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

T.Nikitha, B.Hadhvika, Dr.K.Vaidehi

Student, Dept. of CSE, Stanley College of Engineering and Technology for Women, Hyderabad,

Student, Dept. of CSE, Stanley College of Engineering and Technology for Women, Hyderabad.

Associate Prof, Dept. of CSE, Stanley College of Engineering and Technology for Women, Hyderabad.

Abstract: Outwardly SONAR technique has exploited the discovery of rocks and minerals which would have been very difficult otherwise. The technique exploits certain parameters which will aid to detect the surface targets or obstacle such as a rock or a mine. Machine learning has drawn the attention of maximum part of today's emerging technology, related from banking to many consumer and product based industries, by showing the advancements in the predictive analytics. The main aim is to emanate a capable prediction representative method united by the machine learning algorithmic characteristics, which can deduce if the target of the sound wave is a rock, a mine, any other organism or any kind of foreign body. This proposed work 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 83.17% and area under curve (AUC) came out to be 0.92. With random forest algorithm, the results are further optimized by feature selection to get the accuracy of 90%. Persuade results are found when the fulfillment of the designed groundwork is set side by side with the standard classifiers like SVM, random forest, etc. Different evaluation metrics like accuracy, sensitivity, etc are investigated. Machine learning is performing a major role in improving the quality of detection of underwater natural resources and will tend be better paradigm.

Keywords: feature selection, data analytics, rocks and mines, machine learning, prediction, SONAR.

I. Introduction : There is a lot to explore under the deep waters, 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 [2]. 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, C4.5, Adabag etc. Advantageous results are achieved, when we compare the performance of the classifiers in the framework like standard classifiers like SVM, random forest, adabag, 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 [3]. Researchers are utilizing the results of machine learning for building the prediction models in different domains [4]. 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, C4.5, Adabag etc. Advantageous results are achieved, when we compare the performance of the classifiers in the framework like standard classifiers like SVM, random forest, adabag, neural networks, etc., using various evaluating metrics like accuracy, area under curve, sensitivity, specificity etc. 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.

II. Processing Methods: The process and methods used for proposing the prediction model is discussed in this section.

1) Dataset: In our proposed method dataset has been collected from UCI Repository. It has come across 61 features which define and differentiate Rocks and Mines and comprises of 209 samples.

2) Experimental Setting: In building of this system we are using WEKA tool for the purpose of implementing the varied feature selection and model systems. The main motive is to measure 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.

3). Machine Learning Classifiers:

In below Section explains Machine Learning classifiers with different algorithms.

a). Neural Network: An unreal neural network linked group of nodes, known as perceptron's, and is like a colossal network of neurons in a human brain. In this, the perceptron algorithm has been used to train the machine. It is for a managed learning of two-fold classifiers that can decide if an input belongs to some unique category or not [7].

b). Support Vector Machine (SVM) 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 [8].

c). 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 [6].

d). Adaboost is capable of being used in a partnership with many other types of classifiers to boost performance. Adaboost is often said to be best the out of-the-box classifier. Information collected at each stage of the Adaboost algorithm of each training sample is stuffed into the tree viable algorithm in such a way that later trees favor to focus on harder to classify examples [9].

e). 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 [10].

III. 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 1. Figure 1 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.

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

ii. 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.

iii. Prediction Model: Different ML classifiers are explored and implemented to find the best possible solution. Random forest, being an ensemble model has shown the highest performance with 83.17% 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 90.20% after optimization. The outcome of this proposed framework helps to predict the targeted surface to be a Rock or a Mine.

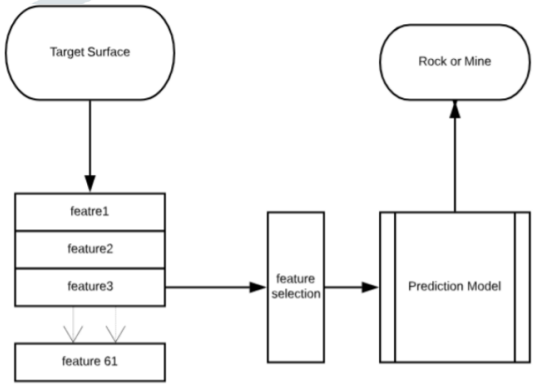
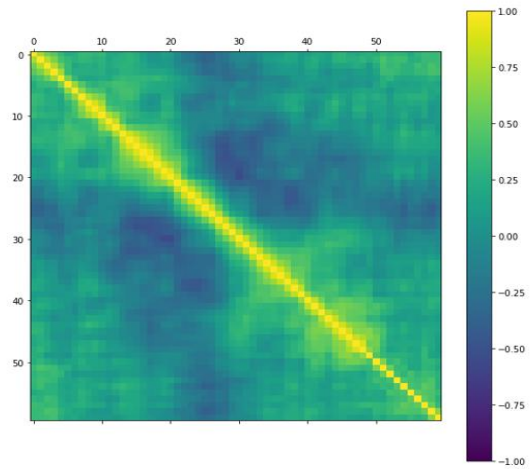


Figure 1: Prediction Framework

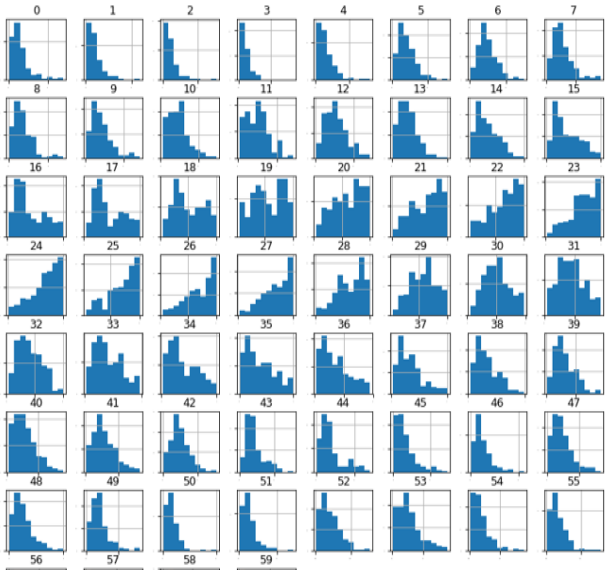
IV. Experimental Results and Discussion: This section discusses parameter evaluation metrics to measure the performance of various machine

learning algorithms. The results of 10-fold cross validation method are presented graphically and discussed much in detail.

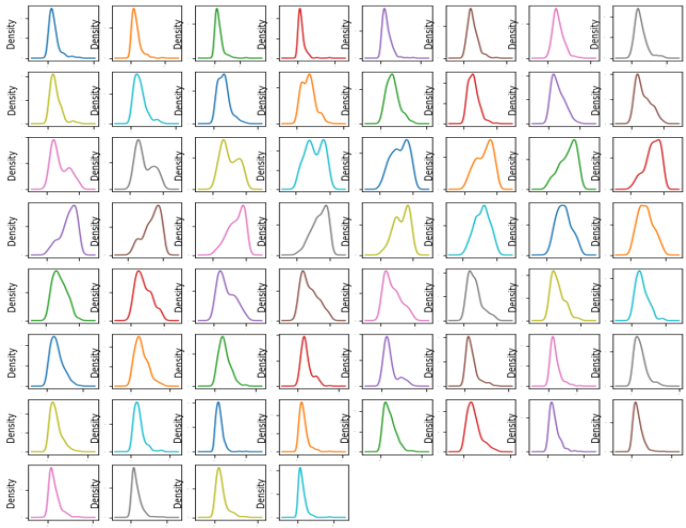
a) Data Correlation Representation



b) Different dimensions of frequency in Vertical axis and Horizontal Axis



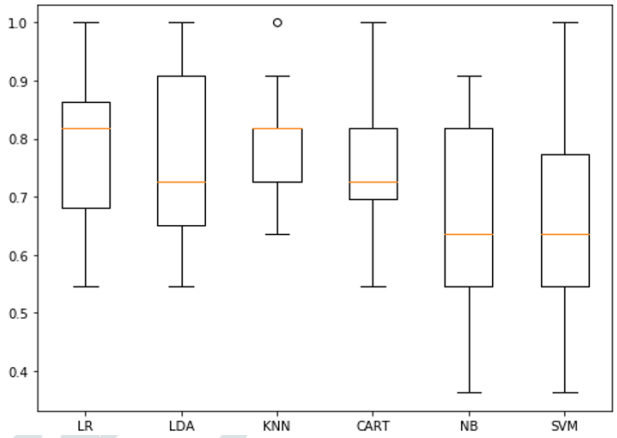
c) Data Distribution process in Density plots Representation:



Density Graphical representation

d) 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.

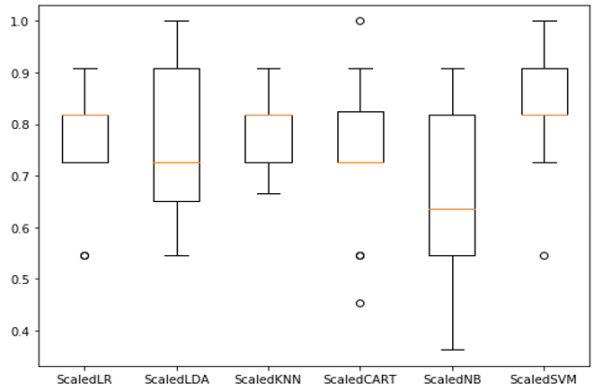
NB: Naive Bayes classifier.

SVM: Support Vector Machine.

e) Scaled Algorithms Accuracy Values

```
ScaledLR: 0.758586 (0.098421)
ScaledLDA: 0.765657 (0.130205)
ScaledKNN: 0.789899 (0.075427)
ScaledCART: 0.752525 (0.143171)
ScaledNB: 0.655556 (0.157505)
ScaledSVM: 0.836869 (0.106380)
```

f) Scaled Algorithm Comparison



ScaledLR: Scaled Logistic Regression.

ScaledLDA: Scaled Linear discriminant analysis.

ScaledKNN: Scaled k-nearest neighbors.

ScaledCART: Scaled Classification and Regression Trees.

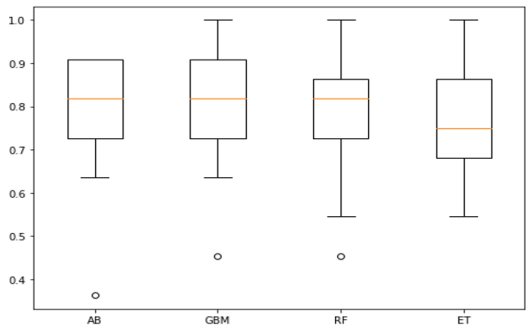
ScaledNB: Scaled Naive Bayes classifier.

ScaledSVM: Scaled Support Vector Machine.

g) Bosting Algorithms Accuracy Values

```
AB: 0.770707 (0.144575)
GBM: 0.783838 (0.134096)
RF: 0.783333 (0.135570)
ET: 0.777273 (0.131792)
```

f) Ensemble Algorithm Comparison



AB: Adabosting

GBT: Gradient bosting Machine.

RF: Random Forest.

ET: Extra Tree Classifier.

g) Confusion Matrix Classification

	precision	recall	f1-score	support
M	0.92	0.85	0.88	27
R	0.76	0.87	0.81	15
accuracy			0.86	42
macro avg	0.84	0.86	0.85	42
weighted avg	0.86	0.86	0.86	42

V.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 random forest model, which is an ensemble tree-based classifier in machine learning with the highest accuracy rate of 83.17% and giving the best ROC-AUC rate 0.93, with least error for better elaboration of this prediction model. For future work more, complex data will be handled using big data Hadoop framework. With random forest algorithm, the results are further optimized by feature selection to get the accuracy of 91.15%.

References

1). N.Hooda, Nishtha et al. "Fraudulent Firm Classification: A Case Study of an External Audit." Applied Artificial Intelligence 32.1: 48-64. (2018).

2). Ho, Tin Kam. Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. (1995).

3). 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).

4). Erkmen, Burcu, and TülayYıldırım. "Improving classification performance of sonar targets by applying general regression neural network with PCA." Expert Systems with Applications 35.1-2: 472-475. (2008).

5). Bacardit, Jaume, and Martin V. Butz. "Data mining in learning classifier systems: comparing XCS with GAssist." Learning Classifier Systems. Springer, Berlin, Heidelberg. 282-290. (2007).

6). N. Hooda et al. “B 2 FSE framework for high deimensional imbalanced data: A case study for drug toxicity prediction”, Neurocomputing, (2018)

7). Corinna, Cortes; Vladimir N., Vapnik. "Support-vector networks". Machine Learning. 20 (3): 273–297. doi: 10.1007/BF00994018. (1995).

8). Kégl, Balázs. "The return of AdaBoost.MH: multiclass Hamming trees". arXiv:1312.6086. (20 December 2013).

9).Pearl, Judea. Causality: Models, Reasoning, and Inference. Cambridge University Press. ISBN 0-521-77362-8. OCLC 4229125. (2000).

10). Huang, Jin. Performance measures of machine learning. University of Western Ontario, (2006).

11). Bradley, Andrew P. "The use of the area under the ROC curve in the evaluation of machine learning algorithms." Pattern recognition 30.7: 1145-1159. (1997).

About Authors:

T.Nikitha is currently pursuing her B.E in Computer Science & Engineering Department, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana. Her areas of interest include Machine Learning and Neural Networks.

B.Hadhvika is currently pursuing her B.E in Computer Science & Engineering Department, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana. Her areas of interest include Machine Learning and Artificial Intelligence.

Dr.K.Vaidehi is currently working as an Associate Professor in Computer Science & Engineering Department, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana. She received her Ph.D. from Anamalai University, Tamil Nadu. Her areas of interest include Image Processing, Data Mining, Signal Processing and Pattern Classification Techniques.