**Rock-Vs-Mine-Prediction Using Machine Learning Algorithms**.

1Anubhav Kumar

1Kumaraa8@rknec.edu

1Shri Ramdeobaba College of Engineering and Management, Nagpur

***ABSTRACT:* - There is a conflict between two countries, and one country's submarine is traveling underwater in the enemy's territory. The enemy has placed explosive mines in the ocean, and the submarine needs to differentiate between a mine and a rock to avoid danger.**

**To achieve this, the submarine uses sonar signals that emit sound waves and receive echoes to determine if an object is a mine or a rock. Various algorithms, such as decision trees ,KNN ,logistic regression, Random forest ,Naïve Bayes and support vector machines (SVM) can be used to classify the objects.**

**I] INTRODUCTION**

**UNDERWATER MINES:** Underwater mines, also known as naval mines, are explosive devices that are placed in bodies of water to destroy enemy surface ships or submarines. They have been in use since the mid-19th century and were introduced by David Bushner during the American Civil War in 1877. It is estimated that there are still around 5,000 naval mines remaining from the two World Wars in the Adriatic Sea.

Traditionally, mines were activated only by physical contact, but modern mines can be activated by various methods, such as acoustic, pressure, and magnetic changes in the water. These are known as influence mines. Mines are usually classified as either offensive or defensive weapons in warfare.

One challenge with identifying underwater mines is that they can be mistaken for rocks due to their similar shape, length, and width. Therefore, it is important to use accurate detection methods to distinguish mines from rocks. One such method is SONAR, which uses sound waves to detect and locate underwater objects.

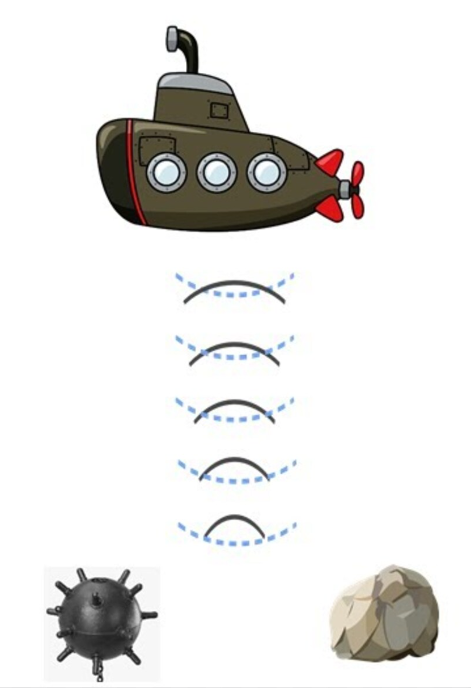
In conclusion, underwater mines are a dangerous threat to naval operations, and their detection and removal are critical for ensuring safe passage for ships and submarines. The use of advanced technology such as SONAR can help improve detection and mitigation of this threat.

Top of Form

**SONAR:** The SONAR system works by using sound waves to navigate and detect objects underwater. It is commonly used for military purposes such as acoustic mine detection, but it is also used for non-military purposes like finding fish, mapping the sea bottom, and locating sea divers.

The frequencies used for a specific underwater sonar application depend on the range required, as the sound wave attenuation increases rapidly with frequency and limits the reachable distance. In mine hunting, the frequencies of underwater SONARs typically vary between 0.1 and 1 MHz, with a range between 1 and 0.1 km.Active SONAR systems use a sound transmitter and receiver. When a sound wave is transmitted from the transmitter, it travels to the target and reflects back as an echo. The receiver then decodes the echo and records the target object's frequencies. The frequency of active SONAR is typically in the 20 kHz range.To determine if a target is a mine or a rock, we use the frequencies obtained by active SONAR at 60 different angles as our input. Passive SONAR, on the other hand, is used only to detect noises and is sometimes referred to as Listening SONAR.

In conclusion, SONAR is a versatile technology that has both military and non-military applications. Its ability to use sound waves to detect underwater objects makes it an important tool for mine hunting, fish finding, and mapping the sea bottom.



**Figure 1: Rock-Vs-Mine-Prediction**

**II] EXISTING SYSTEMS**

In the current system, various methods have been used to detect underwater mines such as explosive ordnance disposal divers, marine mammals, video cameras on mine neutralization vehicles, and laser systems. However, these methods can be risky, time-consuming, and may cause harm to marine life.

To address these issues, modern technology such as SONAR is being increasingly used as a primary tool for mine detection. SONAR technology uses sound waves to navigate and detect underwater objects, making it an efficient and accurate way to detect underwater mines without putting human divers or marine animals at risk.

Compared to traditional methods, SONAR technology enables the detection of mines at greater depths and in larger areas. It is also a non-invasive and eco-friendly way to detect mines, minimizing harm to marine life.

Continued investment in research and development of SONAR technology will further improve its accuracy and effectiveness in detecting underwater mines, making mine clearance operations safer and more efficient.

**III] PROPOSED SYSTEM**

We have proposed the development of a predictive system that provides accurate results and outcomes for underwater mine detection. To achieve this, we have utilized a dataset from "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets" by R. Paul Gorman and Terrence J. Sejnowski. In their study, they used SONAR to conduct trials in a simulated region with metal cylinders in place of mines. The object was struck with sonar signals from 60 different angles, and the results were recorded. The dataset was then used to train the evaluated models.

The Sonar output frequencies obtained from the dataset are used as input for the predictive system. We utilize classification machine learning techniques to predict whether the object is a rock or a mine. By using this dataset and machine learning techniques, we can predict the presence of a mine with a high degree of accuracy, thus improving the safety and efficiency of mine clearance operations. This approach minimizes the risk to human divers and marine life and allows for the detection of mines in a larger area and at greater depths.

**IV] WORKING PROCEDURE**

**ALGORITHM:**

To develop our predictive system for underwater mine detection, we have devised the following steps:

Step 1: We first gather the dataset and perform data preparation and exploratory data analysis to clean the dataset and ensure that it is suitable for model training.

Step 2: We then split the data into training and test datasets and use them to evaluate the classification models.

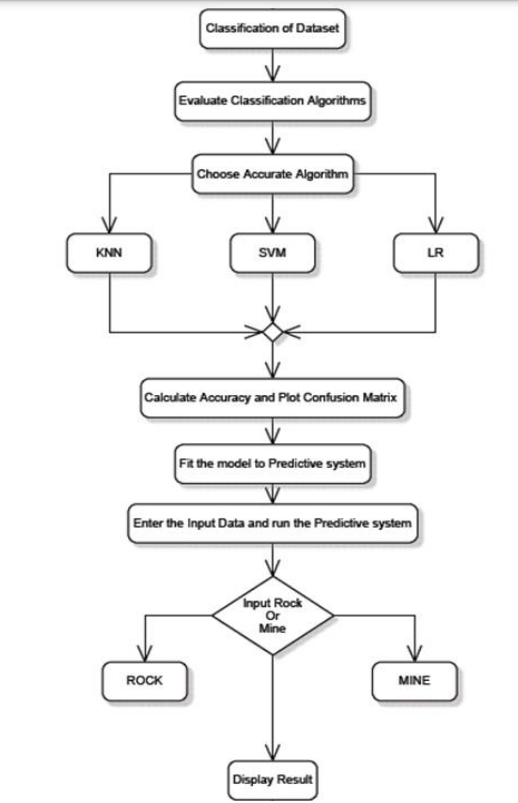
Step 3: Following the evaluation, we identify the top-performing models, which may include SVM, KNN, Logistic Regression, and others.

Step 4: We evaluate the accuracy of these models and generate a classification report to assess their performance.

Step 5: We then fit the selected models to create a prediction system that is both accurate and efficient.

Step 6: Using the predictive system, we can finally determine if the object detected by SONAR is a mine or a rock, and take appropriate action to ensure the safety of divers and marine life.

By following these steps, we can create a reliable and effective system for underwater mine detection, which can significantly improve the safety and efficiency of mine clearance operations.



**V] TOOLS AND ALGORITHMS USED IN**

**ROCK VS MINE PREDICTION**

In this paper supervised classification problem to be trained with algorithms like:

**1.Random Forest**

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to make predictions.In a random forest model, multiple decision trees are trained on randomly selected subsets of the training data and with random subsets of the features. This process helps to reduce overfitting and improve the generalization performance of the model.

During prediction, the random forest algorithm aggregates the predictions of all individual trees to make a final prediction. This aggregation can be done through either a majority voting mechanism (in classification tasks) or an averaging mechanism (in regression tasks).Random Forest has many advantages, including its ability to handle high-dimensional datasets, its robustness to noisy data, and its ability to capture complex relationships between features and target variables. Additionally, Random Forest models can provide estimates of feature importance, which can be useful in feature selection and data exploration.

**2. Logistic Regression (LR)**

Logistic Regression is a popular machine learning algorithm used for binary classification tasks. It models the probability of a binary response variable (e.g., yes/no, true/false) based on one or more predictor variables .In logistic regression, the response variable is modeled using the logistic function, which maps any value from the real number line to the range [0, 1]. The logistic function transforms the output of a linear combination of the predictor variables (called the logit) into a probability value that can be interpreted as the probability of the positive class.  
  
Logistic regression estimates the parameters of the model using a maximum likelihood estimation approach. The algorithm seeks to find the set of coefficients that maximize the likelihood of observing the training data, given the model.  
Logistic regression is a relatively simple algorithm that can be trained quickly and works well when the relationship between the predictor variables and the response variable is roughly linear. It can also handle both continuous and categorical predictor variables.

**3. Support Vector Machine (SVM).**

Support Vector Machine (SVM) is a popular machine learning algorithm used for classification and regression tasks. SVMs work by finding the hyperplane in a high-dimensional space that best separates the different classes.

In a binary classification task, SVMs try to find a hyperplane that maximizes the margin between the two classes. The margin is the distance between the hyperplane from and the closest data points from each class. The SVM algorithm then classifies new data points by which side of the hyperplane they fall on.

SVMs can also handle non-linear classification tasks by mapping the original feature space into a higher-dimensional space using a kernel function. The SVM algorithm then finds a hyperplane in this higher-dimensional space that separates the classes.

SVMs have several advantages, including their ability to handle high-dimensional data, their robustness to noisy data, and their ability to find non-linear decision boundaries. However, SVMs can be sensitive to the choice of hyperparameters, and their training time can be slow for large datasets.

**4. K-Nearest Neighbors (KNN)**

K-Nearest Neighbors (KNN) algorithm is a type of supervised machine learning algorithm used for classification and regression tasks. It is a non-parametric and instance-based algorithm that does not make any assumptions about the underlying data distribution.

In KNN algorithm, the classification of a new data point is based on the k nearest data points in the training set. The value of k is typically a positive integer and is chosen by the user. The algorithm computes the distances between the new data point and all the training data points and selects the k nearest points. The new data point is then assigned to the class that occurs most frequently among the k nearest points.

KNN algorithm is relatively simple and easy to understand. It can be used for both classification and regression tasks. However, it can be computationally expensive for large datasets since it requires storing all the training data points in memory. Additionally, the choice of k can significantly affect the algorithm's performance, and selecting the optimal value of k can be challenging.

**5. Decision Tree**

The Decision Tree algorithm is a type of supervised machine learning algorithm used for both classification and regression tasks. It creates a decision tree by recursively splitting the data into subsets based on the features that best separate the target variable.

The decision tree consists of nodes that represent the features, branches that represent the decisions or outcomes, and leaves that represent the final predictions or target variable values. The algorithm selects the feature that maximizes the information gain at each node, which is a measure of the reduction in entropy or impurity of the target variable.

The decision tree algorithm continues recursively to build the tree until it reaches the predefined stopping criteria, such as a maximum tree depth or a minimum number of samples in each leaf. Once the tree is built, it can be used to predict the target variable for new data by traversing the tree based on the values of the features.

**6. Naive Bayes**

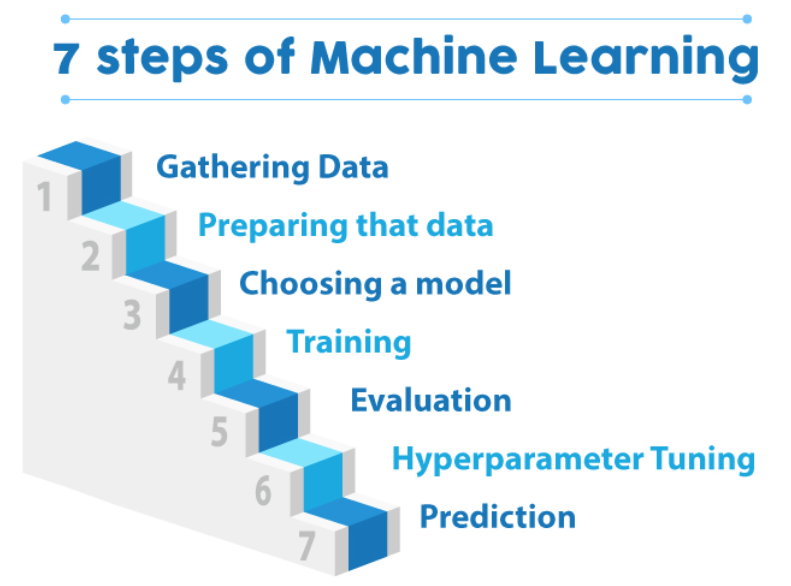
Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It is called "naive" because it assumes that the features are independent of each other, which is not always the case in real-world scenarios. The algorithm is commonly used in natural language processing, spam filtering, sentiment analysis, and other text classification tasks.

The Naive Bayes algorithm calculates the probability of a sample belonging to a particular class by multiplying the prior probability of the class and the likelihood of the sample given that class. The class with the highest probability is chosen as the predicted class.

The algorithm is trained on a labeled dataset, where each sample is labeled with its corresponding class. The prior probability of each class is calculated by dividing the number of samples in that class by the total number of samples in the dataset. The likelihood of a sample belonging to a class is calculated by multiplying the probabilities of each feature given that class.

During the prediction phase, the algorithm calculates the probability of the sample belonging to each class and chooses the class with the highest probability as the predicted class.

One of the advantages of Naive Bayes is that it requires a small amount of training data compared to other algorithms. It is also computationally efficient and can handle high-dimensional data. However, the assumption of feature independence may not always hold true in real-world scenarios, leading to inaccurate predictions.

****

**Figure 2: Methodology for Typical Machine Learning approach.**

**VI] PROPOSED FRAMEWORK**

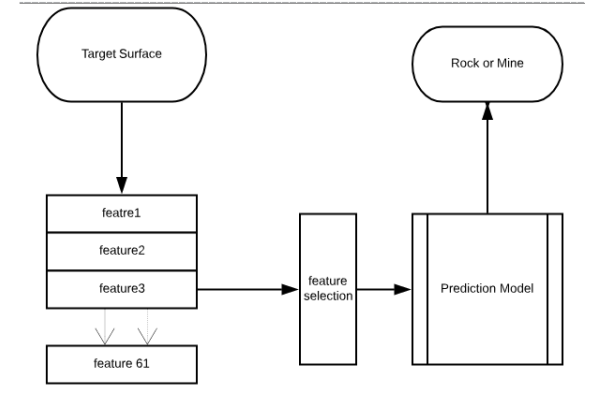
The main concern of analysis in the field of machine learning is being to form a scheduled computational machine for the categorizing the forecast of the objects, based on the attainable information. The outcome of proposed framework helps to predict the triggered sound waves reflect from surface Rock or a Mine.

**1) Proposed framework methods:** Broadly in physical world or realistic issues, there is no curb over the types of data. Some dire pre-processing like removal of missing values, feature selection, etc. are always required. Machine learning focuses on taking up contemporary techniques to process huge amount of complex data with lower expense.

The abstract view of proposed framework has been represented in Figure 3. Figure 3 describes the framework of the prediction model created to determine the surface to be a rock or a mine based on about 61 factors or features , which give outputs with an acceptable accuracy and precision percentage.

**i. Preprocessing:** Missing values are removed by replacing them by mean value imputation.

**iii. Prediction Model:** Different ML classifiers are explored and implemented to find the best possible solution. The outcome of this proposed framework helps to predict the targeted surface to be a Rock or a Mine.



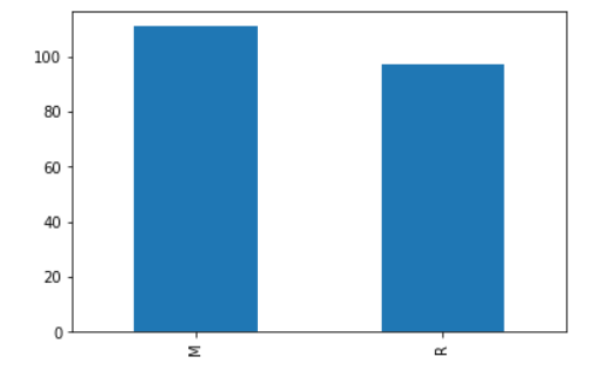
**Figure 3: Prediction Framework**

**VII] EXPERIMENTAL RESULTS AND DISCUSSION**

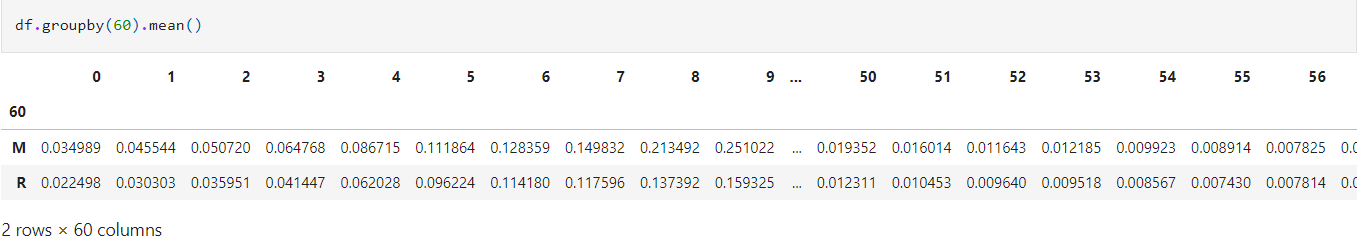
This section discusses parameter evaluation metrics to measure the performance of various machine learning algorithms.

Firstly, Lets look out the how much values have rock and mine in target table

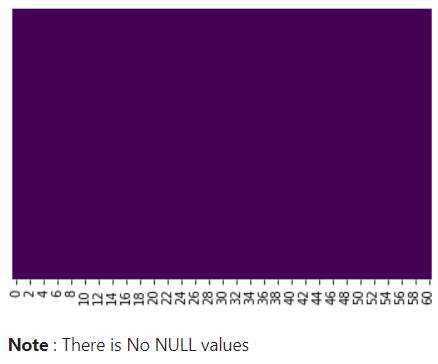
1. **Mine and Rock Representation**



‘**M**‘ represents mines, and ‘**R**‘ represents rocks. Now, let us group this data of mines and rocks through the mean function.

****

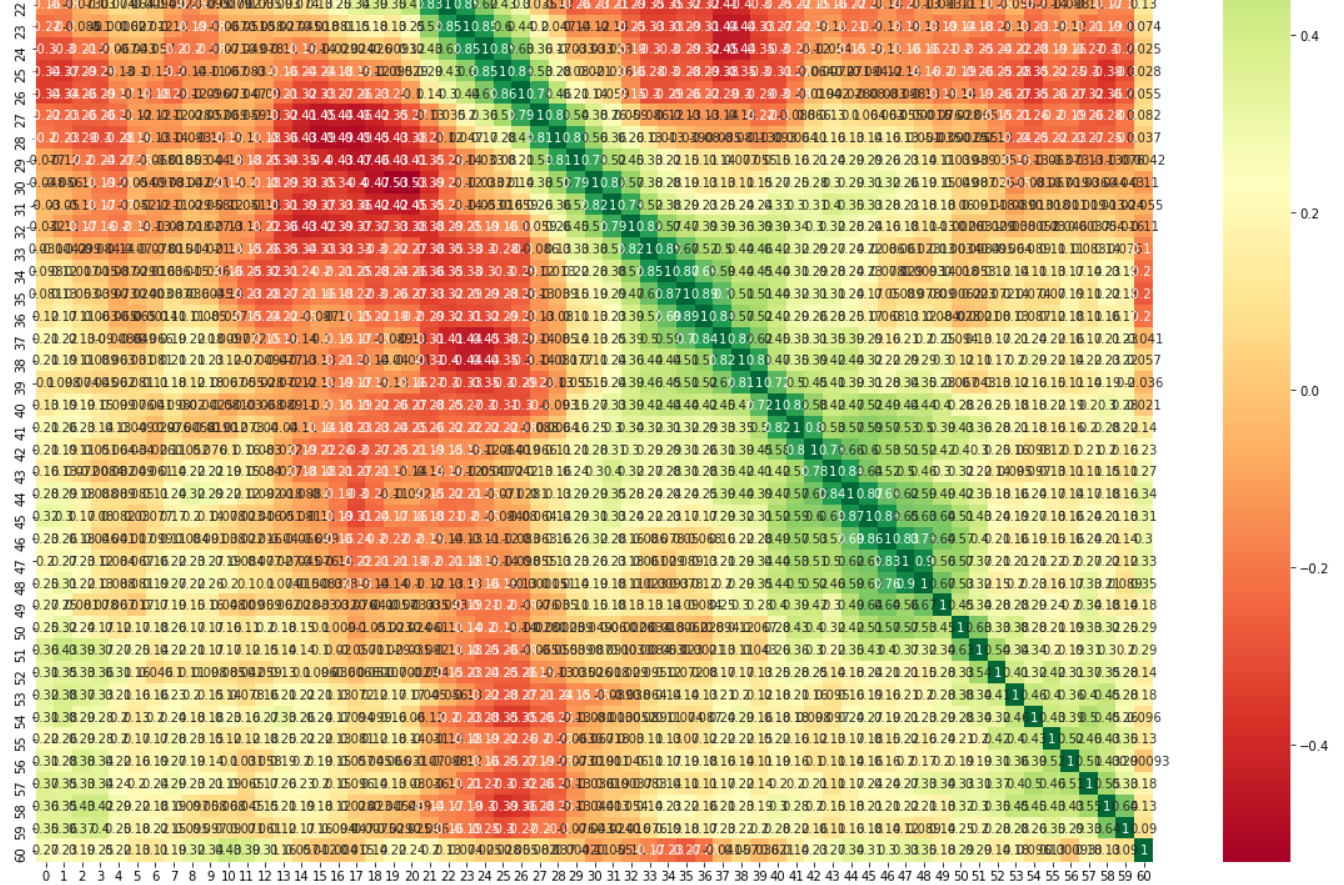
1. **Check the null values using heatmap**



1. **Correlation Matrix with Heatmap**

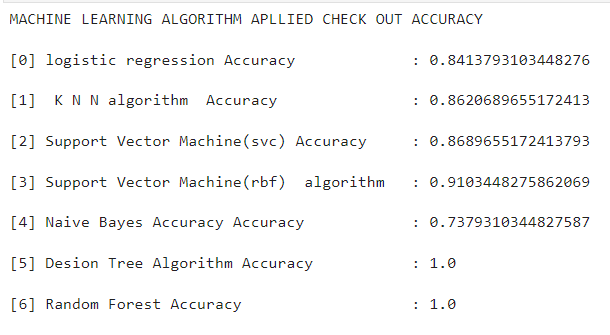
Correlation is a statistical measure that quantifies the relationship between two variables, such as the features or the target variable. A positive correlation between two variables means that an increase in one variable is associated with an increase in the other variable, while a negative correlation means that an increase in one variable is associated with a decrease in the other variable.

One way to visualize the correlation between variables is to use a heatmap, which can be created using the seaborn library. The heatmap allows us to quickly identify which features are most strongly correlated with the target variable, as well as which features are highly correlated with each other. This can be useful in understanding the underlying relationships between variables and in identifying potential areas for further analysis.Top of Form

****Bottom of Form

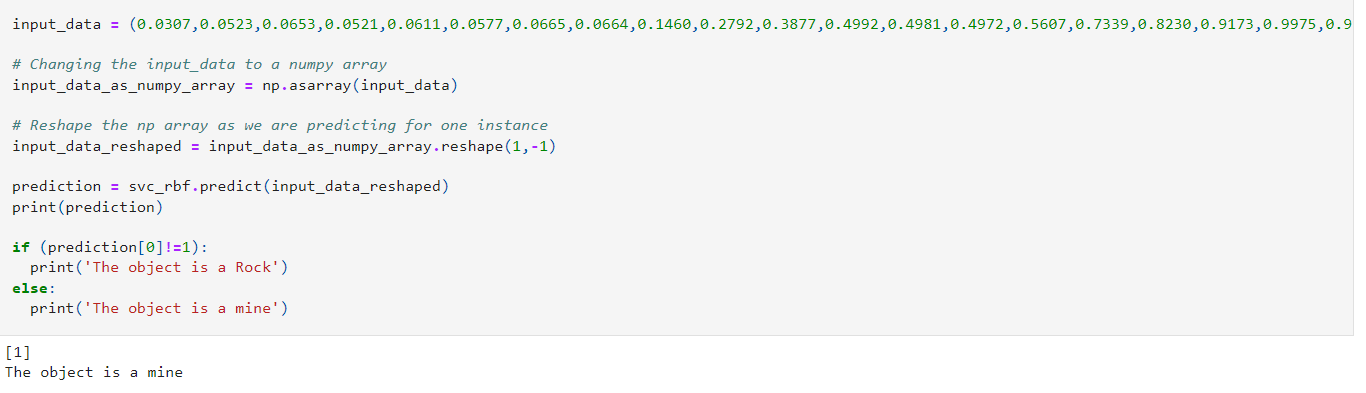
**VIII] RESULT**

In this paper, we used various machine learning algorithms, The results of the model are shown below with their classification report for better understanding of the accuracy and other scores of the models.



NOTE:- The project involved the application of six machine learning algorithms to detect rocks and mines in the ocean bed. However, it was observed that two algorithms, namely Decision Tree Algorithm and Random Forest, resulted in overfitting, and hence they were not considered for the final prediction model. On the other hand, the Support Vector Machine (RBF) algorithm yielded good accuracy (91.0%) and was selected for the prediction model.

**Result for Mine: -**

****

**Result for Rock: -**

****

**CONCLUSION**

Our project aims to detect rocks and mines in the ocean bed using machine learning algorithms. Naval mines can be used to block ships and restrict naval operations, resulting in significant negative economic and environmental impacts. Currently, there are two ways to detect mines, one using sonar signals and the other using manpower. Sonar signals are preferred as they carry less risk. The collected data is stored in a CSV file.

We use different machine learning techniques to evaluate and compare their accuracies, which helps us build a better-performing prediction model. Python is an open-source software with fast computation speed and lower costs, making it a suitable choice for our project. The goal of our project is to simplify and streamline the detection process.

**REFERENCES**

[[1]https://www.irjmets.com/uploadedfiles/paper/issue\_4\_april\_2022/21936/final/fin\_irjmets1651487003.pdf](%5b1%5dhttps:/www.irjmets.com/uploadedfiles/paper/issue_4_april_2022/21936/final/fin_irjmets1651487003.pdf)

[[2] https://www.kaggle.com/code/sugamkhetrapal/project-3-sonar-mines-vs-rocks/notebook](https://www.kaggle.com/code/sugamkhetrapal/project-3-sonar-mines-vs-rocks/notebook)

[[3] https://www.ijser.org/researchpaper/Techniques-for-classification-sonar-rocks-vs-mines.pdf](https://www.kaggle.com/code/sugamkhetrapal/project-3-sonar-mines-vs-rocks/notebook)

[4]<https://www.analyticsvidhya.com/blog/2022/04/predicting-sonar-rocks-against-mines-with-ml/>

[5]<https://www.irjmets.com/uploadedfiles/paper/issue_4_april_2022/21936/final/fin_irjmets1651487003.pdf>

[6] <https://medium.com/@harsh.vardhan7695/rock-v-s-mine-prediction-using-ml-models-233207d1207f>

[7] “Connectionist Bench (Sonar, Mines vs. Rocks).” Connectionist Bench (Sonar, Mines vs. Rocks) | Kaggle, www.kaggle.com, <https://www.kaggle.com/datasets/armanakbari/connectionist-bench-sonar-minesvs-rocks>.

[8] “Underwater Mine Detection Using Symbolic Pattern Analysis of Sidescan Sonar Images.” Underwater Mine Detection Using Symbolic Pattern Analysis of Sidescan Sonar Images, ieeexplore.ieee.org, <https://ieeexplore.ieee.org/document/5160102>.

[9] <https://www.jetir.org/papers/JETIR1907H24.pdf>

[10] Dura, Esther, et al. "Active learning for detection of mine-like objects in side-scan sonar imagery." IEEE Journal ofOceanic Engineering 30.2: 360-371 (2005).

[11]<https://www.kaggle.com/code/aryanandakumar/sonar-rock-vs-mine-prediction>

[12]<https://www.google.com/search?q=rock+vs+mine+prediction+project&tbm=isch&ved=2ahUKEwjjmKSwj5n9AhXk63MBHVdODQUQ2-cCegQIABAA&oq=rock+vs+mine+prediction&gs_lcp=CgNpbWcQARgBMgQIIxAnMgcIABCABBAYMgcIABCABBAYUABYAGDlEmgAcAB4AIABjgGIAY4BkgEDMC4xmAEAqgELZ3dzLXdpei1pbWfAAQE&sclient=img&ei=uaXtY6PYO-TXz7sP15y1KA&bih=656&biw=1536#imgrc=ax3QTFV9S9ak-M>

[]Data Set Link:-

<https://drive.google.com/file/d/1pQxtljlNVh0DHYg-Ye7dtpDTlFceHVfa/view>