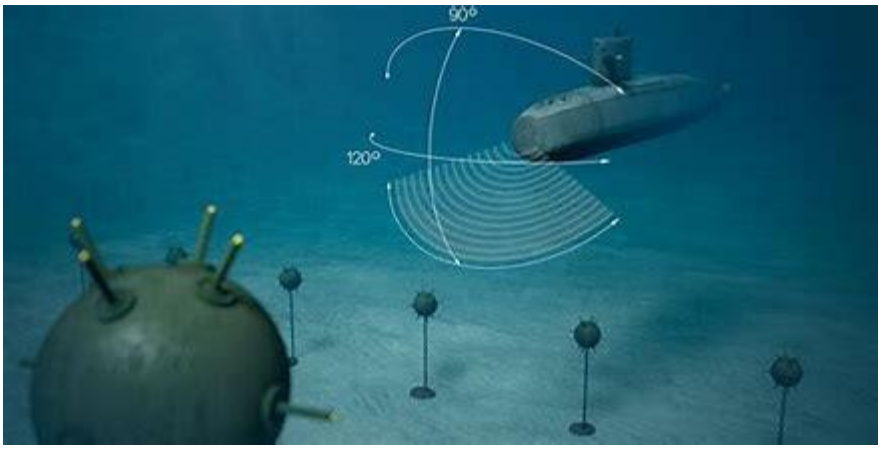


Prediction of Underwater Surface Target through SONAR by using the Machine Learning Algorithms.

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 emanate a capable prediction representative, 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 91.03%.

Assuring results are found, when the fulfillment of the designed groundwork is setside by side with the standard classifiers like SVM, random forest, 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 to 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 the similar way these sound 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, Naïve Bayes Logistic Regression, KNN** etc.

Advantageous results are achieved, when we compare the performance of the classifiers in the framework like standard classifiers like SVM, random forest, Naïve Bayes, Logistic Regression 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. This paper further consists of the following sections: Section 2: Briefly describes the classification methods which have been implemented in the desired plan. Section 3 : Confers about the data, its features and the experimental setup. Section 4: Discusses the experimental outcomes and their accomplishment comparisons. At the end, Section 5: Summarizes the paper with a conclusion and the future scope of this prediction model.

Material And Methods:

The material and methods used for proposing the prediction model is discussed in this Section.

Dataset:

The dataset has been collected from UCI Repository. It has 61 features which define and differentiate Rocks and Mines and comprises of 209 such samples.

Experimental Setting:

The WEKA tool has been used for implementing the varied feature selection and model building systems. The main motive is to calculate the predicting efficiency of the classifier when it is functional and operating and then classifying new samples outside the benefit of perceiving the bona fide class of the samples. The comparators have been designed to implement a 10-fold cross validation trial. The dataset is split into 10 equally distributed subsets. The most exact machine learning classifier is chosen as a base classifier to instruct the nine-subset layer and examine it on the last subset layer. To measure the durability of crafted groundwork, the step is repeated. To appraise the performance of the considered framework, seven different specifications listed as, F measure, accuracy, MCC, error rate, True and False Positive rates, and area under curve (AUC) are used.

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.

- **Support Vector Machine (SVM):** These networks are super visional 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.

- **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:** It is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

- **The Naive Bayes:** It separates data into different classes according to the Bayes' Theorem, along with the assumption that all the predictors are independent of one another. It assumes that a particular feature in a class is not related to the presence of other features.

- **The k-nearest neighbors algorithm:** It's also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

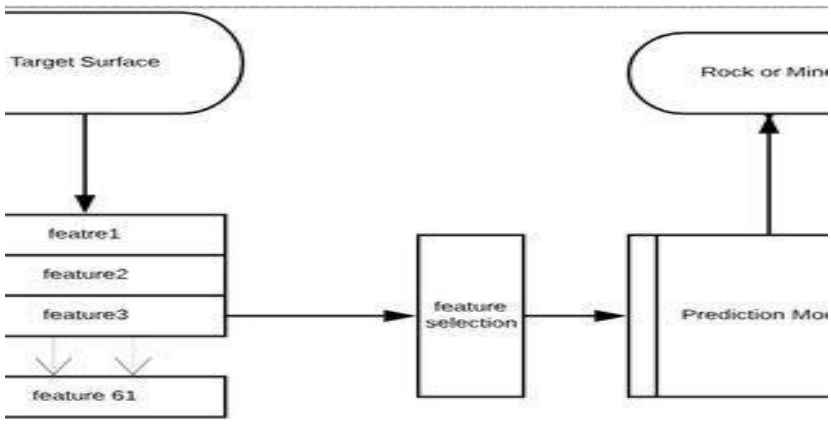
Proposed Framework:

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 back from what surface: Rock or a Mine.**

Steps for Proposed Framework:

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 below. Figure 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, processed by 10 different classifier models, which give outputs with an acceptable accuracy and precision percentage.

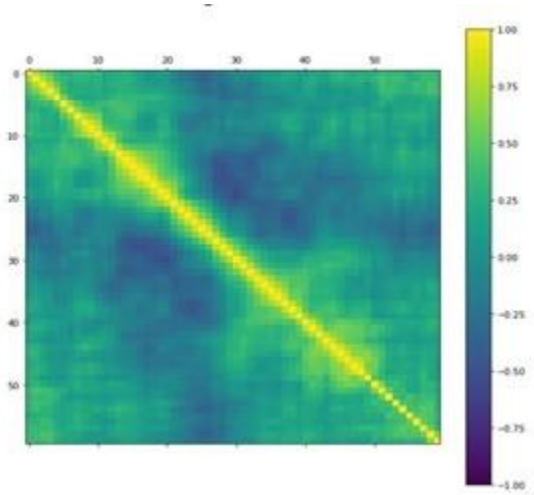
1. **Preprocessing:** Missing values are removed by replacing them by mean value imputation.
2. **Feature Selection:** Mean Gini index is used to rank the important features. The top 50 features ranked by mean gini index is selected and fed to the prediction model.
3. **Prediction Model:** Different ML classifiers are explored and implemented to find the best possible solution which is the SVM(rbf), being an ensemble model has shown the highest performance with 91.03% of accuracy after doing the performance tuning or hyperparameter tuning selection technique to feed the prediction model with the best features and accuracy.The outcome of this proposed framework helps to predict the targeted surface to be a Rock or a Mine.



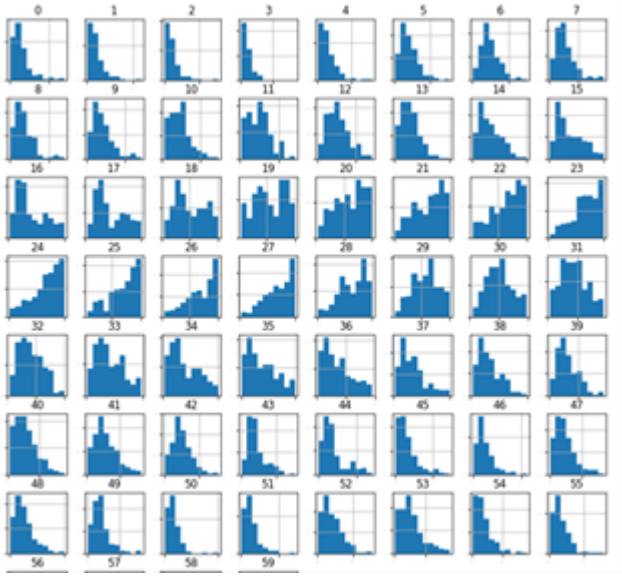
Results and Discussions:

This section discusses parameter evaluation metrics to measure the performance of various machine learning algorithms. The results of 10-fold cross validation method are presented graphically and discussed much in detail.

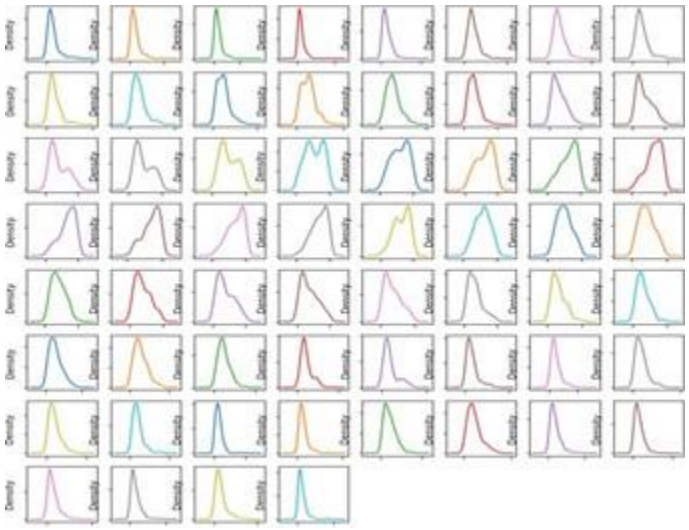
Data Correlation Representation:



Different dimensions of frequency in Vertical axis and Horizontal Axis:



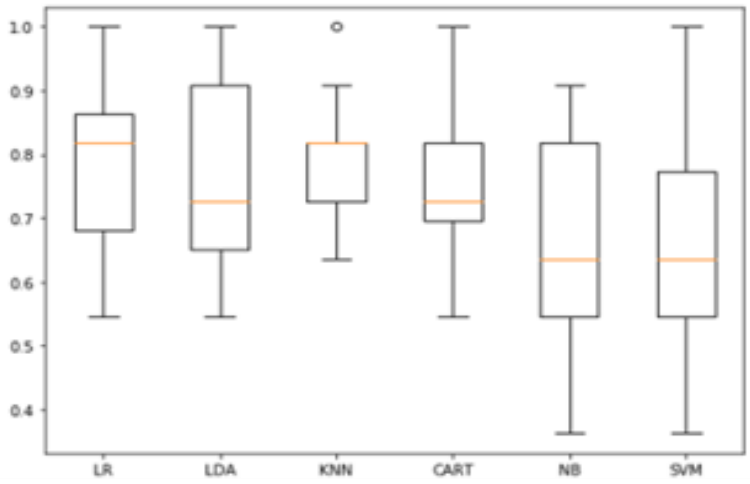
Data Distribution process in Density plots Representation:



Density Graphical representation

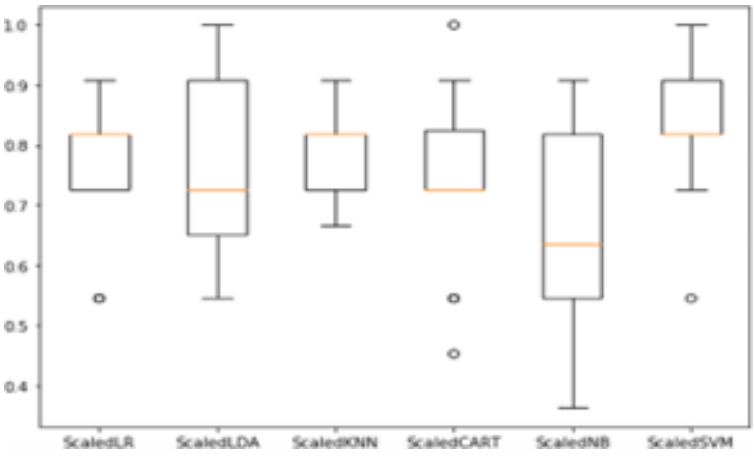
Algorithms Comparison:

In this module, we compare Numerical Data based on Quartile Values with different algorithms.

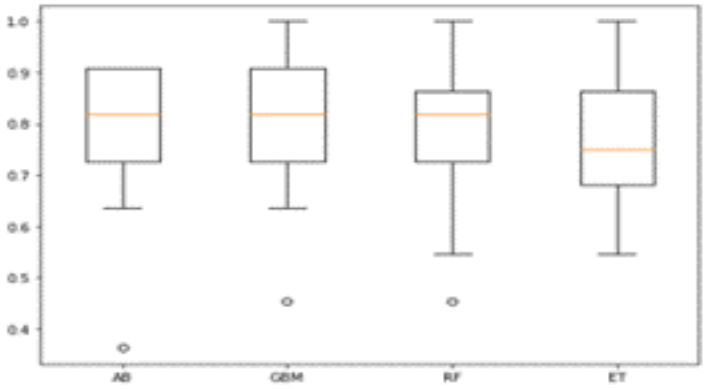


LR: Logistic Regression.
LDA: Linear discriminant analysis.
KNN: k-nearest neighbors
CART: Classification and Regression Trees.
NB: Naive Bayes classifier.
SVM: Support Vector Machine.

Scaled Algorithm Comparison:



Ensemble Algorithm Comparison:



AB: Adabosting
GBT: Gradient bosting Machine.
RF: Random Forest.
ET: Extra Tree Classifier.

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.
Out of all these chosen models, SVM(rbf) turned out to be the efficient one in terms of performance and accuracy. Apart performing the hyperparameter tuning/performance tuning, SVM(rbf) turns out to be 91.03%.
For future work more, complex data will be handled using big data Hadoop framework.