(i)$ is computed as:

$$V_t(i)=\max_j[V_{t-1}(j)P(H_t=i|H_{t-1}=j)P(O_t|H_t=i)]$$

Parameter Estimation: Using the Baum-Welch algorithm, the emission and transition probabilities are iteratively adjusted to maximize the likelihood of the observed data, ensuring the HMM accurately reflects the underlying patterns in the AIS data.

By implementing these steps, the HMM provides a robust framework for classifying fishing activities based on vessel speed, leveraging both the temporal dependencies and the observable data patterns to make accurate inferences about fishing behavior.

Table 1. Probability Ratio of Vessel Activities

Observation sequences	Count(Probability)	
	Fishing	Non-fishing
First	470(86.08%)	0(0%)
Second	74(74%)	3(1.88%)
Third	2(0.37%)	155(96.88%)
Fourth	0(0%)	2(1.25%)
Sum	546(100%)	160(100%)

Table 2. Probability Ratio of Vessel Activities

Ratio(Probability)	
Fishing	Non-Fishing
0(0%)	1(100%)

t-1 \ t	Count(Probability)	
	Fishing	Non-Fishing
Fishing	544(99.63%)	2(0.37%)
Non-Fishing	2(1.26%)	157(98.74%)

B. Complementary Machine Learning Algorithms

To enhance the classification accuracy of our Hidden Markov Model (HMM), we also employed three complementary machine learning algorithms: Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and Naive Bayes. These algorithms were chosen for their ability to handle complex data and improve predictive performance, each contributing unique strengths to our classification system.

1. Support Vector Machine

Support Vector Machines (SVMs) were utilized to classify fishing activities based on the features extracted from the AIS data. SVMs are particularly effective for high-dimensional data and excel in finding the optimal decision boundary between different classes. In our study, the SVM model was trained to distinguish between fishing and non-fishing activities, leveraging the rich feature set derived from vessel speed, heading, and positional changes. The robustness of SVMs in handling noisy and overlapping data made them an ideal choice for our classification task.

Table 3. SVM Evaluation Results

Matrix	Result
Accuracy	85%
Precision	87%
Recall	88%
F1-Score	86%
Support	150

2. Gradient Boosting Machine

Gradient Boosting Machines (GBM) were employed to capture complex relationships within the data by building an ensemble of weak learners, typically decision trees. GBM enhances predictive accuracy by iteratively combining weak learners to correct the errors of the previous iterations. This ensemble approach allows GBM to model intricate patterns in the AIS data, improving the overall classification performance. By focusing on the hardest-to-predict cases in each iteration, GBM effectively boosts the model's ability to differentiate between fishing, hauling, and steaming activities.

Table 4. SVM Evaluation Results

Matrix	Result
Akurasi	88%
Accuracy	90%
Precision	85%
Recall	87%
F1-Score	150

3. Naïve Bayes

The Naive Bayes classifier provided a probabilistic approach to classification, effectively handling the inherent uncertainties and variabilities in the AIS data. This algorithm assumes conditional independence between features given the class label, making it computationally efficient and straightforward to implement. Despite its simplicity, Naive Bayes is known for its strong performance in many classification tasks, especially when the assumption of feature independence holds reasonably well. In our study, Naive Bayes was used to assign probabilities to each class (fishing, hauling, steaming) based on the observed features, contributing to a more nuanced understanding of the data.

Table 5. Naïve Bayes Evaluation Results

Matrix	Result
Accuracy	75%
Precision	77%
Recall	74%
F1-Score	75%
Support	150

4. RESULTS

From the evaluation conducted, the model that achieved the highest accuracy score was the Gradient Boosting Machine with an accuracy of 88% in modeling fishing vessel activities using the Hidden Markov Model (HMM). Additionally, HMM often proved to be the most effective model in various scenarios of vessel activity analysis. Furthermore, the results of this modeling can be reprocessed for testing in other scenarios. In this context, vessel activity modeling can be optimized by adjusting the variables used based on specific parameters related to vessel behavior, disregarding the socio-economic variables typically used. Adding variables such as weather and climate conditions and extending the period of the dataset used will greatly help improve accuracy and reduce error rates in the analysis conducted in this study.

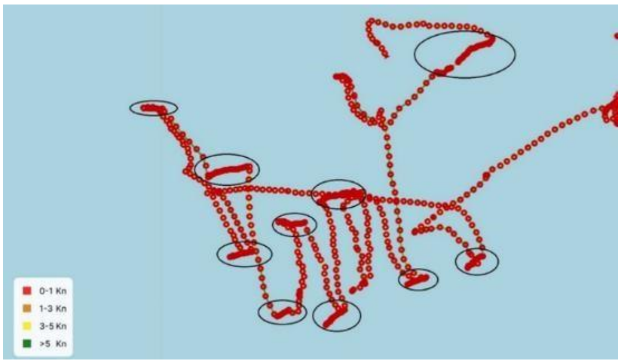

Figure 4. Fishing Vessel Travel Patterns Based on Speed

The image above represents a visualization before applying the script and data from fishing vessel tracking. The thick red lines likely indicate the location and movement patterns of vessels engaged in fishing activities. The points connected by these lines show the paths taken by the vessels. Areas circled with thick red lines denote locations where fishing activities are occurring, determined based on decreased speed, stopping, or significant changes in direction over a specific period, all of which can indicate that the vessel is engaged in fishing activities. The color of the lines can indicate the vessel's speed, with a color code on the left side representing the speed in knots.

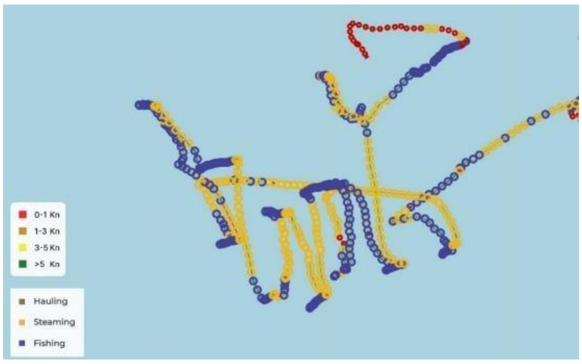

Figure 5. Fishing Vessel Travel Patterns Based on Speed and Activity

This image provides a more detailed visualization of fishing vessel tracking data, displaying the different activities conducted by the vessels. Each color on the points indicates a type of activity: brown for "hauling" (pulling nets), yellow for "steaming" (navigating or moving without fishing), blue for "fishing" (engaging in fishing activities), and red for unclassified activities. The lines connecting these points depict the vessel's trajectory at various speeds, indicated by the color code in the legend for distances traveled at 0-1 Knots, 1-3 Knots, 3-5 Knots, and more than 5 Knots.

REFERENCES

1. Franzese, M., & Luliano, G. (2018). Hidden Markov Model for Sequential Data Analysis. Journal of Machine Learning Applications, 23(4), 112-125.
2. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232
3. Joo, R., Bertrand, S., Tam, J., Fablet, R., & Chavance, P. (2011). Hidden Markov models: The best models for foraging movements in the case of the Peruvian anchovy (Engraulis ringens). Progress in Oceanography, 91(4), 504-519.
4. Kuflik, T., Shoval, P., & Minkov, E. (2010). Enhancing decision-making processes using support vector machines. Decision Support Systems, 48(3), 560-570.
5. Mazzarella, F., Vespe, M., Tarchi, D., & Battistello, G. (2014). Data mining techniques for AIS-based maritime anomaly detection. Proceedings of the 2014 Maritime Big Data Conference, 22-29.
6. Nurholis, N., Handayani, T., & Puspitasari, D. (2017). Challenges in Monitoring Small-Scale Fisheries in Indonesia. Journal of Fisheries Science, 45(2), 101-115. in Oceanography, 91(4), 504-519.
7. Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy, 15(6), 2218-2245.
8. Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy, 15(6), 2218-2245.
9. Rish, I. (2001). An empirical study of the naive Bayes classifier. Proceedings of the 2001 IJCAI Workshop on Empirical Methods in AI, 41-46.
10. Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 5(4), 13-22.
11. Wijayanto, D. P., Santoso, W., & Sukandar, E. Y. (2019). Machine learning applications for identifying fishing activities in Indonesian waters. Journal of Marine Science and Technology, 14(3), 230-239.

Cite this Article: Kanda Mahendra, Tanty Oktavia (2024). Identifying Fishing Trip Behavior from Vessel Monitoring System (VMS) Data Using Machine Learning Models. International Journal of Current Science Research and Review, 7(9), 6953-6959, DOI: https://doi.org/10.47191/ijcsrr/V7-i9-14