

Cranes Varsity

‘***Where Technology meets Excellence’***

**Bangalore, India**

**Project on:**

Prediction of Underwater Surface Target through SONAR: A Case Study and project report of Machine Learning.

**Submitted by:** (Batch No:1)

Arun Prasad P

**Under the guidance of:**

Mr. Jithesh Kurian (Trainer)

---------------------------------------------------------------------------ACKNOWLEDGEMENT

---------------------------------------------------------------------------------

At this pleasing moment of having successfully completed our project, we wish to convey our sincere thanks and gratitude to the Trainer of the Cranes Varsity, Mr. JITHESH KURIAN who provided us the logics and implementation of the algorithms in the project.

Also, I express indebt thanks to our TEACHING and NON-TEACHING staffs of CRANES VARSITY, Bangalore.

**---------------------------------**

**TABLE OF CONTENTS:**

**---------------------------------**

|  |  |  |
| --- | --- | --- |
| **Sno** | **TITLE** | **Page no** |
| 1. | Abstract | **4** |
| 2. | Introduction | **5** |
| 3. | Material and the Methods | **6** |
| 4. | Procedure | **8** |
| 5. | Experimental Setup on G-colab Notebook | **10** |
| 6. | Conclusion | **15** |
| 7. | References | **16** |

--------------------

**ABSTRACT:**

--------------------

The discovery of rocks and minerals would have been very difficult past the development of the SONAR technique, which relays on certain parameters to be able to detect the obstacle or the surface is a rock or a mine. Machine learning has drawn the attention of maximum part of the technology related and based industries, by showing advancements in the predictive analytics.

The main aim is to show the model prediction representation, united by the machine learning algorithmic characteristics, which can figure out if the target of the sound wave is either a rock or a mine or any other organism or any kind of other body. This attempt is a clear-cut case study which comes up with a machine learning plan for the grading of rocks and minerals, executed on a huge, highly spatial and complex SONAR dataset.

The attempts are done on highly spatial SONAR dataset and achieved an accuracy of 90.4%(approx.). With Support Vector Machine algorithm, the results are further optimized by feature selection to get the maximum accuracy of 90.4%. Assuring results are found, when the fulfillment of the designed groundwork is set side by side with the standard classifiers like SVM, random forest, Decision tree and etc. using different evaluation metrics like accuracy, sensitivity, etc.

Machine learning is performing a major role in improving the quality of detection of underwater natural resources, and will tend be better in the near future.

-----------------------------

INTRODUCTION:

-----------------------------

* There is a lot to explore under the deep waters (seas and oceans), rocks and mines are two of those crucial natural resources, and this would have been very difficult to find these resources past the development of the SONAR technique, which is an acronym for Sound Navigation and Ranging, and is used to measure the depth of the sea or the ocean or the distances in the water.
* In this probe, after the pre- processing of the input, different machine learning classifiers are trained to check the achievement of classification. The conduct for the finest classifier included comparison with some standard up-to-date classifiers like Random Forest, SVM, Decision Tree etc.
* Advantageous results are achieved, when we compare the performance of the classifiers in the framework like standard classifiers like SVM, random forest, K-neural networks, etc., using various evaluating metrics like accuracy, area under curve, sensitivity, specificity etc. Waves can be used to make predictions for the underwater surfaces, mines and rocks.

# **MATERIAL AND METHODS**

The material and methods used for proposing the prediction model is been discussed below.

## **Dataset:**

The dataset has been collected from **Kaggle**. It has 208 x 61 features which define and differentiate Rocks and Mines and comprises of 208 such samples.

## **Machine Learning Classifiers**

**Random Forest:** Random Forest comes under the category of tree type classifiers, in this the dataset values are inspected separately and by the same distribution of all the trees in the forest. Internal valuation monitors strength, errors and the correlations which are implemented to display the response to the growing number of features that have been used in splitting

**K-Nearest Neighbour:**

An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If **k = 1**, then the object is simply assigned to the class of that single nearest neighbour.

KNN algorithm is one of **the simplest classification algorithm**. Even with such simplicity, it can give highly competitive results. KNN algorithm can also be used for regression problems.

**Support Vector Machine:**

(SVM) networks, are supervised learning algorithms that figure out the data used for classification and backsliding analysis. SVM model is a depiction of the examples as points in space, charted to create separate categories, divided by a clear chasm. New samples are then mapped into that same space and then concluded to belong to a category based on the side of the chasm they fall [8].

**Logistic Regression**:

Bayesian networks are aimed acyclic graphs whose nodes show variables in the Bayesian sense. Each node is correlated with a probability function that takes a specific set of values, as input, for the node's parent variables, and gives the probability distribution of the variable represented by the node.

**Decision Tree:**

Some of the decision tree algorithms include Hunt's Algorithm, ID3, CD4. 5, and CART. ... In this example, the class label is the attribute i.e., “loan decision”. The model built from this training data is represented in the form of decision rules.

As the goal of a decision tree is that **it makes the optimal choice at the end of each node it needs an algorithm that is capable of doing just that.**

--------------------

PROCEDURE:

--------------------

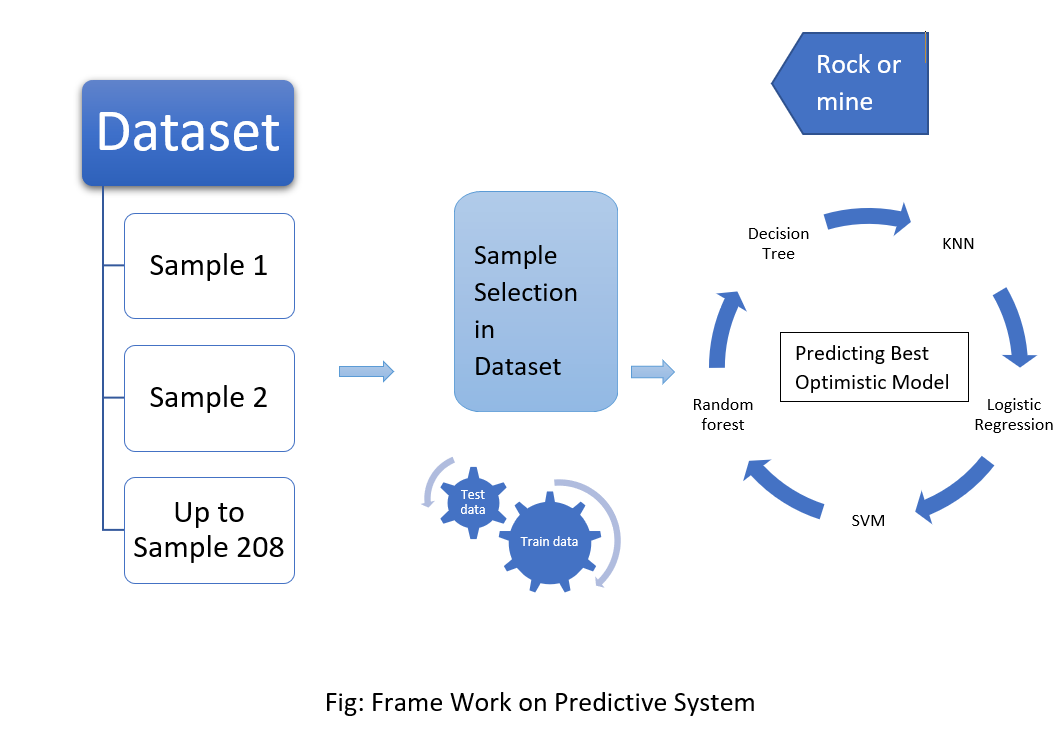
The main concern of analysis in the field of machine learning is being to form a scheduled computational machine for the categorizing the forecast of the objects, based on the attained information. The outcome of proposed framework helps to predict the triggered sound waves reflect back from what surface: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

## **Steps of Proposed framework:**

Broadly in physical world or realistic issues, there is no curb over the types of data. Some direct 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. The Figure describes the framework of the prediction model created to determine the surface to be a rock or a mine based on about 208 x 61 factors or features, processed by classifier models, which give outputs with an acceptable accuracy and precision percentage.

1. ***Pre-processing*:** Missing values are removed by replacing them by mean value imputation.
2. ***Prediction Model*:** Different ML classifiers are explored and implemented to find the best possible solution. Logistic Regression, being an ensemble model has shown the highest performance with 77% of accuracy. The results are further optimized by applying feature selection technique to feed the prediction model with the best features and accuracy reached at 81% after optimization. The outcome of this proposed framework helps to predict the targeted surface to be a Rock or a Mine.

**Frame-Work:**



Rock or mine

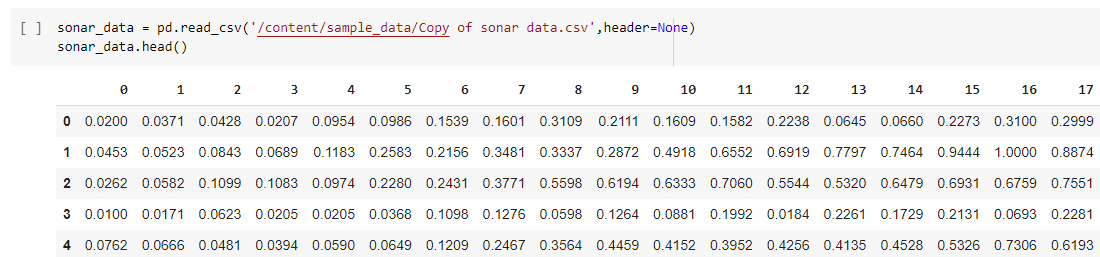
**Experimental-Setup on google Colab Notebook:**

The Colab link:

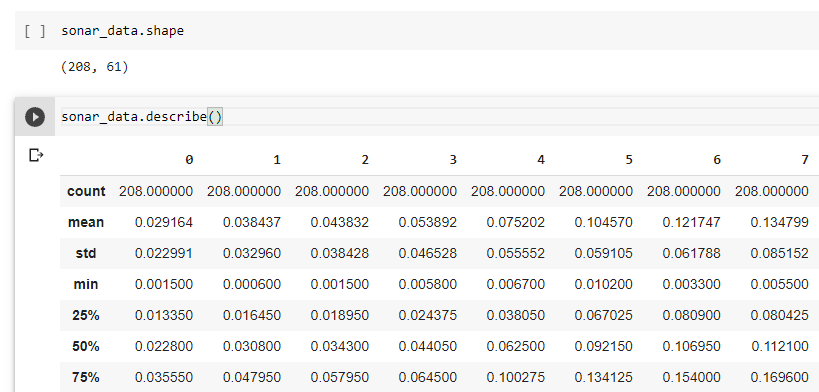
<https://colab.research.google.com/drive/1unlcIyjs6OX1DJaWK5CQTfbA3FsUagGU?usp=sharing&hl=en_GB#scrollTo=PAZ7MVZOCX44>

Experimental Code:

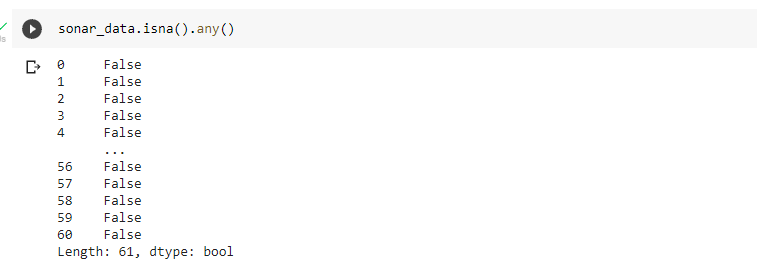
* First, we should import the necessary dependencies required for our model.
* Then load our desired dataset to a pandas Data Frame.



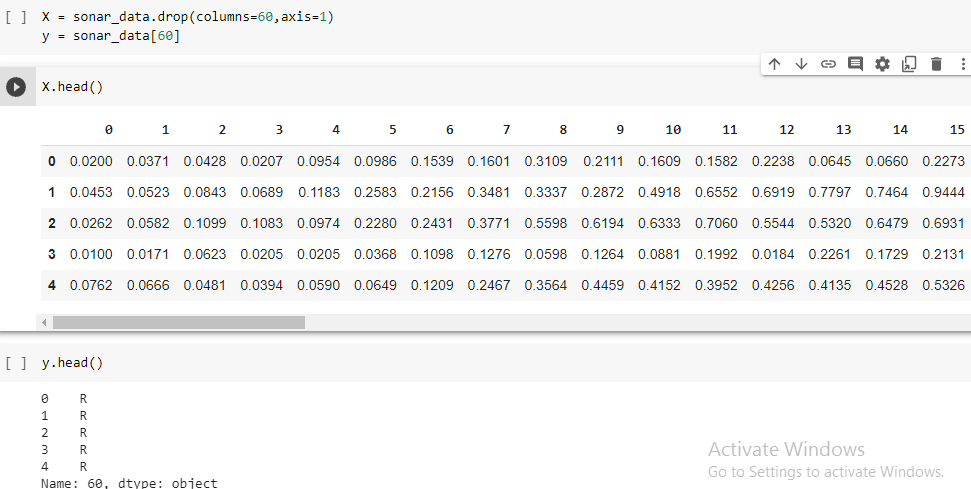
* In order to know the number of columns, rows and other statistical details of our data we can use functions like shape and describe.



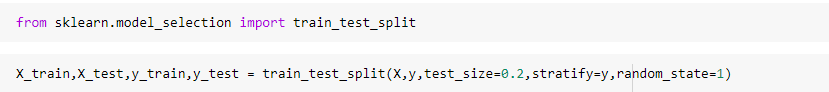
* Data cleaning is the process of detecting and correcting inaccurate records in a given dataset by replacing, modifying or deleting the irrelevant data.



* As our model is a supervised learning model, we are separating the data and labels by putting the data into **X** variable and labels into **y** variable.



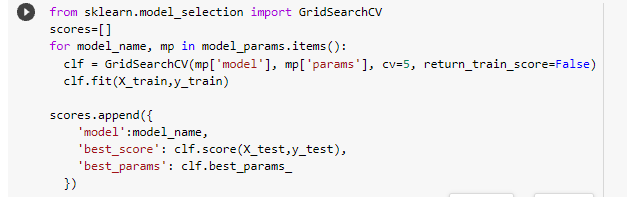
* After assigning our data to desired variables we are training and testing our data with a minimum test size.



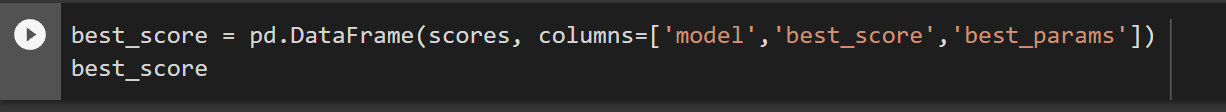
* By completing the training and testing part we proceed forward to implement Logistic Regression, K-Nearest Neighbors, Random Forest Classifier and Decision Tree classifiers for choosing the best model.



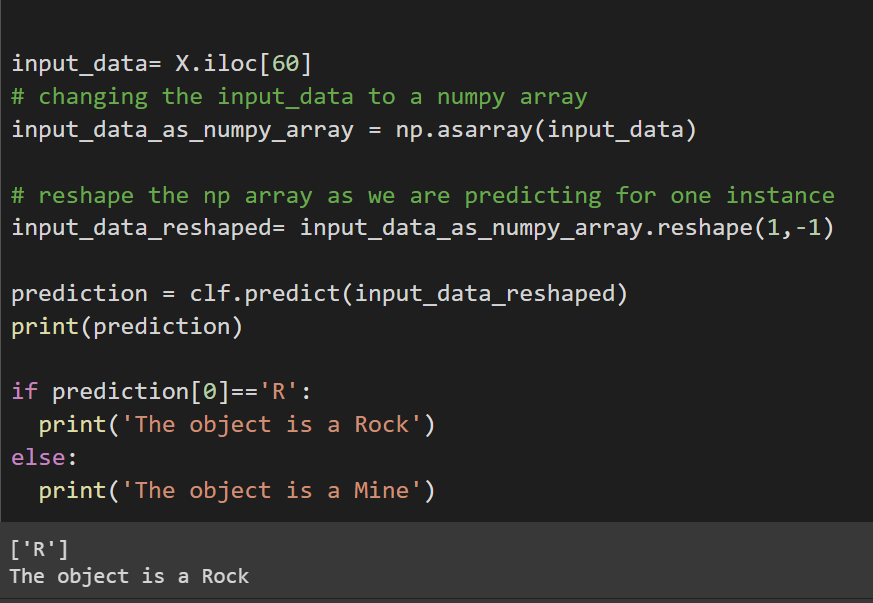
* After importing the necessary algorithms, we continue with Grid Search CV and append the scores of the model by which we can determine the best algorithm.



* Finding out the best and Optimistic value among those Classifiers.



* Finally, the object is predicted successfully, by giving the test data as input data by the concept of list slicing.



* Predicted the Output is Rock as per the given test input data for Prediction.

**Tabular and Pict Representation:**

From the above output,

|  |  |  |
| --- | --- | --- |
| **Sno** | **Algorithm** | **Accuracy in (%)** |
| 1 | Random Forest Classifier | 76.1 |
| 2 | K- Nearest Neighbour | 69.0 |
| 3 | Decision Tree | 73.8 |
| 4 | Logistic Regression | 69.0 |
| 5 | SVM | 90.4 (Max) |

--------------------------

CONCLUSION:

---------------------------

* An adequate prediction miniature, united with the machine learning classifying features, is proposed which can conclude if the target of the sound wave is either a rock or a mine or any other organism or any kind of other body. Research is carried out for predicting the best possible result for the target to be a rock or a mine, which is found to be best through the Support Vector Machine (SVM) model, which is an ensemble tree-based classifier in machine learning with the highest accuracy rate of 90.4%, with least error for better elaboration of this prediction model.
* For future work more, complex data will be handled using big data framework.
* It is found to be further Suffix best through the Random Forest algorithm, the results are only optimized by feature selection to get the accuracy of 76%.

**References:**

* codebasics – YouTube Channel.

Link: <https://www.youtube.com/c/codebasics/playlists>

Website: <https://www.skillbasics.com/>

* Siddhardhan-YouTube Channel.

Link: <https://www.youtube.com/channel/UCG04dVOTmbRYPY1wvshBVDQ>

* Learn with whiteboard-YouTube channel

Link:

<https://www.youtube.com/channel/UCF-uUxo43IPpnEwnR62WGlA>

* Kaggle Sonar-Dataset

<https://drive.google.com/file/d/1mHFPUIsaS9j8uchMTeSBZJZllSm_mf5i/view?usp=sharing>

* SONAR DATASET

[Link: https://drive.google.com/file/d/1mHFPUIsaS9j8uchMTeSBZJZllSm\_mf5i/view?usp=sharing](https://drive.google.com/file/d/1mHFPUIsaS9j8uchMTeSBZJZllSm_mf5i/view?usp=sharing)

* SUPERVISED MACHINE-LEARNING

[Link:](https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590)

<https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590>

* GRID SEARCH CV

[Link:](https://towardsdatascience.com/gridsearchcv-for-beginners-db48a90114ee)

<https://towardsdatascience.com/gridsearchcv-for-beginners-db48a90114ee>