ISSN: 2278-4632
Vol-13, Issue-08, August 2023

Enhancement of Accuracy in Predicting Marine Accidents Using Machine Learning Algorithms

V Ramalakshmi[1], Dr. K. Madhavi[2]

[1]PG-Scholar, Department of CSE (CS), JNTUA College of Engineering (Autonomous) Ananthapuramu, India.
[2]Professor, Department of CSE, JNTUA College of Engineering (Autonomous) Ananthapuramu, India.
vadiveluramalakshmi@gmail.com[1], kasamadhavi.cse@jntua.ac.in

Abstract

Accidents in marine injure a lot of people and damage many things, like ships, every year Thus, the authors of this study propose a machine learning-based maritime incident prediction model that may be used to avoid marine mishaps by anticipating and evaluating them. This work gets around the issues with different works, which aren't valuable since they aren't pragmatic. To predict marine incidents using machine learning techniques The purpose of this research is to enhance the accuracy of the maritime incident forecast system by employing Random Forest, Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbors (KNN). The Random Forest gives a better result compared to the other algorithms like SVM, DT, and KNN. Random Forest performs better in terms of precision, recall, accuracy, and F1 score. When compared to the current system, this is a substantial improvement.

Keywords: Marine accidents, Machine learning, Accident prediction, Random Forest.

I. INTRODUCTION

Any marine accident is defined as an occurrence, or series of incidents, that has resulted in any of the following occurring related to via the regular operation of a marine vessel: the escape of a human being compared to a boat, the loss, feared loss, or loss of a marine vessel, the loss of life or serious injury of a human being, a boat suffering substantial damage, becoming stuck or disabled, or being involved in an accident, significant environmental harm, or a possibility for serious damage to the environment.

Marine accidents can be caused by a variety of factors, including collisions, grounding, entanglement with suspended solids, equipment or machinery failure, fire, and explosion. Some causes of marine accidents are man-made and can be prevented by taking appropriate precautions. For example, regular inspections of marine equipment can help prevent equipment failures. However, some accidents are caused by natural events that are beyond human control.

A ship, for example, could sink in a storm with strong winds or waves. Marine accidents with unpredictable causes are omitted from the model's research and construction forecast to reduce unclear noise. The bulk of marine accidents, on the other hand, happen by hidden factors that may be measured and employed. Overall, adrift accidents are common, and they may be stopped here and there. In fact, quantitative logs are used in many papers that investigate the risk of marine accidents in order to build models that can predict mishaps. This paper presents an approach for predicting marine accidents based on accident records, most of which are thought to be somewhat predictable.

However long enough wave and wind information is gathered, more exact expectation models for impacting variables could be worked to give a more precise gauge of the possibilities of a marine accident occurring [1]. In this work, stowed away gamble factors remembered to cause marine

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

accidents were tracked down through a careful investigation of the writing and numerous surveys with specialists in the field.

To forecast marine accidents, machine learning techniques are utilized. Random forests, Machine Learning with Support Vector Machines, decision trees, and KNN algorithms are examples of machine learning algorithms. These algorithms are used to boost the prediction System's accuracy. When compared to the other methods, Random Forests produce superior outcomes.

The Random Forest algorithm is used to predict marine accidents. It gives more results to accuracy, precision, Recall and F1 scores. The random forest algorithm is the best algorithm to implement the prediction of marine accidents.

II. RELATED WORK

Funda Uğurlu et al. [2] proposed the significant business of business fishing helps many individuals overall in either straightforwardly or in a roundabout way bringing in cash. Without a boat, fishing on this colossal size is inconceivable. Along these lines, the primary piece of the continuous fishing business is fishing boats. For getting, shipping, and keeping fish, it is vital to fish boats. A huge number of individuals bite the dust every year while fishing boats crash. Disasters on fishing boats should be explored and steps should be taken to stop them if we genuinely want to keep fishing in a safeguarded way. Hence, in this survey, Bayesian association and chi-square strategies were used to look at crashes that happened on fishing boats with a full length of 7 m or all the more someplace in the scope between 2008 and 2018. Subsequently, suggestions were made to forestall future accidents. Similarly shown is the Accident (Bayes) Network, which gives an overview of how regularly setbacks happen on fishing boats. Through these organizations, it is feasible to fathom how accidents happen on fishing boats and to appraise the recurrence of accidents in different conditions. Moreover, it was found that there were areas of strength between the kind of accident, vessel misfortune, length, age, and number of lives lost.

Willem A. Wagenaar et al. [3] developed that human errors are the most significant of the numerous things that contribute. In this way, forestalling human slip-ups is a decent objective for forestalling accidents. The ongoing investigation of 100 marine accidents shows that human missteps were not as clear before the accidents occurred. Thus, a general ascent in drive or information on security will not tackle the issue. The most common factors that lead to accidents are bad habits, incorrect diagnoses, not paying enough attention, not receiving sufficient training, and having the wrong mindset. You need to take specific steps to stop unwanted behaviors in order to prevent these issues from occurring. Individuals shouldn't need to comprehend the association between their activities and occasions that happen a short time later for these progressions to occur.

Pedro Antao et al. [4] proposed a survey that used Bayesian Belief Networks (BBN) to endeavor to make a model for marine incidents. The indispensable bits of the model would be the kind of boat disaster, the sort of setback, and the results. The Portuguese Marine Authority's Mullai et al. [5] proposed a study to think about a strategy for inspecting marine crashes using a determined model. The model relies upon a lot of veritable information, for instance, the Swedish Ocean Association database, which was carefully looked at. This informational index has a summary of the boat and variable-based plans for marine crashes. Since there are countless qualities, a portion of the variables

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

have never been inspected, and most of them are not estimated. While looking at the data, once-over figures were used. To make a mental model, the database factors were accumulated into eleven key social occasions, or fabricates, which were then coordinated by their qualities and associated with a way guide of affiliations. To delineate, one non-metric variable and five measurement factors were picked: passing, boat's components (like age, total record capacity, and length), The overall number of persons prepared, marine disasters, and number of people killed. The structural equation modelling (SEM) methodology was used to look at these. 65% of the variety in the number of passings could be anticipated by the free factors "boat's properties" and "number of individuals record of marine accidents was used for this. From the past decade, 857 affirmed occasions have been accounted for. In light of different conditions, forecasts of the probability of marine accidents in Portuguese oceans were made with the help of Bayesian examination. For the assessment of the Conditional Probability Tables (CPT), certified data from the informational index was used as opposed to very capable appraisal. The standard mistake rate related to this sort of elicitation is decreased accordingly. By and large, the discoveries show the way that genuine information can be utilized to make a clear model that makes sense of the meaning of significant gambling factors and helps leaders.

Mullai et al. [5] proposed a study to think about a strategy for inspecting marine crashes using a determined model. The model relies upon a lot of veritable information, for instance, the Swedish Ocean Association database, which was carefully looked at. This informational index has a summary of the boat and variable-based plans for marine crashes. Since there are countless qualities, a portion of the variables have never been inspected, and most of them are not estimated. While looking at the data, once-over figures were used. To make a mental model, the database factors were accumulated into eleven key social occasions, or fabricates, which were then coordinated by their qualities and associated with a way guide of affiliations. To delineate, one non-metric variable and five measurement factors were picked: passing, boat's components (like age, total record capacity, and length), The overall number of persons prepared, marine disasters, and number of people killed. The structural equation modelling (SEM) methodology was used to look at these. 65% of the variety in the number of passings could be anticipated by the free factors "boat's properties" and "number of individuals ready." The Swedish data set gave a ton of the data that was utilized to make the model. In any case, since this informational index and various informational collections in the region and all around the planet share different components, the approach presented in this work might be applied to different datasets. The model is useful both on a fundamental level and, in reality. Moreover, there are suggestions for working on the library.

Andrea Coraddu et al. [6] proposed Adrift, accidents are muddled occasions welcomed on by a great many elements. Thusly, it is hard to figure out what mix of factors provoked an incident, especially when human components are involved. Since they need the help of human specialists, who each have their own impediments, predispositions, and significant expenses, present-day strategies like the Human Unwavering quality Appraisals, the Human Variable Examination and Arrangement Framework, and simple factual examination don't function admirably in many occasions. The creators need to utilize an information-driven approach that can distinguish the main human elements from records of past ocean accidents. Subsequently, a two-step strategy is given: starting, a data-driven guess the design consists of manufactured that can forecast the sort of setback considering the contributing components, and a short time later the contributing factors are situated by the sum they impact the projection. The proposed new procedure will be maintained by results from a record of veritable accidents kept by the Marine Setback Assessment Branch, a free unit of the UK Division for Transport.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

III. METHODS

This section discusses the proposed task's implementation as well as the study's resources.

A. Dataset

This paper uses a marine incident dataset containing 250 records, which includes three kinds of marine incidents: crashes, collapses, and different mishaps. The dataset includes parameters such as rain, persons on board, crew members, longitude, and latitude. Machine learning algorithms are used to predict marine accidents using this dataset, which is divided with an 80 per cent to 20 per cent split, both training and testing units are divided. Information about the dataset, including the number of classes, class names, and dataset path, is provided in an Excel file.

B. Proposed Method

The goal of this experiment is to predict marine accidents using machine learning algorithms. Several algorithms, including Random Forest, Decision Tree, KNN, and SVM, are applied to the marine accident's prediction task. Among these algorithms, The Random Forest algorithm delivers the greatest results in predicting marine accidents.

C. Apply Algorithms

A variety of Methods from machine learning can be used on cleaned-up data, with a focus on methods that provide clear and transparent decision-making processes. Some understandable techniques include:

1. Random Forest: A well-known decision tree-based ensemble method that assigns importance to each feature.

2. Decision Tree: Decision trees are transparent and can be visually represented, making it easy to choose a course of action.

3. SVM (Support Vector Machine): SVM is useful for separating data into two groups, and support vectors can be used to understand the decision-making process.

4. K-Nearest Neighbors (KNN): The K-Nearest Neighbors (KNN) algorithm is a straightforward and easily interpretable technique for classifying data points based on their closeness to one another.

Random Forest (RF)

Random forest is a notable ML pattern created by Leo Breiman and Adele Cutler. It faces the results of miscellaneous conclusion shrubs and mixes bureaucracy to follow a unique composition. It has enhanced notably taking everything in mind the evidence that it is foolproof and may be used to handle both accumulation and backslide issues.

Random Forests is a method of reducing variation by averaging many deep decision trees that have been trained on different parts of a single training set. This algorithm is utilized to forecast actions and results in a wide range of industries, including banking and e-commerce.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{n}(fi - yi)^2 \text{ ----------Eq. (1)}$$

Equation 1 shows that the mean squared error formula isused in the random forest algorithm.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

Where N is the number of data points,

fi is the value returned by the model andyi is the actual value for data point i.

Random forests contain the three components intheir architecture as shown in Fig.1 they are Training set,Testing set and Prediction.

In the Training set, there are multiple boxes representing the training samples. These samples are used totrain multiple Decision trees, which are shown as connected by arrows. Decision Trees are a type of algorithm that can beused for Classification and regression tasks.

The testing set section has a single box representingthe testing data. This data is used to test the performance of the trained Decision trees. The testing data is connected by anarrow to the voting section.

The projections from each Decision tree are mergedto create a final forecast during the voting step. For classification problems, the Random Forest outcome is the class chosen by the majority of trees. The regression analysisreturns the mean or average forecast of the specific tree.

The final output of the process is shown in the prediction section. This is where the final made by the Random Forest is displayed. Overall, this diagram provides avisual representation of how a Random Forest algorithm works. It shows how training data is used to train multiple Decision trees, how testing data is used to evaluate their performance, and how their predictions are combined to produce a final prediction.

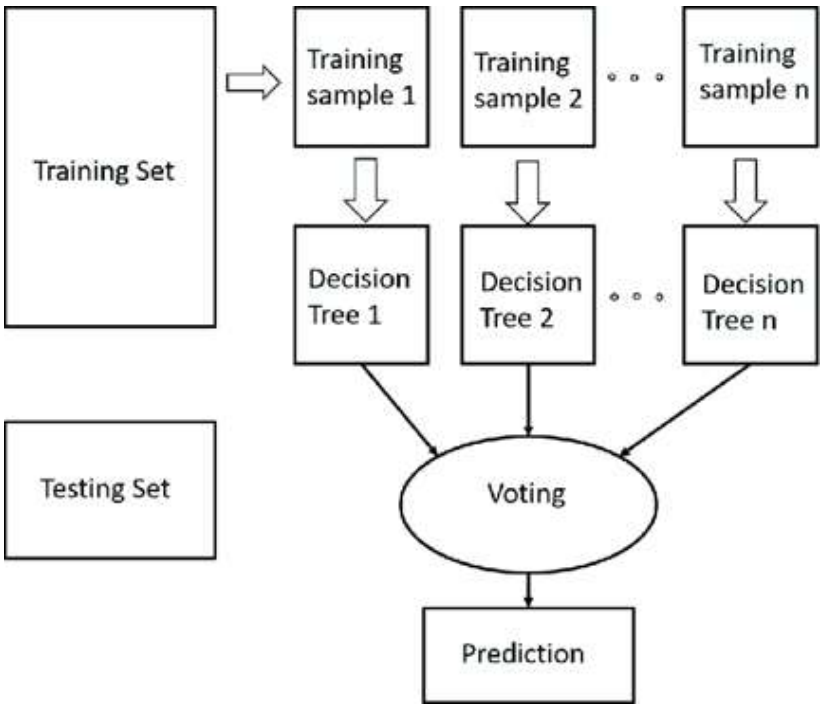


Fig.1. Architecture of Random Forest

Steps for marine accident prediction using Random Forest

1. Collect the input data.

2. Preprocess the data to prepare it for analysis.

3. Divide the data into training and testing sets.

4. Import all necessary Random Forest modules.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

5. Train the dataset using the Random Forestalgorithm.

6. Validate the model using the test data and RandomForest.

7. Use various commands to make predictions withthe trained model.

Decision Tree (DT)

A decision tree is a type of model that uses a flowchart-like structure to make predictions. It splits the datainto branches and assigns outcomes to the leaf nodes. Decision trees are used to create simple models for classification and regression tasks.

The technique operates by continuously dividing theinitial data set into subsets depending on its values for attributes till an interruption requirement is reached, such as the highest level of the hierarchy or the minimal number of examples necessary to divide a node. The decision tree method determines the appropriate attribute to divide the dataduring training based on a measure of quality such as entropyor Gini impurity, which quantifies the amount of impurity orunpredictability in the subgroups.

$$\text{Entropy (S)} = \sum_{i=1}^{c} -pi \ \log_2 pi \quad \text{----------(2)}$$

$$\text{Gain (S, A)} = \text{Entropy(S)} - \sum_{v \in Values(A)} \frac{|Sv|}{|S|} \ Entropy|Sv| \quad \text{----------(3)}$$

Eq. (2) & Eq. (3) are the entropy and information gain formulas used in the Decision tree algorithm. These two equations build the decision tree algorithm.

Where pi is the probability of randomly selecting an examplein class i,

N is the number of classes in a dataset,A is an attribute,

S is the set,

V is the element of attribute A

Sv is the subset of S with value v.

Support Vector Machine (SVM)

SVM is a big field of substance for a calculation thateverything best on little anyhow complex datasets. Support Vector Machine, or SVM, maybe secondhand for two together revert and plan, still a meaningful some moment of truth, it winds up being crude for assemblage.

The purpose of the SVM technique is to find the optimal line or decision boundary for categorizing n- dimensional space, so that we may simply enter fresh data points into the correct category in the future. A hyperplane isthe best choice for a boundary. SVM selects the extreme points/vectors that will aid in the construction of the hyperplane. The Support Vector Machine approach is used, and the exceptional circumstances are known as support vectors.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

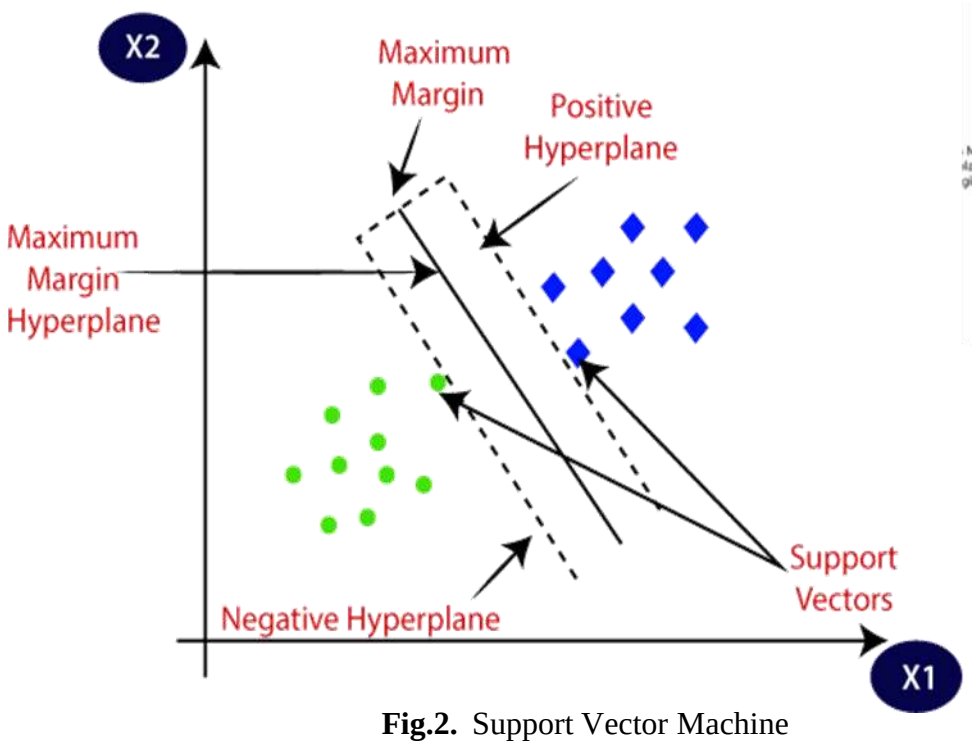


Fig.2. Support Vector Machine

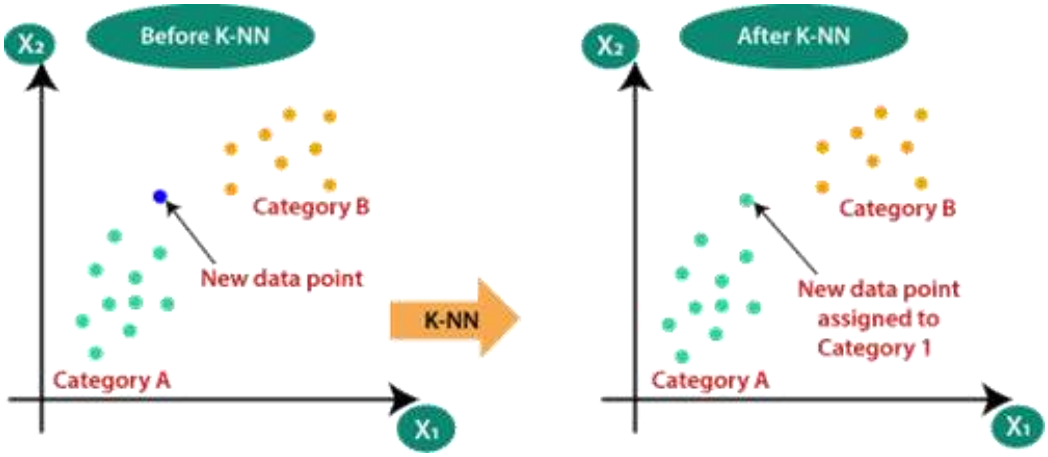
Fig.2 shows that how to make the hyperplane and select the best Boundary.

K- Nearest Neighbors (KNN)

K-Nearest Neighbors, or KNN, is a straightforwardbut effective categorization algorithm used in Machine Learning. It's a supervised learning technique used for patternrecognition, data mining, and intrusion detection. KNN is a non-parametric approach for classifying or predicting how a data point belongs to a group.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad \text{----------(4)}$$

Equation (4) is the Euclidean distance which is the distance between the two points. This method is used for theKNN algorithm. Fig.3. shows that the KNN applies before andafter applying the KNN algorithm. Two features are taken.



ISSN: 2278-4632
Vol-13, Issue-08, August 2023

Fig.3. structure of the KNN

These are the steps in the KNN algorithm

1. Select the number of neighbors, K, to consider whenclassifying a new data point.
2. Calculate the Euclidean distance between the new data point and its K nearest neighbors in the trainingset.
3. Using their Euclidean distances, determine the K closest neighbors.
4. Count the number of training data points in each category among the K nearest neighbors.
5. Place the new data point in the category with the nearest neighbors among its K.
6. The Archetype is now prepared to classify new datapoints.

IV RESULTS AND DISCUSSIONS

In this work, the Random Forest, SVM, KNN, and Decision Tree algorithms were used to predict marine accidents. The Random Forest method outperformed the other algorithms. It was used to train specific classes, and after training, the system could predict the presence of marine accidents in the classes based on the confidence value.

The experiments in this work were done using a PC with 4GB RAM, an Intel Core i5 5th generation CPU, and a Jupyter Notebook with 4GB storage.

The work is measured using the following metrics.

$$\text{Accuracy} = \frac{TP+FP}{TP+FP+FN+TN} \text{ ----------(5)}$$

$$\text{Precision} = \frac{TP}{TP+FP} \text{ ----------(6)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \text{ ----------(7)}$$

$$\text{F1 Score} = \frac{2*Precision*Recall}{Precision+Recall} \text{ ----------(8)}$$

Where,

TP = True Positive, FP = False Positive

TN = True Negative, FN = False Negative

The formulas to determine accuracy, precision, recall,and F1 Score are represented with Eq. (5), Eq. (6), Eq. (7), and Eq. (8), respectively. To calculating Accuracy,Precision, Recall and F1 Score using Confusion Matrix.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

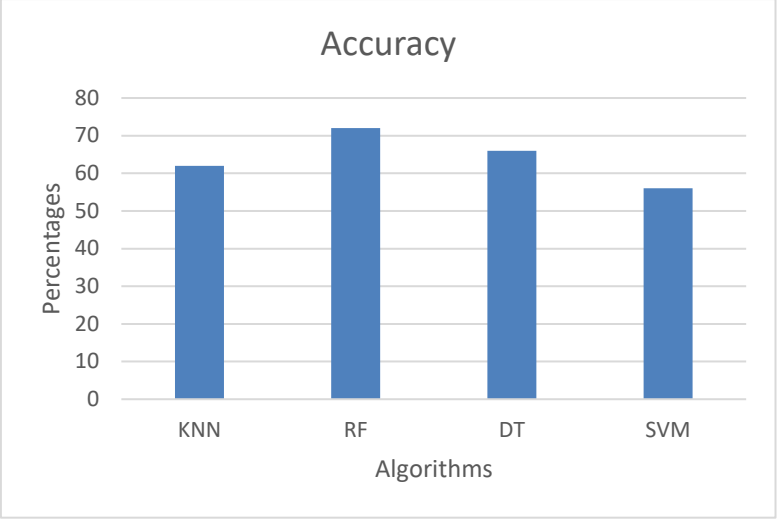


Fig.4. Accuracy for KNN, RF, DT and SVM algorithms

Fig.4 shows the accuracy level of the SVM, DT, RF and KNN. The Random Forest is containing more accuracy compared to the other algorithm.

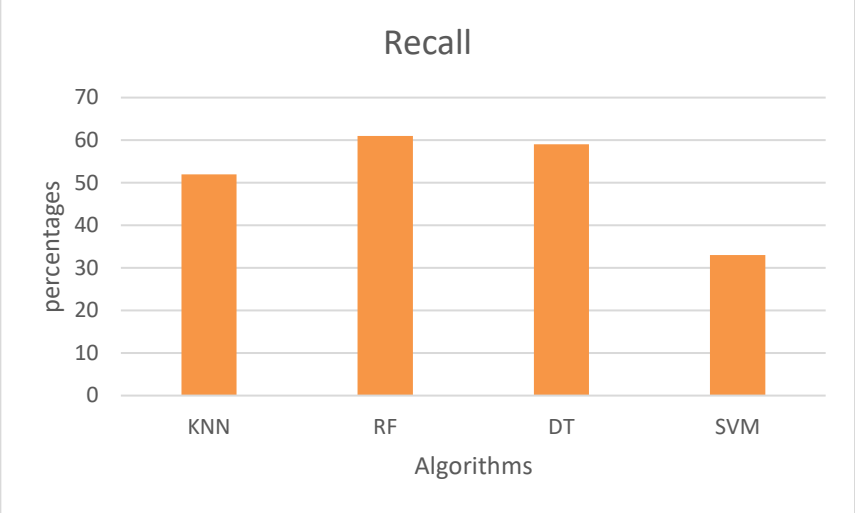


Fig.5. Recall for KNN, RF, DT, and SVM algorithms

Fig.5 shows the recall of the SVM, DT, RF and KNN. TheRandom Forest gives better performance compared to the other algorithm.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

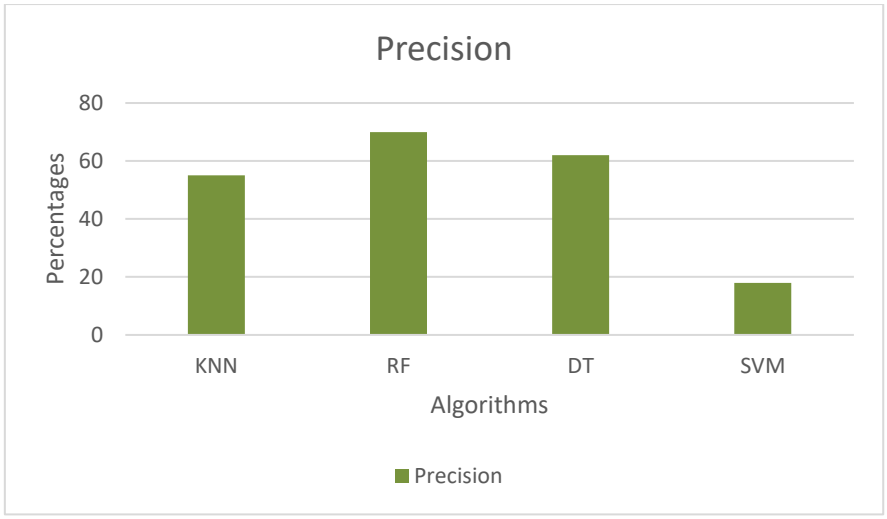


Fig.6. Precision for KNN, RF, DT, and SVM algorithms

Fig.6 shows the precision level of the SVM, DT, RF and KNN. The Random Forest is containing more precision compared to the other algorithm.

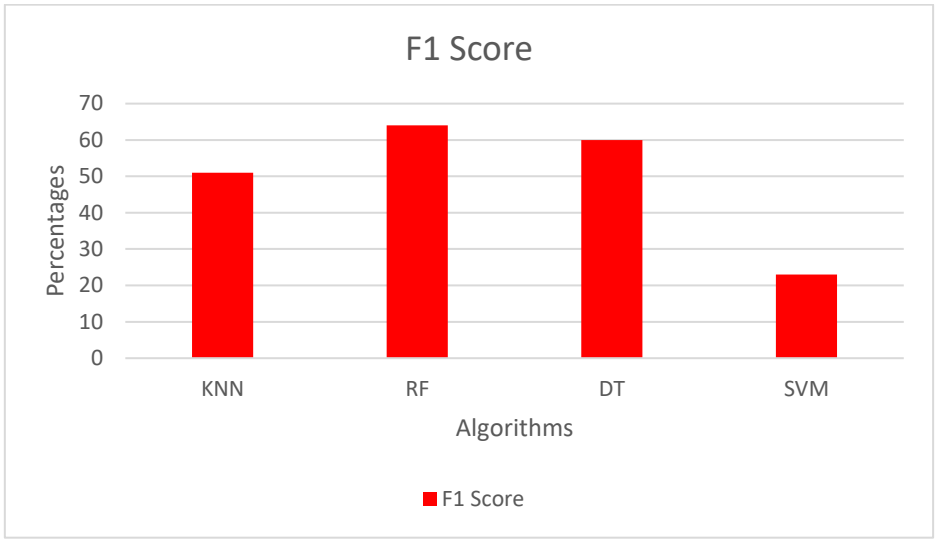
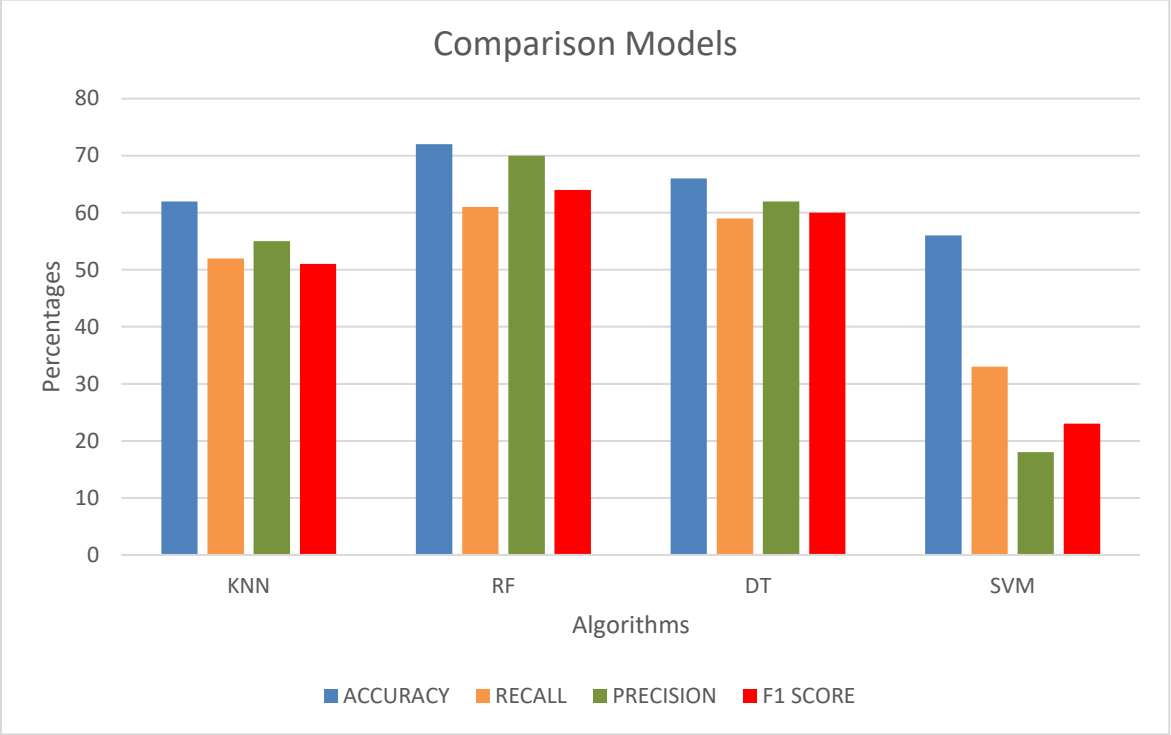


Fig.7 F1 Score for KNN, RF, DT, and SVM algorithms.

Fig.7 shows the precision level of the SVM, DT, RF and KNN. The Random Forest is containing more F1 scores compared to the other algorithm.

Fig.8 shows the comparison of the SVM, DT, KNN, and RF. The Random Forest is improving the better performance compared to other algorithms.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023



Fg.8. Comparison of Graph of KNN, RF, DT and SVM Algorithms.

To give the input data in a given parameter like rain, air temperature, water temperature, longitude, latitude etc., and process the data to produce the output.

Fig.9. Output for Marine accident Prediction

Fig.9 shows that displays the output of which accident is possible like collision, grounding, and others. The marine accident prediction can change the input data inthe given dataset and produce other types of accidents.

TABLE.1 shows that the accuracy, Precision, Recall, F1 score results comparing with the algorithms are SVM, KNN, DT and RF.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

TABLE. 1. Comparing current algorithms

Algorithms	**Accuracy %**	**Precision %**	**Recall %**	**F1score %**
SVM	56	18.6	33.3	23.9
KNN	62	55.8	52.3	51.6
DT	66	62.1	59.5	60.5
RF	**72**	**70.0**	**61.9**	**64.0**

V CONCLUSION & FUTURE WORK

This paper proposes a marine accident prediction system that uses machine learning algorithms to predict marine accidents and improve accuracy performance. Three classes - collision, grounding, and other were used in the work. Among all the machine learning algorithms used, the Random Forest algorithm improved the accuracy, recall, precision and F1 score.

Future efforts will focus on increasing the data in the dataset and adding more classes to improve the precision, recall, and F1 score.

REFERENCES

[1] F. U§urlu, S. Yıldız, M. Boran, O. U§urlu, and J. Wang, ‘‘Analysis of fishing vessel accidents with Bayesian network and chi-square methods,’’ Ocean Eng., vol. 198, Feb. 2020,Art. no. 106956, doi: 10.1016/j.oceaneng.2020.106956.

[2] W. Wagenaar and J. Groeneweg, ‘‘Accidents at sea: Multiple causes and impossible consequences,’’ Int. J. Man-Mach. Stud., vol. 27, nos. 5–6, pp. 587–598, 1987, doi: 10.1016/S0020-7373(87)80017-2.

[3] P. Antªo, O. Grande, P. Trucco, C. Soares, S. Martorell, and J. Barnett, ‘‘Analysis of maritime accident data with BBN models,’’ Saf. Rel. Risk Anal., Theory, Methods Appl., vol.2, pp. 3265–3274, Sep. 2008.

[4] A. Mullai and U. Paulsson, ‘‘A grounded theory model for analysis of marine accidents,’’ Accid. Anal. Prevention, vol. 43, no. 4, pp. 1590–1603, 2011, doi: 10.1016/j.aap.2011.03.022.

[5] A. Coraddu, L. Oneto, B. N. de Maya, and R. Kurt, ‘‘Determining the most influential

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

human factors in maritime accidents: A datadriven approach,'' Ocean Eng., vol. 211, Sep. 2020, Art. no. 107588, doi: 10.1016/j.oceaneng.2020.107588.

[6] X. Shi, H. Zhuang, and D. Xu, ''Structured survey of human factorrelated maritime accident research,'' Ocean Eng., vol. 237, Oct. 2021, Art. no. 109561, doi: 10.1016/j.oceaneng.2021.109561.

[7] J. Zhang, A. He, C. Fan, X. Yan, and C. G. Soares, ''Quantitative analysis on risk influencing factors in the Jiangsu segment of the Yangtze River,'' Risk Anal., vol. 41, no. 9, pp. 1560–1578, Sep. 2021, doi: 10.1111/risa.13662.

[8] B. Wu, J. Zhang, T. L. Yip, and C. G. Soares, ''A quantitative decisionmaking model for emergency response to oil spill from ships,'' Maritime Policy Manage., vol. 48, no. 3, pp. 299–315, Apr. 2021, doi: 10.1080/03088839.2020.1791994.

[9] J.-N. Zhao and J. Lv, ''Comparing prediction methods for maritime accidents,'' Transp. Planning Technol., vol. 39, no. 8, pp. 813–825, Nov. 2016, doi: 10.1080/03081060.2016.1231901.

[10] M. Luo and S.-H. Shin, ''Half-century research developments in maritime accidents: Future directions,'' Accident Anal. Prevention, vol. 123, pp. 448–460, Feb. 2019, doi: 10.1016/j.aap.2016.04.010