ORIGINAL RESEARCH (AR) DOI: 10.4274/jems.2024.99075

Journal of ETA Maritime Science 2024;12(4):404-417

A Dynamic Discretization Algorithm for Learning BN Model: Predicting Causation Probability of Ship Collision in the Sunda Strait, Indonesia

Iis Dewi Ratih[1], Ketut Buda Artana[1], Heri Kuswanto[2], Dhimas Widhi Handani[1], Renata Zahabiya[3]

[1]Institut Teknologi Sepuluh Nopember, Department of Marine Engineering, Surabaya, Indonesia
[2]Institut Teknologi Sepuluh Nopember, Department of Statistics, Surabaya, Indonesia
[3]Institut Teknologi Sepuluh Nopember, Department of Business Statistics, Surabaya, Indonesia

Abstract

Ship collisions represent a significant category of maritime accidents with far-reaching consequences that cause damage to the involved ship and neighboring vessels. This poses a threat to the marine environment, leading to potential oil spills and the triggering of additional maritime accidents. Therefore, predicting the frequency of ship collisions by identifying the contributing factors is crucial as an initial step in preventing and mitigating their occurrence. Causation probability refers to the likelihood of events resulting from a ship collision. The contributing factors to ship collisions include weather conditions, technical failure, insufficient resources, navigation errors, human error, and the failure of other vessels. The Bayesian Network (BN) machine learning method is capable of predicting ship collisions. This method delineates the relationships among diverse and complex random variables in the form of a diagram grounded in conditional probability theory. It considers both categorical and continuous variables. The prediction of ship collisions through the application of the BN involves the use of a dynamic discretization algorithm, which offers advantages over static discretization. In this research, the causation probability of ship collisions in the Sunda Strait, Indonesia was predicted. This endeavor is necessary because of the distinct characteristics inherent to each geographical area, which implies the likelihood of varying causation probabilities across regions. The resulting predictive model for the likelihood of ship collisions in the Sunda Strait, Indonesia, derived from the implementation of the BN with the dynamic discretization algorithm, yields causation probabilities of head-on collision at 2.74×10^{-4}, overtaking at 9.84×10^{-4}, and crossing at 8.41×10^{-5}. The model demonstrated an overall accuracy of 94.74%.

Keywords: Dynamic discretization, BN, Ship collision

1. Introduction

Indonesia, a maritime nation consisting of an archipelago where two-thirds of its territory is water, is facing distinctive challenges. Economic activities within the country heavily rely on maritime transportation routes, resulting in significant maritime traffic, as noted by Arfian [1]. This congestion presents operational complexities for ships, frequently leading to accidents resulting in both material and human losses. From 2019 to 2022, a considerable number of maritime accidents occurred in Indonesia, as documented by the National Transportation Safety Committee (KNKT [2]). Among these incidents, 28% involved ship fires or explosions, followed by ship sinkings (27%), ship collisions (23%), ship groundings (12%), and miscellaneous incidents (10%). Ship collisions are particularly concerning due to their notable ramifications, including loss of life, vessel damage, and environmental hazards like oil spills, as noted by Yulianti [3]. Additionally, ship collisions have the potential to trigger subsequent accidents such as sinkings, fires, and explosions. According to data from the Naval Base Command and Control Center (Puskodal) and investigations conducted by KNKT, seven ship accidents were recorded in the Sunda Strait from 2007 to 2019, with collisions being the most prevalent, contributing to four of these accidents.

The Sunda Strait is one of the most important straits in Indonesia because it is located on the shipping route

Address for Correspondence: Ketut Buda Artana, Institut Teknologi Sepuluh Nopember, Department of Marine Engineering, Surabaya, Indonesia
E-mail: kb.artana@gmail.com
ORCID iD: orcid.org/0000-0002-0302-3641

Received: 22.03.2024
Last Revision Received: 27.09.2024
Accepted: 22.11.2024

To cite this article: I. D. Ratih, K. B. Artana, H. Kuswanto, D. W. Handani, and R. Zahabiya, "A Dynamic Discretization Algorithm for Learning BN Model: Predicting Causation Probability of Ship Collision in the Sunda Strait, Indonesia." *Journal of ETA Maritime Science*, vol. 12(4), pp. 404-417, 2024.

categorized as the Indonesian Archipelagic Sea Lane (ALKI) I from south to north with a high-density traffic route from Java Island to Sumatra Island, mostly traversed by passenger ships. As one of the straits traversed by ALKI I, the Sunda Strait has the widest shipping lane in the south with a distance of 52 nautical miles, while the narrowest shipping lane corridor in the Sunda Strait is located in the northern part with a distance of 2.2 nautical miles. The narrowness of this shipping lane is caused by several navigation hazards, such as reefs, shallows, and shipwrecks. The increase in the number of ships traversing the Sunda Strait, categorized as ALKI, prompted the Indonesian government to implement *Traffic Separation Scheme*-TSS on July 1, 2020 (Figure 1).

The noteworthy impact of ship collisions has spurred the International Maritime Organization to institute regulations specifically addressing this issue, known as The International Regulations for Preventing Collisions at Sea 1972 (Collision Regulations/COLREGS). These regulations serve as the guiding framework for all ship operational processes, mandating that ship crews have a comprehensive understanding of the established rules. Despite stringent adherence to these regulations, the practical reality is that ship collisions remain unavoidable because of other contributing factors. According to KNKT investigation reports, three primary factors stand out as the causes of ship collisions in Indonesia: human factors, technical factors, and weather factors. This underscores the necessity for regulations aimed at preventing ship collisions to be complemented by understanding the additional factors that contribute to such accidents. As a proactive measure to prevent and mitigate ship collisions, it becomes imperative to predict them through the identification of the contributing factors.

Several machine learning methodologies, including Decision Tree, Random Forest, Naïve Bayes, Bayesian Network (BN), Ensemble Bagging, and XGBoost. A comprehensive literature review conducted by Chen et al. [4] systematically compared analytical approaches to assess the probability of ship collisions, including those based on historical data, Fault Tree Analysis (FTA), and BN modeling. Their study revealed that BN is the most effective method for estimating the likelihood of ship collisions. This conclusion finds support in the work of Hasugian et al. [5], which underscores the suitability of the BN methodology for discerning cause-and-effect relationships among factors influencing maritime accidents.

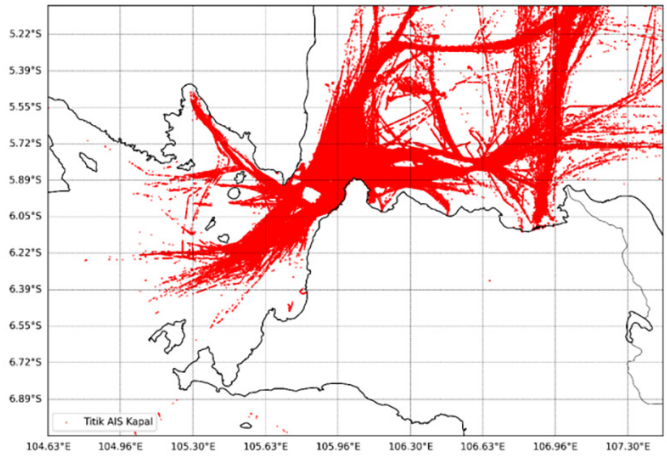

Figure 1. *Ship trajectories along the Sunda Strait*

Research on ship collisions utilizing BNs has traditionally emphasized categorical variables while overlooking the impact of continuous variables, such as wind speed, wave height, and ship length, in contributing to these incidents. Although some studies have incorporated continuous variables, their use of static discretization algorithms, converting continuous variables into predefined categories, has been prevalent. An illustrative instance was observed in the study conducted by Zamzuri and Isa [6], wherein wind speed and wave height variables were discretized into categorical variables with predefined categories.

Fenton and Neil [7] posited that static discretization algorithms can compromise the accuracy of a model, which limits its applicability to real-world scenarios. To overcome these limitations, dynamic discretization algorithms offer a solution by analyzing continuous variables within BNs without requiring transformation into predefined categories. The modeling framework employing the BN method, coupled with dynamic discretization algorithms, articulates the intricate relationships among random variables through a diagram based on conditional probability theory. This method not only accommodates categorical variables and integrates continuous variables seamlessly into the modeling process.

Ship collision incidents can be broadly classified into two categories: collisions between ships and stationary objects and collisions involving two or more moving ships. This study specifically focused on the latter category, focusing on collisions among multiple moving ships. The main objective is to predict the probability of such ship collisions in the Indonesian context. This prediction can be assisted by applying a BN employing a dynamic discretization algorithm. After deriving the prediction outcomes and determining the pivotal factors contributing to the incidence of ship collisions, the resulting model can be effectively utilized for calculating the causation probabilities (Pc). Causation Probability refers to the likelihood of events resulting from ship collisions. Karlsen and Kristiansen [8] outlined that the contributing factors to ship collisions include natural elements, technical malfunctions, insufficient resources, navigation errors, human errors, and the failure of other vessels.

Earlier investigations on the frequency of ship collisions in Indonesia relied on default values derived from IWRAP, which were determined by analyzing the causation probability in the Akashi and Dover Straits. Nevertheless, Montewka et al. [9] asserted that causation probability values should be customized according to the particular conditions of the studied waters, as each region possesses distinct characteristics, conditions, and cultures. Consequently, this study employs BN modeling to characterize and compute the causation probability in Indonesian waters, with a specific focus on the Sunda Strait as a case study. Numerous research studies have investigated the risk of ship collisions. For instance, the works of Nurmawati et al. [10], Sutrisno and Dinariyana [11], and Wuryaningrum and Handani [12] delved into ship collision risk analysis, employing default Pc values from Fuji and Shiobara [13] and Macduff [14]. These studies posit that the Pc values derived from analyses of the Akashi and Dover Strait can be universally applied. In the context of analyzing ship collisions in the Sunda Strait, several studies, including Pratiwi et al. [15], Arfian et al. [16], and Sukma et al. [17], utilized Pc values derived from modeling, employing the FTA modeling method. The modeling methodologies used in these studies entailed the determination of influencing factors and their respective probabilities, primarily referencing existing literature rather than relying on historical data from Sunda Strait accidents. Furthermore, investigations by Mulyadi et al. [18] and Purnomo et al. [19] applied the BN method to model causation probability. However, these studies adopted a simplified approach to BN modeling due to limited reference data, underscoring the need for subsequent development and updating based on the latest available data and the creation of more intricate networks, tasks that will be undertaken in the present study.

The novelty of this research, compared to previous studies, lies in the use of a BN model to capture the complex relationships between various factors influencing ship accidents, with an approach that adjusts causation probability values based on the specific conditions of the studied waters. This approach is in contrast to previous research, which often relied on default values or values derived from other regions without specific consideration of local characteristics. This research aimed to elucidate the application of BN for modeling the determinants of ship collisions and deriving Pc values for various collision scenarios (head-on, overtaking, and crossing) in Indonesian waters. Subsequently, these values were used to calculate the frequency of ship collisions in the Sunda Strait. The selection of the Sunda Strait as the focal area for this study was predicated on its status as the second-largest area in Indonesia with a notable history of ship collisions. Moreover, the strait experiences a considerable volume of ship crossings, transporting numerous passengers daily. Therefore, a meticulous analysis of the ship collision frequency in this region is imperative to uphold safety standards and mitigate the potential losses resulting from ship collisions.

2. Materials and Methods

2.1. BN

BN is one of the simple Probabilistic Graphical Models built from graph theory and probability theory, and it serves as a well-established machine learning method by utilizing conditional probability as its foundation. In its development, the BN method, which is a BN consisting of both categorical and continuous variables, was developed [7]. A BN consists of two main parts: a qualitative part in the form of a graphical structure called a Directed Acyclic Graph (DAG) and a quantitative part in the form of a set of conditional probabilities. Figure 2 depicts the structure of the BN, which consists of nodes and edges, where X_j represents a categorical variable and Z_k represents a continuous variable. If an edge from node X_l points to node X_j then X_l is referred to as the parent and X_j is the child of X_l. Nodes without parents are called root nodes, while nodes without children are referred to as leaf nodes Korb and Nicholson [20].

In constructing a BN model, key assumptions should be considered. First, the network graph structure is directed and acyclic, facilitating the understanding of causal relationships among variables. Second, conditional independence is assumed, meaning that each variable is independent of others given its parent variables. Complete information on variable relationships is presumed to be available, facilitating risk estimation. The variables depend solely on their parent variables, without considering other variables in the network. The independence of parameters implies

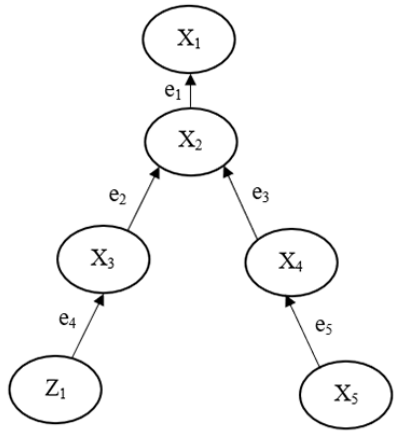

Figure 2. *Example structure of a Directed Acyclic Graph (DAG) for a BN*

that they are unrelated unless determined by the network structure. Finally, unmeasured confounding factors were assumed to be absent. Adhering to these assumptions improves the interpretation of the results, leading to more accurate conclusions about the variable relationships.

2.1.1. Estimation of probability values for nodes with categorical variables

BN involves estimating the probability values, starting with the determination of the prior probability values. The prior probabilities for categorical variables can be computed by utilizing straightforward probability functions, as outlined in the

$$P(X_{l(j)}) = \frac{n_{l(j)}}{n} \quad (1)$$

with $P(X_{l(j)})$ defining the probability of the occurrence of the l-th categorical variable or the probability of the parent node's occurrence for variable X_j where $j = 1, 2, \ldots$,p and $l = 1, 2, \ldots$, p with $l \neq j$, $n_{l(j)}$ indicating the count of the occurrence of the l-th categorical variable, and n representing the total count of all events [21].

The structure of a BN is built using a statistical approach known as conditional probability, which is defined as the probability of an event occurring based on other events that have already occurred. If analogized in terms of parent and child, the conditional probability of the child is obtained based on the conditions experienced by the parent earlier, Sari [21]. Conditional probability values for categorical variables are presented in Equation (2).

$$P(X_j|X_{l(j)}) = \frac{P(X_j, X_{l(j)})}{P(X_{l(j)})} \quad (2)$$

with X_j indicating the categorical child node with values $j = 1, 2, \ldots$,p, $X_{l(j)}$ representing the parent node of variable X_j with values $l = 1, 2, \ldots$, p where $l \neq j$, $P(X_j|X_{l(j)})$ defining the conditional probability of X_j given the value of $X_{l(j)}$, $P(X_j, X_{l(j)})$ defining the joint probability of X_j and $X_{l(j)}$, and $P(X_{l(j)})$ defining the probability value of the parent node [7].

The joint probability distribution is the probability of the simultaneous occurrence of all events. In a BN, the joint probability distribution for categorical variables is presented in Equation 3.

$$P(X_1, \ldots, X_p) = P(X_1|X_2, X_3, \ldots, X_p) \ldots P(X_{p-1}|X_p)P(X_p) \quad (3)$$

With $P(X_{p-1}|X_p)$ defining the conditional probability of X_{p-1} given the value of X_p and $P(X_p)$ indicating the probability value of variable X_p Fenton and Neil [7]. If additional information is available that, when event X_j has occurred, there may be a change in the initial estimate regarding the likelihood of event $X_{l(j)}$ occurring, the probability of the occurrence of event $X_{l(j)}$ now is the conditional probability due to the occurrence of event X_j and is referred to as the posterior probability. The posterior probability calculation for categorical variables is as follows:

$$P(X_{l(j)}|X_j) = \frac{P(X_j|X_{l(j)})P(X_{l(j)})}{P(X_j)} \quad (4)$$

with $P(X_{l(j)}|X_j)$ defining the probability of $X_{l(j)}$ given the value of X_j or the posterior of $X_{l(j)}$ and $P(X_{l(j)})$ indicating the prior probability of $X_{l(j)}$ Fenton and Neil [7].

2.1.2. Estimation of probability values for nodes with contionous variables

The prior probability values for continuous variables in the BN follow the probability distribution of the data. The data distribution pattern must first be determined using a statistical goodness of fit test such as Kolmogorov-Smirnov test. This test is used to determine the deviation or the largest difference between the observed probability or empirical probability and the theoretical probability, Basuki et al. [22]. Conditional probability values for continuous variables in the BN are analogized in the form of parent and child, as found in the following:

$$f(Z_k|Z_{m(k)}) = \frac{f(Z_1, \ldots, Z_k, \ldots Z_q)}{f(Z_{m(k)})} \quad (5)$$

with Z_k indicating the continuous child node with values k = 1, 2, …, q, $Z_{m(k)}$ representing the parent node of variable Z_k with values $m = 1, 2, \ldots$, q where $m \neq k$, $f(Z_k|Z_{m(k)})$ defining the conditional probability of Z_k given the value of $Z_{m(k)}$, $f(Z_1, \ldots, Z_k, \ldots Z_q)$ defining the joint probability, and $f(Z_{m(k)})$ indicating the probability of the parent node. The values of the joint probability distribution for continuous variables in the BN can be expressed as follows:

$$f(Z_1, \ldots, Z_q) = f(Z_1|Z_2, Z_3, \ldots, Z_q) \ldots f(Z_{q-1}|Z_q)f(Z_q) \quad (6)$$

With $f(Z_{q-1}|Z_q)$ defining the conditional probability of Z_{q-1} given the value of Z_q and $f(Z_q)$ indicating the probability value of variable Z_q, Fenton and Neil [7]. Bayes' theorem is also used to determine the posterior probability values for continuous variables, as follows:

$$f(Z_{m(k)}|Z_k) = \frac{f(Z_k|Z_{m(k)})f(Z_{m(k)})}{f(Z_k)} \quad (7)$$

With $f(Z_{m(k)}|Z_k)$ defining the probability of $Z_{m(k)}$ given the value of Z_k or the posterior of $Z_{m(k)}$ and $f(Z_{m(k)})$ indicating the prior probability of $Z_{m(k)}$.

2.1.3. Estimation of probability values for mixed variables

The probability of mixed variables in the BN can be differentiated into two cases, namely, when the continuous child node has a categorical parent node and when the categorical child node has a continuous parent node. The first condition can be explained using equation 11, and the second condition can be explained using Equation 8, Pati [23].

$$f(Z_k, X_j) = f(Z_k|X_j)P(X_j) \tag{8}$$

$$P(X_j, Z_k) = P(X_j|Z_k)f(Z_k) \tag{9}$$

Then, the posterior probability for mixed variables can be expressed in two equations, adjusting to the two conditions mentioned earlier, as found in Equations (10) and (11).

$$f(Z_k|X_{l(k)}) = \frac{P(X_{l(k)}|Z_k)f(Z_k)}{P(X_{l(k)})} \tag{10}$$

$$P(X_j|Z_{m(j)}) = \frac{f(Z_{m(j)}|X_j)P(X_j)}{f(Z_{m(j)})} \tag{11}$$

with Z_k indicating the continuous child node with values k = 1, 2, …, q, X_j indicating the categorical child node with values j = 1, 2, …, p, $Z_{m(j)}$ indicating the continuous parent node for categorical variable and $X_{l(k)}$ indicating the categorical parent node for the continuous variable.

2.1.4. Dynamic discretization algorithm

The dynamic discretization algorithm defines a continuous node as a simulation node. Suppose Z is a node with a continuous variable in the BN structure, where the range of Z is denoted by Ω_Z and the Probability Density Function (PDF) of Z is denoted by f_Z. The idea behind the dynamic discretization algorithm is to estimate the value of f_Z by partitioning Ω_Z into a set of intervals $\Psi_Z = \{w_u\}$ and defining the local constant function $\tilde{f}_Z$ for the formed interval sets. This algorithm performs two main tasks: determining the optimal discretization set and determining the optimal values for the local function $\tilde{f}_Z$ that approximates the actual value of f_Z, Fenton, and Neil [7]. The use of a dynamic discretization algorithm can improve model accuracy compared to static discretization.

2.2. Confusion Matrix and Sensitivity Analysis

The confusion matrix serves as a structured table that presents the performance of a model or algorithm in a specific manner. In this matrix, each row represents the actual class of the data, and each column denotes the predicted class of the data (or vice versa), as described by Saputro and Sari [24]. A comprehensive elucidation of this matrix is presented in Table 1.

Table 1. *Confusion matrix*

Actual	Predicted	
	Collision	No collision
Collision	True Positive (TP)	False Negative (FN)
No collision	False Positive (FP)	True Negative (TN)

The four values in the confusion matrix can be utilized to compute the performance indicators of the classification models : accuracy, sensitivity, and specificity. Accuracy = (TP + TN) / n, Sensitivity = TP / (TP + FN), and Specificity = TN / (TN + FP).

Sensitivity analysis is a method used to determine the sensitivity of a model to changes in parameters. The main advantage of sensitivity analysis is its ability to assess model accuracy when applied to a real system. By experimenting with changes in parameters within variables, one can identify where the most significant changes occur. Sensitivity analysis was performed by altering the prior probability distribution of each node within the range of 0%-100%, Ahmadi and Manurung [25].

2.3. Frequency Analysis

The frequency analysis was conducted following the framework of the IALA Waterway Risk Assessment Program (IWRAP MK II). In this study, the examined frequencies encompass ship collision occurrences in the head-on, overtaking, and crossing scenarios. IWRAP is a comprehensive program designed to analyze various aspects of ship traffic movements. The proposed method takes into account factors such as hydrographic conditions, navigation channel usage, characteristics, collision risk, and other factors that influence navigational safety in specific waterways. Particularly beneficial for mapping ship movement geometries to illustrate traffic density and compute the number of potential candidate ships at risk of collision, this program is employed in our research to facilitate the calculation of the analyzed ship collision frequencies.

The computation of the collision frequency involves the use of geometric probability values, which were obtained through analysis facilitated by the IWRAP software, along with the causation probability derived from the conducted analysis. The mathematical model used to calculate the frequency is expressed as follows:

$$\lambda_{Col} = N_G \times P_c \tag{12}$$

In the given equation, λ_{Col} denotes the frequency of ship collisions, N_G represents the number of ships potentially at risk of collision, and P_c represents the causation probability.

The geometric probability calculations for each type of ship collision were formulated as follows:

2.3.1. Head-on collision

A head-on collision is a type of ship collision that occurs at the bow section between two ships moving in opposite directions. Based on the "IWRAP Mk II Working Document: Basic Modeling Principles for Prediction of Collision and Grounding Frequencies" by Hansen [26], the value of geometric probability for sailing ships potentially experiencing head-on collisions along a specified route segment is modeled as follows:

$$N_G{}^{head-on} = L_W \sum_{i,j} P_{G\,i,j}^{head-on} \frac{V_{ij}}{V_i^{(1)} V_j^{(2)}} (Q_i^{(1)} Q_j^{(2)}) \quad (13)$$

Where $N_{Ghead-on}$ is the number of ships potentially at risk of a collision, L_W is Segment Length (m), P_G is the probability that two ships will collide in a head-on meeting situation, $V_{i(1)}$ is speed of the ship on route I (m/s), $V_{j(2)}$ is speed of the ship on route j (m/s), V_{ij} is the relative speed between the vessels (m/s), $Q_{i(1)}$ and $Q_{j(2)}$ is the number of passages per time unit for each ship type and size in each direction. Then to determine $P_{Gi,jhead-on}$ can use the following equation:

$$P_{Gi,j}{}^{head-on} = \Phi\left(\frac{\overline{B_{ij}} - \mu_{ij}}{\sigma_{ij}}\right) - \Phi\left(-\frac{\overline{B_{ij}} + \mu_{ij}}{\sigma_{ij}}\right) \quad (14)$$

Where $\Phi(x)$ is standard normal distribution function, $\mu_{ij} = \mu_{i^{(1)}} + \mu_{j^{(2)}}$ is the mean sailing distance between the two vessels, $\sigma_{ij} = \sqrt{(\sigma_i^{(1)})^2 + (\sigma_j^{(2)})^2}$ is the standard deviation of the joint distribution, and $\overline{B_{ij}} = \frac{B_i^{(1)} + B_j^{(2)}}{2}$ is the average vessel breadth.

2.3.2. Overtaking collision

Overtaking collision is a type of ship collision that occurs when a ship is behind another and moves at a higher speed with the intention of passing the ship ahead of it in the same lane and direction. For overtaking collisions, the number of geometric collision candidates for ships sailing along the route segment in direction (1) is expressed by equation (13) using the relative speed $V_{ij} = V_{i(1)} - V_{j(1)}, V_{ij} > 0$. The geometric probability of meeting [Equation (14)] becomes:

$$P_{Gi,j}{}^{overtaking} = P\left[y_i^{(1)} - y_j^{(1)} < \frac{B_i^{(1)} + B_j^{(1)}}{2}\right] - P\left[y_i^{(1)} - y_j^{(1)} < -\frac{B_i^{(1)} + B_j^{(1)}}{2}\right] \quad (15)$$

For normally distributed variables, the mean value in equation (14) should be replaced by $\mu_{ij} = \mu_{i(1)} - \mu_{j(1)}$ to handle overtaking.

2.3.3. Crossing collision

Crossing collision is a type of ship collision that occurs between two ships moving in opposite directions relative to each other (at an angle between $10° < [\theta] < 270°$). The frequency of crossing collisions depends on the angle between two lanes. The geometric amount of crossing collision candidates for crossing waterways can similarly to Equation (13) be expressed as follows:

$$N_G{}^{crossing} = \sum_{i.i} \frac{Q_i^{(1)} Q_j^{(2)}}{V_i^{(1)} V_i^{(2)}} D_{ij} V_{ij} \frac{1}{\sin\theta} \qquad \text{For } 10^o < |\theta| < 270^o \quad (16)$$

Where $V_{ij} = \sqrt{(V_i^{(1)})^2 + (V_j^{(2)})^2 - 2V_i^{(1)} V_j^{(2)} \cos\theta}$ is the relative speed between the vessels and D_{ij} defines the apparent collision diameter. The sinus term stems from the variable transformation when integrating over the area of the joint probability distribution. Note that, contrary to head-on and overtaking collisions, the distribution of the traffic spread is not relevant for crossing collisions, except for the sinus term of course. When the crossing angle approaches zero, the length of the crossing (or the time of the crossing) goes to infinity and hence does the number of collisions. For practical reasons, it is necessary to limit the crossing angle to an interval of, for example, 10° to 270°.

As mentioned D_{ij} is the geometrical collision diameter. If it is assumed that ships can be approximated by rectangular shapes, then it can be shown that:

$$Dij = \frac{L_i^{(1)} V_j^{(2)} + L_j^{(2)} V_i^{(1)}}{V_{ij}} \sin\theta + B_j^{(2)} \left\{1 - \left(\sin\theta \frac{V_i^{(1)}}{V_{ij}}\right)^2\right\}^{1/2} + B_i^{(2)} \left\{1 - \left(\sin\theta \frac{V_j^{(2)}}{V_{ij}}\right)^2\right\}^{1/2} \quad (17)$$

Where Li is length of ship i , Lj is length of ship j, Bi is width of ship i, Bj is width of ship j.

2.4. Data Sources

The dataset used in this study encompasses ship collision incidents from 2009 to 2021, consisting of a total of 44 collision cases involving two or more ships, resulting in 94 ships being involved in these collision cases. Based on the analysis, 68 ships were classified as collisions, while the remaining 26 ships were considered as non-collisions. In the accident reports, these 26 ships were determined to be innocent because they were navigating in good condition at the time. If these ships had not encountered the colliding vessels, they would not have been involved in the collision. Based on accident reports, innocent ships, namely, those that sailed in good condition or those that were stationary and then collided, were categorized as having no collision. The following gives a detailed account of the ship collision cases:

a. There is 1 case involving 3 ships, providing a total of 3 data points.

b. There are 3 cases involving multiple types of accidents:

1. KMP Safira Nusantara faced a head-on collision with the LCT Sentosa Indah Sejati. Both ships, in a narrow

situation, changed course, causing KMP Safira Nusantara to experience a crossing collision with LCT Sentosa Indah Sejati. LCT Sentosa Indah Sejati overtook the ship in front, resulting in a collision (2 head-on data points, 2 crossing data points, and 1 overtaking data point).

2. KM Mochtar Prabu Mangkunegara faced a head-on collision with another ship but eventually avoided collision. KM Mochtar Prabu Mangkunegara then experienced a head-on collision with KM Sinar Jimbaran (3 head-on data points).

3. MT New Global initially faced a crossing situation with KM Maju IX. MT New Global changed course to avoid a collision, resulting in an overtaking situation that made the collision more severe (2 overtaking data points and 2 overtaking data points).

a. There were 39 cases involving 2 ships, providing a total of 78 data points.

b. There is 1 case involving only 1 ship because the captain and the first officer of the opposing ship could not be questioned as they died in the accident, so data regarding the conditions at the time of the collision could not be obtained.

The ship accident data were sourced from multiple repositories, including the following (Table 2):

The factors causing accidents used in this research are limited to those found in the chronology of ship collisions in Indonesia, as recorded in the KNKT and Maritime Court reports only. All operational definitions and categorizations related to variables are sourced from KNKT investigation reports, reports on the results of the Shipping Court's decision, COLREGS 1972, STCW, and decisions of the Minister of Transportation and Government Regulations (Table 3).

In this study, several dependent variables are treated as independent variables. One example is the ship dimension variable, which is influenced by length, breadth, draft, and Coefficient Block, as well as the maritime environment in which the ship operates. However, in this study, we assumed that this variable is independent of maneuver. This is because the selection of variables was constrained based on the accident chronology documented in the KNKT report.

Table 2. *Link to data source*

Data Source	Link
Directorate General of Sea Transportation	https://hubla.dephub.go.id/
Investigation reports on maritime accidents at the KNKT	https://mahpel.dephub.go.id/web/putusan/s?y=&q=&c=tubrukan
European Center for Medium-Range Weather Forecasts (ECMWF)	https://www.ecmwf.int/.

Table 3. *Research variables*

Variable		Description
Y	Collision	The condition of a ship colliding with another ship
X1	Good seamanship	Crew behavior or habit according to the sailor's habits based on applicable regulations
X2	Crew competence	Eligibility for certification by duty officers based on specified standards
X3	Crew leadership	Conditions of doubt in decision making by duty officers due to limited information regarding conditions existing and/or non-existent agreement of duty officers on each ship
X4	Crew communication	Communication between ships close to each other to agree on collision prevention measures
X5	Understanding ship characteristics	The understanding of the officer on duty regarding the nature and capabilities of the ship in relation to maneuvering.
X6	Understanding environment	Understanding the water condition patterns of duty officers in relation to maneuvering
X7	Inexperience	The watch officer's term of office from the start of service in the same position on the relevant ship (<1 year)
X8	Capacity in decision making	The ability of duty officers to make decisions to avoid critical conditions that cause collisions
X9	Crew health	Health condition of the officers on duty
X10	Number of crew members	Crew size based on ship tonnage and engine power
X11	Dual task	Conditions where the duty officer is responsible for more than one workload.
X12	Crew fatigue	The mental condition of the ship's watch officer refers to the time on duty during one watch (>14 hours)
X13	Situational awareness	Awareness of duty officers to continuously obtain information about the surrounding environment and integrate it into their knowledge of the situation, analyze its impact, predict the situation in the near future and assess its impact on safety

Table 3. *Continued*

Variable		Description
X14	Visual observation	Conditions affecting the visual detection of objects around the ship, influenced by weather, screen lights, and other distractions, as well as the ability or negligence of watch officers in visual observation duties
X15	Daylight	Availability of sunlight when the ship is sailing
X16	Master presence on the bridge	Availability of the captain for guard duty and leadership on the bridge of the ship
X17	Understanding navigation and communication signs	The officer on duty's ability to understand communication and navigation codes from other ships in the form of maneuvers, light signals, or sound signals
X18	Proper utilization of navigation and communication	Navigation and communication tools for maneuvering and monitoring the movements of other ships
X19	Establish navigation and communication equipment	Complete navigation equipment required for the ship
X20	Crew responsiveness	Speed and timeliness when making decisions and taking action to avoid collisions
X21	The presence of the pilot	Availability of guides on the ship bridge when sailing in mandatory pilot waters
X22	Maneuverability	The ship's ability to change its course to avoid collision
X23	Technical failure	The ship's engine was not working properly during the operation to avoid collisions.
X24	Ship type	The type of ship involved in the collision case
X25	The type of water body	The type of water where the ship collision occurred
Z1	Wind velocity	Wind velocity at the time of collision with the ship
Z2	Wave height	The height of the water waves at the time of collision with a ship
Z3	Ship length	The horizontal distance between the leading edge of the bow height and the rear end of the stern height of the ship.
Z4	Ship breadth	The horizontal distance between the outer sides of the hull skin was measured at the main deck line.
Z5	Ship draft	The vertical distance between the waterline and the keel of the ship
Z6	Ship speed	Ship speed at the time of collision

3. Results and Discussion

3.1. Predicting Ship Collisions using BN

The application of BN in predicting ship collisions involves a series of methodical stages. These stages comprise the development of the BN structure, the utilization of the dynamic discretization algorithm for continuous variables, the estimation of probability values for each node, the creation of a confusion matrix, the validation of the model, and the execution of the sensitivity analysis. The BN structure's variable relationships are determined based on an understanding of the collision event sequence, which is informed by prior research and expert insights. The DAG construction for the BN applied to the Indonesian ship collision data was guided by these relationships, as depicted in Figure 3.

Before parameter estimation in the BN, the dataset was partitioned into training and testing data using an 80%:20% ratio. The next step is to estimate the prior probability. The calculation of prior probabilities is only performed on the root node, as presented in Table 4 for categorical variables and Table 5 for continuous variables. In addition, for other nodes, calculations were performed using conditional probability.

The calculation of the prior probability for continuous variables commences with the determination of the data distribution, setting the threshold value, and calculating the

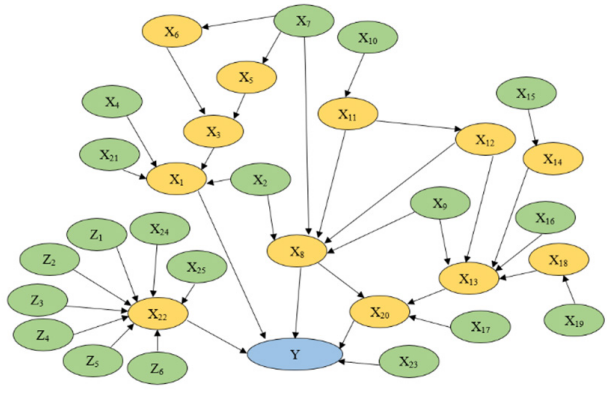

Figure 3. *Ship collision DAG*

DAG: Directed acyclic graph

prior probability based on the data distribution pattern for each variable. The data patterns for each continuous variable suggest several potentially suitable data distributions, including lognormal, gamma, normal, and triangular distributions. The selection of the appropriate distribution for each continuous variable is determined through the Kolmogorov-Smirnov test. Calculating the prior probability of continuous variables using the dynamic discretization algorithm requires threshold values that facilitate the calculation process. The thresholds for each continuous variable were determined by dividing the data for the variable according to the categories of the ship-maneuvering

Table 4. *Prior probability of categorical variables*

Variable	Category	Prior probability
Crew competence (X_2)	Proper	0.760
	Unproper	0.240
Ship communication (X_4)	Good	0.533
	Bad	0.467
Inexperience (X_7)	Yes	0.267
	No	0.733
Crew health (X_9)	Fit	0.947
	Unfit	0.053
Number of crews (X_{10})	Proper	0.893
	Unproper	0.107
Sailing time (X_{15})	Day	0.307
	Night	0.693
Master (X_{16})	Available	0.760
	Not available	0.240
Understanding navigation and communication signs (X_{17})	Good	0.787
	Bad	0.213
Navigation and communication equipment (X_{19})	Proper	0.960
	Unproper	0.040
Pilot (X_{21})	Available	0.320
	Charlie	0.013
	Not available	0.227
	Not required	0.440
Technical failure (X_{23})	Yes	0.027
	No	0.973
Ship type (X_{24})	Tanker	0.173
	Container	0.013
	General cargo	0.227
	Bulk carrier	0.027
	Passenger ship	0.080
	Ro-Ro	0.040
	Fishing ship	0.067
	Barge and tugboat	0.160
	Others	0.213
Type of water body (X_{25})	Open sea	0.493
	River	0.467
	Coastal	0.040

ability variable, namely good and limited. The thresholds are cut-off values representing the median between the good and limited categories for each variable. The threshold values are then used to determine the prior probability of continuous variables using the PDF of each distribution of continuous variables. The obtained prior probability values for each continuous variable are presented in Table 5.

Next is to calculate the conditional probability. The conditional probability values for each child node are calculated using equation 2. The joint probability for categorical child nodes with categorical parent nodes is accomplished through the utilization of Equation 3. Conversely, for categorical child nodes with continuous parent nodes, the determination is performed using equation 10. The final step involves computing the posterior probability. In the context of this study, posterior probability denotes the altered value following the acquisition of new information from the evidence set, constituted by the real-world data pertaining to each variable associated with the ship collision events. The computation of posterior probability for categorical variables is executed through Equation 4, whereas for mixed variables, equation 11 is employed. The outcomes of parameter estimation for each node, which have been acquired, yield a BN structure accompanied by probability values assigned to each node, as

Table 5. *Prior probability of continuous variables*

Variable	Units	Distribution	Parameter	Prior probability
Wind velocity (Z_1)	Knot	Triangular	a = 0.1 m = 4.5 b = 17.50	$f(Z \leq 8.26) = 0.6226$ $f(Z > 8.26) = 0.3774$
Wave height (Z_2)	Meter	Lognormal	$\mu = -0.509$ $\sigma = 0.779$	$f(Z \leq 0.7) = 0.5779$ $f(Z > 0.7) = 0.4221$
Ship length (Z_3)	Meter	Lognormal	$\mu = 4.315$ $\sigma = 0.661$	$f(Z \leq 101.9) = 0.6808$ $f(Z > 101.9) = 0.3192$
Ship breadth (Z_4)	Meter	Gamma	$\alpha = 4.257$ $\beta = 3.691$	$f(Z \leq 16.8) = 0.6183$ $f(Z > 16.8) = 0.3817$
Ship draft (Z_5)	Meter	Gamma	$\alpha = 3.301$ $\beta = 2.003$	$f(Z \leq 6.15) = 0.5223$ $f(Z > 6.15) = 0.4777$
Ship speed (Z_6)	Knot	Normal	$\mu = 6.811$ $\sigma = 2.878$	$f(Z \leq 6.35) = 0.4363$ $f(Z > 6.35) = 0.5637$

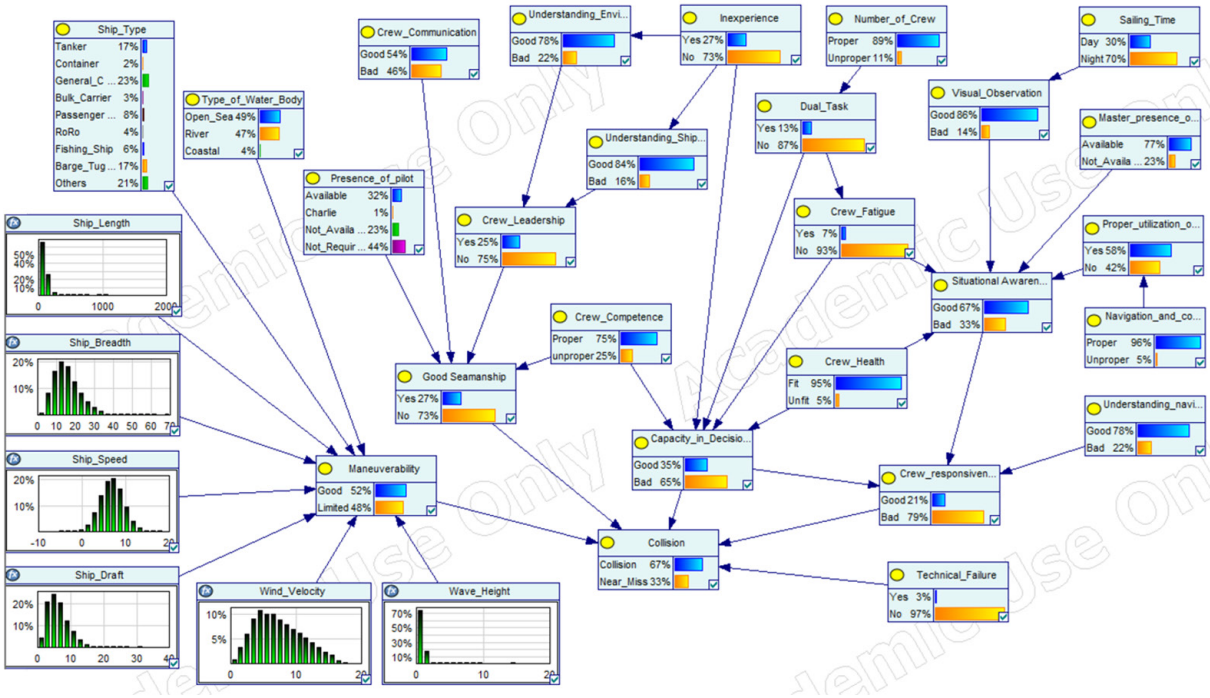

Figure 4. *BN model of ship collision causes*

depicted in Figure 4.

Figure 4 illustrates the configuration of the BN designed for ship collision scenarios. In this BN structure, the probability values are derived from joint probability calculations, forming the foundation for predictions when evidence pertaining to a collision case is identified to acquire posterior probability values. The construction of the BN structure is grounded in the training data, with the resultant probability values indicating a probability of 0.67 for a ship collision and a probability of 0.33 for no collision. The collision probability of 0.67 and the non-collision probability of 0.33 are derived from the sample data used to develop the BN model, which includes 68 ships that experienced collisions and 26 ships that did not (as explained in subsection 2.5). These results do not imply that the overall probability of a collision is 0.67. Instead, 0.67 represents the probability of a collision based on the causative factors identified from the sample data. Therefore, to calculate the causation probability, adjustments were made based on the number of ships passing through the Sunda Strait after obtaining this model.

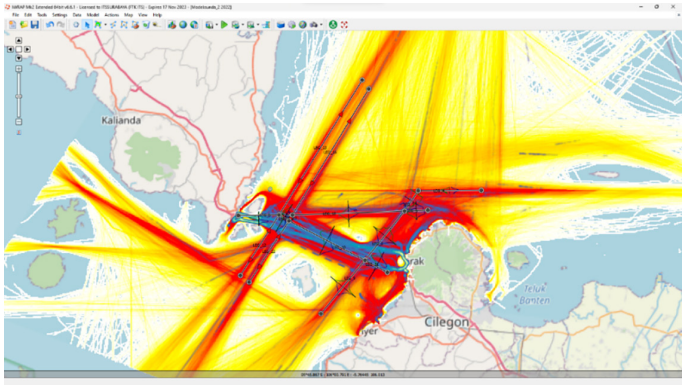

Figure 6. *Mapping of traffic density in Sunda Strait waters by IWRAP MKII*

Model validation was performed to evaluate the performance of the proposed classification model. Testing data consisting of 19 data points, is used to calculate the performance of the BN model by predicting ship collisions. The model accuracy was 94.74%. The sensitivity reached 100%, indicating that all ships involved in collisions were correctly predicted. The specificity is 80%, indicating that 20% of the ships not involved in collisions are incorrectly predicted as being involved, but this can serve as anticipation to avoid collisions.

The tornado diagram is presented in Figure 5 to illustrate the sensitivity analysis, where the diagram includes the top 10 scenarios that contribute the most to increasing or decreasing the probability of collision with a change of ±100% for each scenario. Figure 5 illustrates that the occurrence of a collision scenario is most influenced by the watch officer failing to perform duties in line with good maritime practices, making poor decisions, being slow to take evasive action, the ship not experiencing engine failure, and possessing good ship-maneuvering abilities.

Based on the tornado diagram above, the right side is the area of increasing target probability values, whereas the left side is the opposite. The bar chart shows the impact of changes in the condition probabilities listed above on changes in the target probability values. The green and red bars indicate that the probability values of the listed conditions are increasing and decreasing, respectively. If the green bar is on the right side of the baseline, then the collision probability value will also increase. Conversely, if the green bar is on the left side, then the collision probability value will decrease.

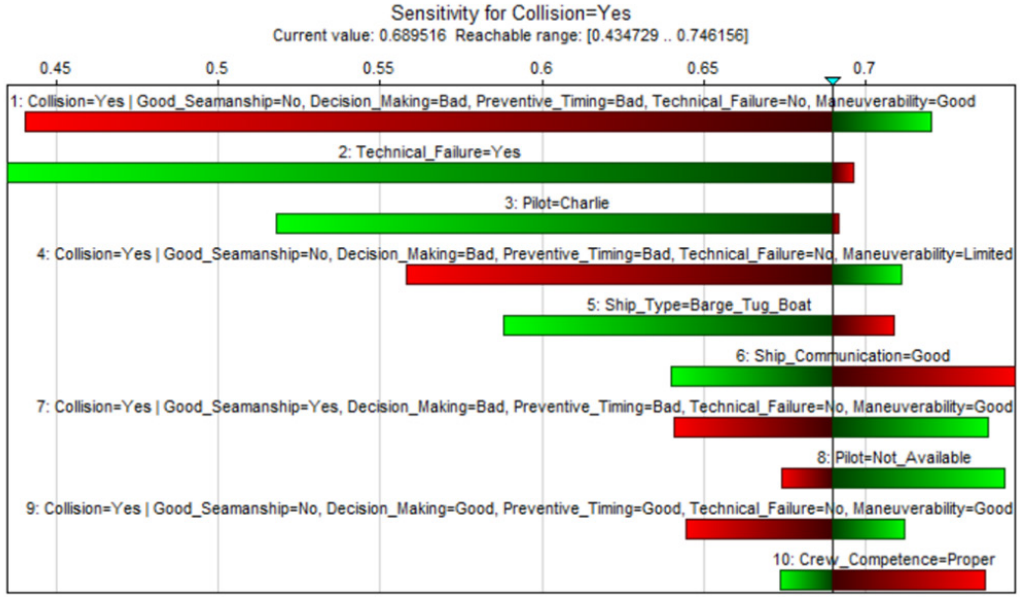

Figure 5. *Tornado diagram*

There exists a scenario with a change in the probability values of collisions that contradicts theoretical expectations, specifically, the scenario of technical failure with a "yes" category. Increasing the probability of this scenario from 2.7% to 100% results in a decrease in the probability of ship collisions from 68.9% to 43.5%. This discrepancy arises because of the limited data for the technical failure scenario with the "yes" category, which causes the sensitivity analysis results for this scenario to be suboptimal.

3.2. Causation Probability and Frequency Analysis

The causal probability (Pc) was calculated based on a BN model that was adjusted according to the annual traffic volume of ships. Therefore, the conditional probability values for each combination of scenarios are proportional to the conditional probability values for the actual conditions that occur. In the calculation of probabilities for each state, the data are separated into data for ships that have experienced collisions and data for ships that have not experienced collisions. Probability calculations are performed separately for these segmented data, followed by the accumulation of probabilities. Upon completion of the probability calculations, akin to the methodology employed in modeling a BN classification model, the subsequent step involves determining the joint probability using a consistent approach. Table 6 presents a succinct overview of the resulting Pc values (Pc analysis) corresponding to each category of collision.

In this research, to apply the obtained Pc values, a ship collision frequency analysis was conducted in one of Indonesia's waters, the Sunda Strait. To perform this analysis, Automatic Identification System data on ship movements in the Sunda Strait in 2022 were required. The data were processed to create a mapping of ship density in the Sunda Strait to understand the distribution of ship density along each existing route. This step was necessary to determine the frequency values for each type of ship collision. In this process, data processing was performed with the assistance of the IWRAP MKII software to map the water density, shipping traffic distribution, and frequency values of each type of ship collision. The density mapping results in IWRAP MKII are displayed below.

From the mapping and density distribution analysis of shipping lanes in the waters of the Sunda Strait, the frequencies of ship collisions for each collision type were acquired as follows.

The results indicate that the total frequency of ship collisions is approximately 0.025 collisions/year, but according to the historical data quoted, there have been 4 collisions in 12 years (0.33 collisions/year). The very large differences between the obtained results and the actual data could occur because the causation probability used to calculate the frequency of ship collisions is obtained based on certain factors, so it does not rule out the possibility that there are other factors that also influence the causation probability and ultimately influence the frequency of ship collisions. The findings of the analysis reveal that in the Sunda Strait, the frequency of ship collisions exhibits that the causation probability for head-on collisions is notably higher than that for overtaking and crossing collisions. The following are the causation probability values obtained from the research and other regions.

Table 6. *Frequency values of ship collisions in the Sunda Strait*

Collision type	Pc default IWRAP	Pc analysis	Frequency (Pc default IWRAP)	Frequency analysis
Head-on	$5x10^{-5}$	$2.74x10^{-4}$	0.004	0.0219
Overtaking	$1.1x10^{-4}$	$9.84x10^{-6}$	0.002	0.000179
Crossing	$1.3x10^{-4}$	$8.41x10^{-5}$	0.0046	0.003

Table 7. *Causation probabilities from literature studies*

Location	Pc ($x10^{-4}$)	Comment	References
Dover Strait	5.18	Head-on: no traffic separation	MacDuff [14]
Dover Strait	3.15	Head-on with traffic separation	MacDuff [14]
Dover Strait	1.11	Crossing: no traffic separation	MacDuff [14]
Dover Strait	0.95	Crossing with traffic separation	MacDuff [14]
Orsund, Denmark	0.27	Head-on	Karlson et al. [27]
Japanese Strait	0.49	Head-on	Fuji and Mizuki [28]
Japanese Strait	1.23	Crossing	Fuji and Mizuki [28]
Japanese Strait	1.10	Overtaking	Fuji and Mizuki [28]
Great Belt, Denmark	1.30	At bends in the lanes, the	Pedersen et al. [29]

The BN model for calculating accident causation probabilities has broad potential applications across various regions, not limited to just the Sunda Strait, provided that the values of condition variables and traffic density are adjusted according to the specific characteristics of the region. This adjustment is necessary to ensure that the BN model provides accurate and contextually relevant analyses. Thus, this model can serve as an effective tool for understanding and mitigating accident risks in various waterways by accounting for variations in environmental conditions and maritime activities.

4. Conclusion

The conclusions from the analysis regarding the BN modeling for estimating Pc values for each type of ship collision to calculate the frequency of ship collisions in the Sunda Strait are as follows:

- In conclusion, the implementation of the BN using the dynamic discretization algorithm yielded an accuracy of 94.74%, sensitivity of 100%, and specificity of 80%. The predominant factors contributing to ship collisions are good seamanship, decision-making, preventive timing, technical failure, and maneuverability.
- Based on the analysis and BN modeling, the causation probability values obtained for the Sunda Strait are as follows: Pc Head-on, Pc Overtaking, and Pc Crossing are 2.74×10^{-4}, 9.84×10^{-6}, and 8.41×10^{-5}, respectively, with a model accuracy of 93.75%.
- The frequency of ship collisions in the Sunda Strait for each type of collision (Head-on, Overtaking, and Crossing) using the default Pc values from IWRAP is as follows: 0.004 collisions/year, 0.002 collisions/year, and 0.0046 collisions/year, respectively. Meanwhile, based on the BN modeling results, the frequency values are 0.0219, 0.000179, and 0.003 collisions/year, respectively.

The research findings offer practical recommendations, emphasizing the importance of ship crew members fulfilling their duties in line with good maritime practices, possessing the ability to make sound decisions, acting promptly and accurately to avert collisions, conducting regular inspections and maintenance of ship engines, and maintaining good maneuverability to mitigate the potential for ship collisions.

The study acknowledges limitations due to data constraints, which result in some unavailable combinations in conditional probability calculations. As a result, the BN model may not be applicable to certain ship collision scenarios. Future research should address these limitations by expanding the dataset and/or involving expert opinion to encompass all potential collision scenarios for more robust and optimal outcomes.

Footnotes

Authorship Contributions

Concept design: I. D. Ratih, K. B. Artana, and H. Kuswanto, Data Collection or Processing: I. D. Ratih, and R. Zahabiya, Analysis or Interpretation: I. D. Ratih, K. B. Artana, H. Kuswanto, D. W. Handani, and R. Zahabiya, Literature Review: I. D. Ratih, K. B. Artana, H. Kuswanto, and D. W. Handani, Writing, Reviewing and Editing: I. D. Ratih, K. B. Artana, H. Kuswanto, and D. W. Handani.

Funding: The authors did not receive any financial support for the research, authorship and/or publication of this article.

References

[1] Z. Arfian, *Penilaian Risiko Tubrukan Kapal Akibat Instalasi Anjungan Lepas Pantai di Dekat Alur Pelayaran Barat Surabaya*, Surabaya: Institut Teknologi Sepuluh Nopember, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:115277737

[2] KNKT, "Laporan dan Informasi Statistik Kecelakaan", 2023. [Online]. Available: https://knkt.go.id/statistik

[3] Yulianti, "Prosedur dan Penanggulangan Keadaan Darurat di Kapal KM. Mutiara Ferindo III", UNIMAR AMNI, Semarang, 2019. [Online]. Available: http://repository.unimar-amni.ac.id/id/eprint/2482

[4] P. Chen, Y. Huang, J. Mou, and P. V. Gelder, "Probabilistic risk analysis for ship-ship collision: State-of-the-art". *Safety Science*, vol. 117, pp. 108-120, Aug 2019.

[5] S. Hasugian, M. Rahmawati, A. I. S. Wahyuni, I. Suwondo, and I. Sutrisno, "Analysis the risk of the ship accident in Indonesia with BN model approach". *Annals of the Romanian Society for Cell Biology,* vol. 25, pp. 3341-3356, 2021.

[6] Z. H. Zamzuri and Z. Isa, "Pengukuran risiko menggunakan rangkaian Bayesan: aplikasi kepada data pelanggaran kapal di Malaysia [English: Risk measurement using Bayesian networks: applications to ship collision data in Malaysia]". *Sains Malaysiana*, vol. 51, pp. 2305-2314, Apr 2022.

[7] N. Fenton, and M. Neil, *Risk Assessment and Decision Analysis with BN*, Boca Raton: CRC Press, 2013.

[8] J. E. Karlsen, and S. Kristiansen, *Analysis of Causal Factors and Situation Dependent Factors. Project: Cause Relationships of Collisions and Groundings*, Report 80-1144, Det Norske Veritas, Høvik, Norway, 1980. [Online]. Available: https://api.semanticscholar.org/CorpusID:108479276

[9] J. Montewka, F. Goerlandt, and P. Kujala, "A new definition of a collision zone for a geometrical model for ship-ship collision probability estimation". *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 5, no. 4, pp. 497-504, 2011.

[10] Nurmawati, K. B. Artana, and T. Pitana, "Penilaian risiko tubrukan kapal di sekitar buoy 12 perairan selat madura melalui proses formal safety assessment (FSA), [English: Risk assessment of ship collision around buoy 12 madura strait through formal safety assessment (FSA) process]". 2015.

[11] J. Sutrisno, and A. A. B. Dinariyana, "Risk assessment of ship collision and grounding in Surabaya west access channel due to the existence of shipwrecks". 2018.

[12] N. D. Wuryaningrum, and D. W. Handani, "Frequency analysis of ship collision and its impact on the fulfillment of supporting facilities and route changes due to implementation of Sunda Strait TSS". *IOP Conference Series: Earth and Environmental Science,* 2020.

[13] Y. Fujii, and R. Shiobara, "The analysis of traffic accidents". *Journal of Navigation,* vol. 24, pp. 534-543, 1971.

[14] T. MacDuff, "The probability of vessel collisions". *Ocean Industry*, vol. 9, pp. 144-148, Sep 1974.

[15] E. Pratiwi, K. B. Artana, and A. A. B. Dinariyana, "Ship collision frequency during pipeline decommissioning process on Surabaya west access channel (SWAC)". *Journal of Engineering Science and Technology,* vol. 14, pp. 2013-2033, Aug 2019.

[16] Z. Arfian, K. B. Artana, and A. A. B. Dinariyana, "Penilaian risiko tubrukan kapal akibat instalasi anjungan lepas pantai di dekat alur pelayaran barat Surabaya". Jul 2017.

[17] R. A. Sukma, D. W. Handani, and T. F. Nugroho, "Risk assessment of ship collision on FSO pertamina abherka and oil spill modelling due to structural damage". 2021.

[18] Y. Mulyadi, E. Kobayashi, N. Wakabayashi, T. Pitana, and Wahyudi, "Development of ship sinking frequency model over subsea pipeline for Madura Strait using AIS data". *WMU Journal of Maritime Affairs*, vol. 13, pp. 43-59, 2013.

[19] D. A. Purnomo, A. A. B. Dinariyana, and K. B. Artana, "Formal safety assessment for ship collision in Bali strait". 2019.

[20] K. Korb, and A. Nicholson, *Bayesian Artificial Intelligence*, 2nd ed., CRC Press, 2010.

[21] A. N. Sari, *Model Prediksi Kondisi Perkerasan Jalan dengan Metode Dynamic BN*, Surabaya: Institut Teknologi Sepuluh Nopember, 2016. [Online]. Available: http://repository.its.ac.id/id/eprint/167

[22] I. Basuki, Winarsih, and N. L. Adhyani, "Analisis periode ulang hujan maksimum dengan Berbagai metode". *J Agroment*, vol. 23, pp. 76-92, 2009.

[23] D. Pati, "Jointly distributed random variables". pp. 1-4, 2012.

[24] I. W. Saputro, and B. W. Sari, "Uji performa algoritma naïve bayes untuk prediksi masa studi mahasiswa". *Citec Journal*, vol. 6, pp. 5, Jan-Jun 2019.

[25] B. W. Ahmadi, and O. S. M. Manurung, "Aplikasi model BN dalam perhitungan performansi operasi keamanan laut yang dihasilkan TNI AL di wilayah timur dengan pendekatan causal mapping". *Asro Journal*, p. 63, 2015.

[26] P. Friis-Hansen, "*Basic modelling principles for prediction of collision and grounding frequencies*". Technical University of Denmark, 2007. [Online]. Available: https://iala.int/wiki/iwrap/images/2/2b/IWRAP_Theory.pdf

[27] M. Karlson, F. Rasmussen, and L. Frisk, "*Verification of ship collision frequency model*". Proceeding of the International Symposium on Advances in Ship Collision Analysis, Copenhagen, Denmark, pp. 117-121, 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:197512444

[28] Y. Fujii, and N. Mizuki, "*Design of VTS systems for water with bridges*". Proceeding of the International Symposium on Advances in Ship Collision Analysis, Copenhagen, Denmark, pp. 177-190, 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:188357660

[29] P. T. Pedersen, P. F. Hansen, and L. Nielsen, "*Probabilistic analysis of collision damages with application to passenger Ro-Ro vessels*". Safety of Passenger Ro-Ro Vessels. Dept. of Naval Architecture and Ocean Eng. Doc. pac- 001. 1995