IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sonar Signal Classification: Comparative Analysis of Machine Learning Models for Rock vs Mine Prediction

[1]Chaitanya B K, Student(MCA), Department of Masters Of Computer Application,

[2]Prof. Reshma H Husennaik, Assistant Professor, Department of Masters Of Computer Application

[1]Department of Masters Of Application,
[1]Administrative Management College, Bangalore, INDIA

Abstract: In marine operations, sonar devices are plays a significant role for detecting submerged objects such as rocks and mines. Accurate classification of these signals is necessary to ensure marine safety and security. This study presents a comparative analysis of machine learning algorithms applied to the Sonar dataset. We used a dataset from Kaggle to train a machine- learning model for underwater mine and rock classification, consisting of 208 samples with 60 features. In this work, a comprehensive machine learning pipeline is developed that incorporates data preprocessing, feature scaling, model training, and evaluation. A Support Vector Machine (SVM) model with a Radial Basis Function (RBF) Kernel is optimized using GridSearchCV with cross-validation to achieve optimal performance. Multiple Supervised Machine learning algorithms—including Logistic Regression, Support Vector Machine (SVM), Random Forest, Neural Network (MLP), Stochastic Gradient Descent (SGD), and K-Nearest Neighbors (KNN)- are implemented for comparative analysis.

Model performance is evaluated using various metrics such as accuracy, precision, recall, F1-score, confusion matrix, and Receiver Operating Characteristic (ROC) curve with Area Under the Curve (AUC). Stratified k-fold cross-validation further confirms the robustness and generalization capability of the model.

The results demonstrate that SVM outperforms other models in balancing precision and recall for this dataset. The tuned SVM with RBF kernel achieved 92.85% accuracy and 0.97 AUC, outperforming most baseline models. Neural Networks and KNN also demonstrated competitive performance. Visualizations, including ROC curves, confusion matrices, grouped bar charts, and cross-validation, confirmed model robustness. The proposed system was deployed as a predictive application, enabling real-time classification of Sonar Signals..

Keywords: sonar signal classification, comparative analysis, supervised machine learning, logistic regression, support vector machine (svm), random forest, neural network (mlp), stochastic gradient descent (sgd), and k-nearest neighbors (knn).

I.INTRODUCTION

The identification of underwater objects using sonar signals is a critical task in naval operations, marine exploration, and defense systems. Distinguishing between harmless objects such as rocks and potentially dangerous objects like underwater mines is essential for ensuring safety and operational efficiency. Traditional methods of manual interpretation of sonar signals are time-consuming, require expert knowledge, and are prone to human error. With the advancement of data-driven technologies, machine learning has emerged as a powerful tool for automating and improving the accuracy of such classification tasks.

Sonar (Sound Navigation and Ranging) systems operate by emitting sound waves and analysing the reflected signals from objects. These reflected signals contain valuable information about the structure, composition, and shape of underwater objects. However, the high dimensionality and complex patterns present in sonar data make it challenging to analyse using conventional techniques. This necessitates the use of advanced computational methods capable of capturing subtle patterns and relationships within the data.

In recent years, machine learning algorithms have shown significant success in pattern recognition and classification problems. Techniques such as Support Vector Machines (SVM), Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Neural Networks have been widely applied across various domains due to their ability to learn from data and make accurate predictions. Among these, SVM is particularly effective for high-dimensional datasets and small sample sizes, making it well-suited for sonar signal classification.

This study focuses on developing an efficient machine learning model to classify sonar signals into two categories: rocks and mines. The dataset[1] used consists of multiple numerical features representing the energy of sonar signals at different frequencies. A structured machine learning pipeline is implemented, including data preprocessing, feature scaling, model training, and performance evaluation. Various models are compared to identify the most effective approach, with special emphasis on optimizing the SVM model using hyperparameter tuning techniques.

The performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) analysis. Special attention is given to recall, particularly for mine detection, as false negatives can have serious consequences in real-world scenarios. This project builds on prior work in sonar classification by integrating preprocessing, PCA, hyperparameter tuning, and comparative evaluation of multiple classifiers. The goal is to achieve high accuracy and reliable real-time predictions for maritime safety. This demonstrates the effectiveness of machine learning techniques in improving classification accuracy and reliability in sonar-based object detection systems.

In addition to model development, the system is deployed as a web-based application to enable real-time predictions and user interaction. The results demonstrate that machine learning techniques, SVM with optimized parameters, can improve classification accuracy and reliability in sonar-based object detection systems.

II.LITERATURE SURVEY

An effective and efficient result has been achieved by Linear Discriminant Analysis (LDA) with a grid search approach followed by a Support vector machine (SVM) with RBF kernel[2].Navigation and marine security are gravely endangered by underwater mines [3].While sonar systems are effective for locating submerged objects, the similarity in acoustic properties often makes it difficult to tell rocks apart from mines[4].By illustrating advances in predictive analytics, machine learning has captured the focus of an important portion of today's evolving technology, ranging from banking to several consumer and product focused enterprises. The research on machine learning (ML) methods has been carried out to improve the precision and effectiveness of sonar rock vs. mine classification[5,6]. A prediction method using three binary classifier models built by supervised machine learning algorithms in Python is proposed by the researchers in the identification of the potential hazards to marine life and submarines when misidentified as rocks.[7]Sreenivasa R L et.al., This foundational work surveys techniques for underwater mine detection, providing insights into discriminating challenges between rocks and mines, crucial for logistic regression modeling.

III. RESEARCH METHODOLOGY

The methodology adopted in this project focuses on developing an efficient machine learning-based system for classifying sonar signals into rocks and mines through a structured sequence of steps, including data collection, preprocessing, feature selection, model training, optimization, and evaluation. The overall process ensures that the system is accurate, reliable, and suitable for real-world applications.

The first step involves data collection using the Sonar dataset, which consists of 208 samples with 60 numerical features representing energy levels of sonar signals at different frequencies. Each sample is labeled as either Rock (R) or Mine (M). The dataset is then preprocessed to improve data quality and model performance. Preprocessing includes separating features and labels, handling input formats, and splitting the dataset into training and testing sets using an 80:20 ratio. Feature scaling is applied using StandardScaler to normalize the data, ensuring that all features contribute equally to the model and preventing bias due to varying feature ranges.

To address the high dimensionality of the dataset, Principal Component Analysis (PCA) is applied as a feature selection and dimensionality reduction technique. This reduces computational complexity, removes redundancy, and helps improve model efficiency and generalization.

The methodology involves implementing multiple supervised machine learning algorithms, including Support Vector Machine (SVM), Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), Neural Networks (MLP), and Stochastic Gradient Descent (SGD). Among these, the SVM model plays a significant role due to its ability to handle high-dimensional and non-linear data using the Radial Basis Function (RBF) kernel. The kernel function maps the input data into a higher-dimensional space, making it easier to separate classes with complex boundaries.

To further enhance model performance, hyperparameter tuning is performed using GridSearchCV. This technique systematically searches through different combinations of parameters such as the regularization parameter (C) and kernel coefficient (gamma) for the SVM model. Cross-validation is used during this process to ensure that the selected parameters provide the best generalization performance across different subsets of the data. Once the models are trained, they are evaluated using various performance metrics, including accuracy, precision, recall, and F1-score. A confusion matrix is used to analyze classification errors, while ROC curves and AUC values are used to assess the model's ability to distinguish between classes. Comparative analysis is performed across all models to identify the best-performing algorithm.

In addition to model evaluation, the methodology includes the development of a web-based application using the Flask framework. This application allows users to input sonar data manually and obtain real-time predictions. The system also generates an input signal visualization by plotting the sonar signal, providing a graphical representation of the data for better interpretability.

Finally, the best-performing model is saved using pickle for deployment, and performance visualizations such as ROC curves and comparison graphs are generated. Cross-validation techniques, such as Stratified K-Fold, are used to ensure the robustness and stability of the model. Overall, the methodology integrates data preprocessing, model optimization, evaluation, and deployment into a cohesive pipeline for effective sonar signal classification.

.

IV. PROPOSED SYSTEM

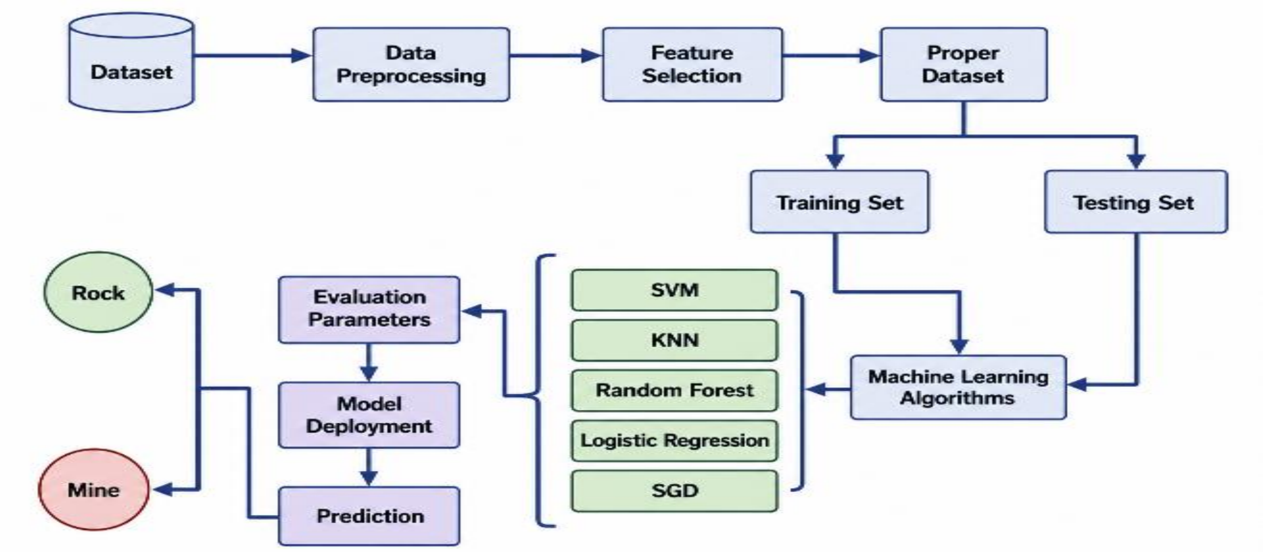

Fig. 1.Architecture of the Proposed System.

From The architecture of the proposed system as shown in Fig.1.

Dataset Loading (Data Collection): The dataset used in this project is the Sonar Mines vs Rocks dataset, which consists of 208 rows and 61 columns. Each row represents a unique sonar signal pattern reflected from an object underwater. Out of the 61 columns, 60 columns are numerical features, and the last column represents the target label.

The 60 input features contain values ranging from 0.0 to 1.0,These values represent the energy of sonar signals in different frequency bands, integrated over specific time intervals. Essentially, each feature captures how much sound energy is reflected back at a particular frequency range, helping in identifying the nature of the object.

The final column contains the class label:
R → Rock
M → Mine
These labels indicate whether the sonar signal was reflected from a natural object (rock) or a man-made object (mine).
This dataset is widely used for binary classification problems in machine learning and is obtained from a reliable source, such as the Kaggle dataset repository. It plays a crucial role in training the model to accurately distinguish between rocks and mines based on sonar signal patterns.
Data Preprocessing: In this study, data preprocessing is fundamental to preparing the Sonar Mines vs Rocks dataset for effective machine-learning classification. The preprocessing steps ensure that the raw data is transformed into a structured and model-ready format. The dataset is obtained from Kaggle and loaded into the Python environment using Pandas. Each instance represents a sonar signal pattern measured across multiple frequency bands. The target variable in the dataset is categorical, consisting of two classes: Rock (R) and Mine (M). Since machine learning algorithms require numerical input, label encoding is applied to convert:
Rock (R) → 0
Mine (M) → 1
Feature Scaling: Although the dataset features are already normalized within the range of 0.0 to 1.0, feature scaling is further applied in some models to standardize the distribution. Standardization using techniques such as StandardScaler ensures that all features contribute equally to the model training process. This step is particularly important for distance-based and margin-based classifiers such as Support Vector Machines (SVM).
Model Training : The dataset is first divided into input features (X) and target labels (y). The features represent sonar signal energy values across different frequency bands.
Train-Test Split: The dataset is then split into training and testing sets using an 80:20 stratified split. Stratification ensures that both classes (Rock and Mine) are evenly distributed in training and testing data, which is important for balanced learning.
Model Selection, Pipeline Construction, Hyperparameter Tuning, and Model Fitting: the primary machine learning model selected for classification is the Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. SVM is chosen due to its strong capability in handling small datasets (208 samples) and high-dimensional feature spaces (60 sonar features). It is particularly effective in binary classification problems, as it works by finding an optimal hyperplane that maximally separates the two classes, i.e., Rock and Mine, with the largest possible margin.
To ensure a structured and error-free workflow, a Machine Learning Pipeline is constructed. The pipeline integrates two main components: StandardScaler for feature scaling and SVC (Support Vector Classifier) for classification. The inclusion of StandardScaler ensures that all input features are normalized to a standard distribution, preventing features with larger numerical influence from dominating the learning process. By embedding both preprocessing and model training into a single pipeline, consistency is maintained during both training and testing phases, reducing the risk of data leakage.
To further enhance model performance, Hyperparameter Tuning is performed using GridSearchCV. This step systematically searches for the best combination of parameters to optimize model accuracy and generalization
The key hyperparameters tuned include:
C (Regularization parameter): Controls the trade-off between maximizing the margin and minimizing classification error.
Gamma: Defines how far the influence of a single training example reaches, affecting the flexibility of the decision boundary.
Kernel: The RBF(Radial Basis Function) kernel is selected for its ability to handle non-linear relationships in sonar signal data.
Through 5-fold cross-validation, GridSearchCV evaluates multiple parameter combinations and identifies the optimal configuration as:
C = 10
Gamma = Scale
Kernel = RBF
This process ensures that the model is not only accurate but also robust and less prone to overfitting.
Model Evaluation: The main objective is to determine how effectively the model can classify sonar signals into Rock (R) and Mine (M) categories and how well it generalizes beyond the training dataset. After training the Support Vector Machine (SVM) model, predictions are made on both training and testing datasets.

Training Accuracy = 100%
Testing Accuracy = 92.85%
The perfect training accuracy indicates that the model has learned the training data patterns completely.
The confusion matrix provides a detailed breakdown of correct and incorrect predictions:

[[15 1]
[2 24]]

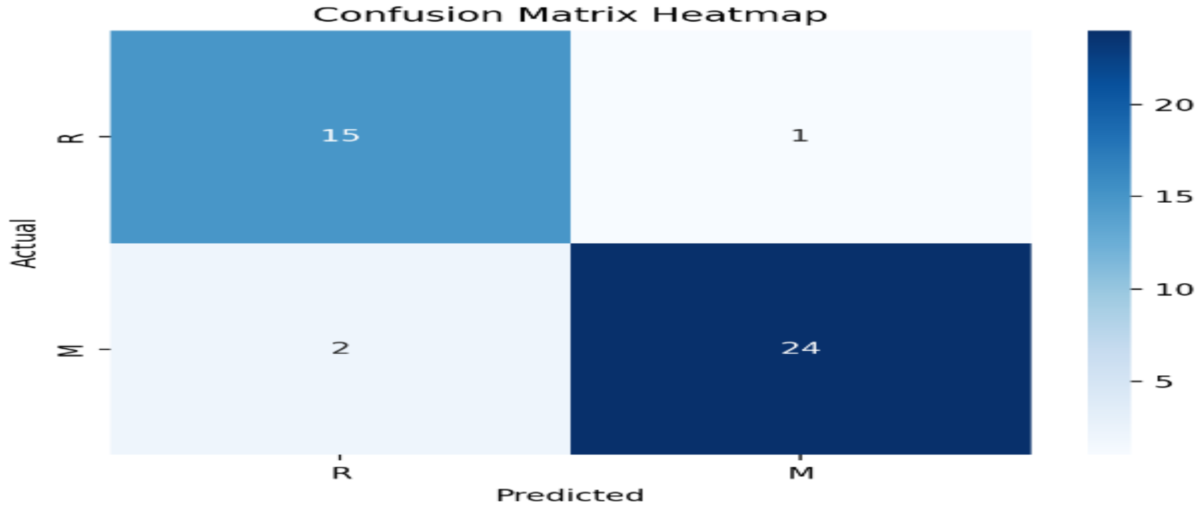

Fig.2. Confusion Matrix

Cross-Validation Analysis : To ensure model stability and reliability, Stratified 5-Fold Cross Validation is performed. The dataset is divided into five equal parts, and the model is trained and tested iteratively.
Fold accuracies: 0.85 to 0.92
MeanCross-Validation Accuracy = 87.06%

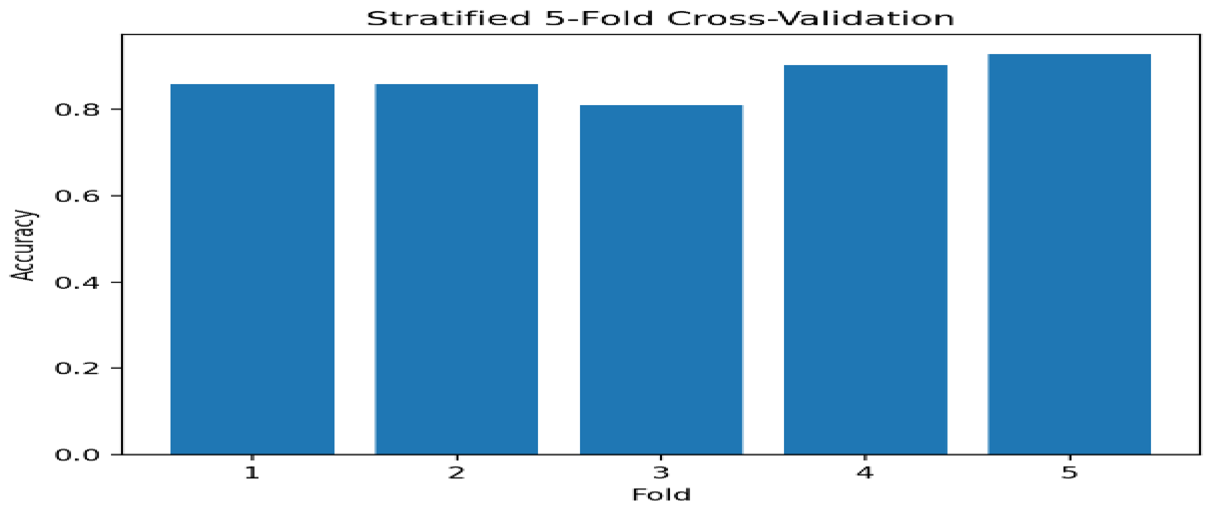

Fig. 3. Stratified F-fold Cross Validation.

Stratified F-fold Cross Validation. Fig 3. confirms that the model performs consistently across different data splits and is not dependent on a specific training-test division.

V. MODEL COMPARISON

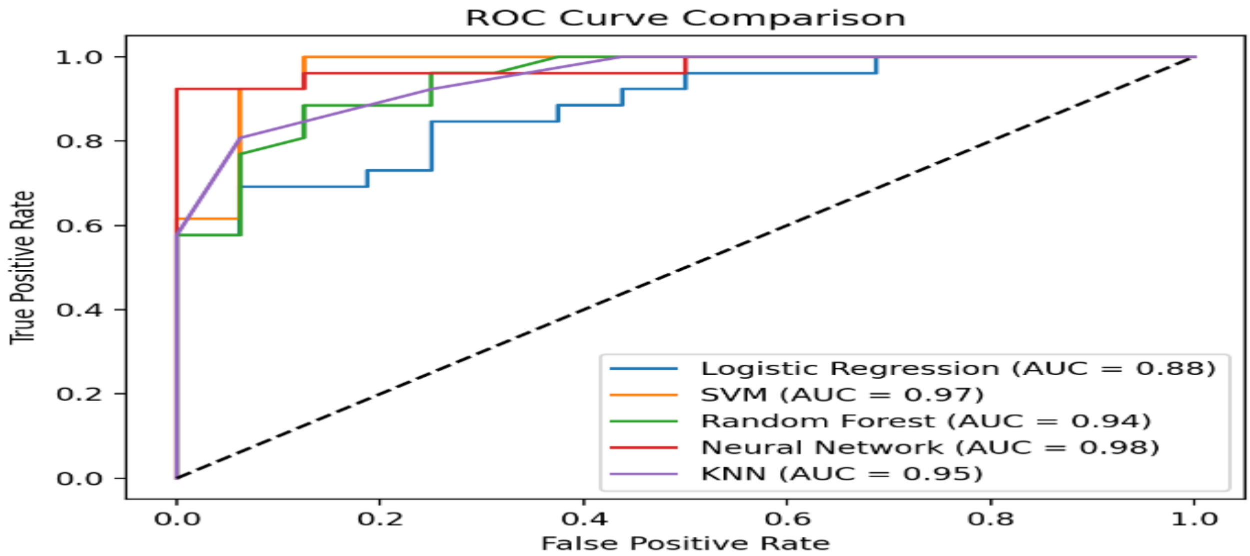

Fig. 4. ROC Curve Comparison.

Multiple machine learning models are evaluated for performance comparison evaluation Parameters Fig4.: The analysis of the model's performance is carried out in evaluation parameters.

In this project, the performance of the machine learning model for classifying sonar signals into rocks and mines is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.Fig5.Model accuracy. These metrics are based on four key outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

True Positive (TP): Correctly classified mines

True Negative (TN): Correctly classified rocks

False Positive (FP): Rock incorrectly classified as mine

False Negative (FN): Mine incorrectly classified as rock .

MODEL EVALUATION

Accuracy : Accuracy measures the overall correctness of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision : Precision measures how many predicted mines are actually correct.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall : Recall measures how many actual mines are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score : F1-score is the harmonic mean of precision and recall, providing a balanced performance measure.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Model	Accuracy	Recall	Precision	F1_score
SVM	92.857143	92.307692	96.000000	94.117647
Neural Network	90.476190	84.615385	100.000000	91.666667
KNN	85.714286	80.769231	95.454545	87.500000
Random Forest	80.952381	76.923077	90.909091	83.333333
SGD	80.952381	76.923077	90.909091	83.333333
Logistic Regression	73.809524	73.076923	82.608696	77.551020

Fig. 5. Model Accuracy.

VI .PREDICTIVE SYSTEM

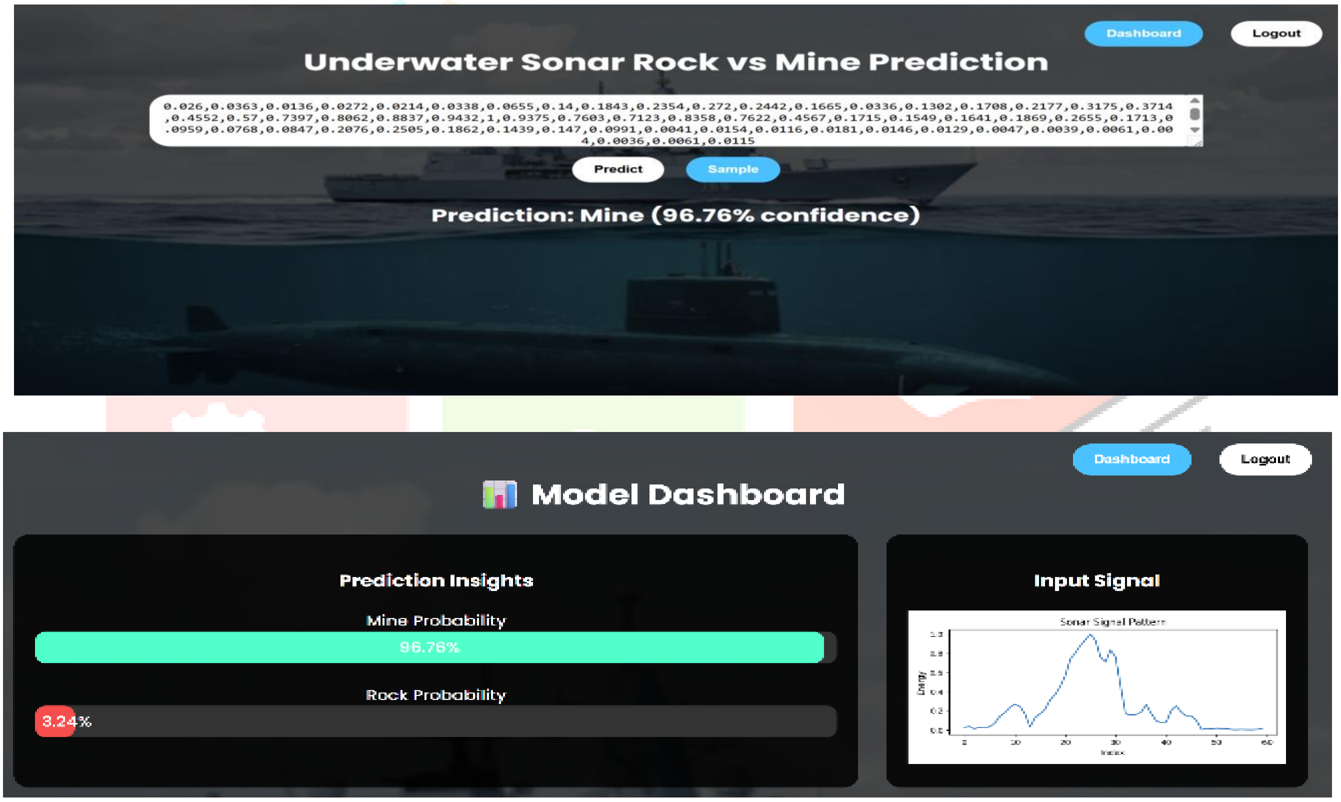

Fig 6. Web-Based Application for Rock vs Mine Prediction

To enhance the practical usability of the proposed machine learning model, a web-based application fig 6. is developed using the Flask framework. The objective of this application is to provide a user-friendly interface that allows real-time prediction of sonar signals, enabling users to classify underwater objects as either rocks or mines without requiring technical expertise. This deployment bridges the gap between theoretical model development and real-world application.

The web application is designed with an interactive dashboard that allows users to input sonar data in the form of 60 numerical values, corresponding to signal energy levels across different frequency bands. The system ensures robust input validation by restricting the user to enter exactly 60 numeric values, thereby preventing incorrect or incomplete data submissions. Any invalid input triggers a warning message, improving the reliability and usability of the interface.

Once valid input data is submitted, the trained Support Vector Machine (SVM) model processes the data and generates a prediction. The output is displayed in a clear and interpretable format, indicating whether the signal corresponds to a rock or a mine. In addition to the predicted class, the system also provides prediction confidence scores in the form of probability values for both classes. These probabilities are visually represented using progress bars, labeled as Mine Probability and Rock Probability. This feature enhances transparency by allowing users to understand how confident the model is in its prediction, rather than providing only a binary output.

A key feature of the application is the Input Signal Visualization, which graphically represents the sonar signal pattern entered by the user. The visualization is generated using Matplotlib and displayed as a line graph, where the x-axis represents the feature index (frequency bands) and the y-axis represents the corresponding energy values. This graphical representation helps users analyze the structure and variation of the input signal. Since sonar signals differ significantly between rocks and mines, this visualization can provide intuitive insights into the nature of the object being analyzed.

The input signal graph is dynamic in nature, meaning it updates every time a new set of data is provided. This ensures that the visualization always reflects the current input rather than previously stored data. Such real-time visualization is particularly useful for understanding how variations in signal patterns influence model predictions.

VII. RESULTS

This work integrates the overall system workflow through a structured pipeline consisting of data collection, preprocessing, model training, evaluation, and prediction for sonar-based classification of rocks and mines Fig6. The dataset used in this study involves identifying data characteristics using visualization techniques such as confusion matrix heatmaps and ROC curves, which help in understanding model behavior and classification performance. Unlike traditional studies, explicit outlier removal is not required as sonar signal data is already preprocessed and noise-reduced. In this work, multiple machine learning algorithms are evaluated to determine the most effective classifier for sonar signal classification. The models include:

Support Vector Machine (SVM) with RBF kernel
Logistic Regression
Random Forest
K-Nearest Neighbors (KNN)
Neural Network (MLP)
Stochastic Gradient Descent (SGD)

Among these models, SVM with RBF kernel is selected as the primary model due to its superior balance between accuracy and generalization on small datasets. GridSearchCV is applied for hyperparameter tuning, optimizing parameters such as C and gamma, which significantly improves model performance.

The classification report indicates that the model performs better in detecting mines (high recall), which is critical for real-world safety applications. A confusion matrix analysis further confirms that only a small number of misclassifications occur. Cross-validation using Stratified 5-Fold validation gives a mean accuracy of 87.99%, proving that the model is stable and generalizes well across different data splits.

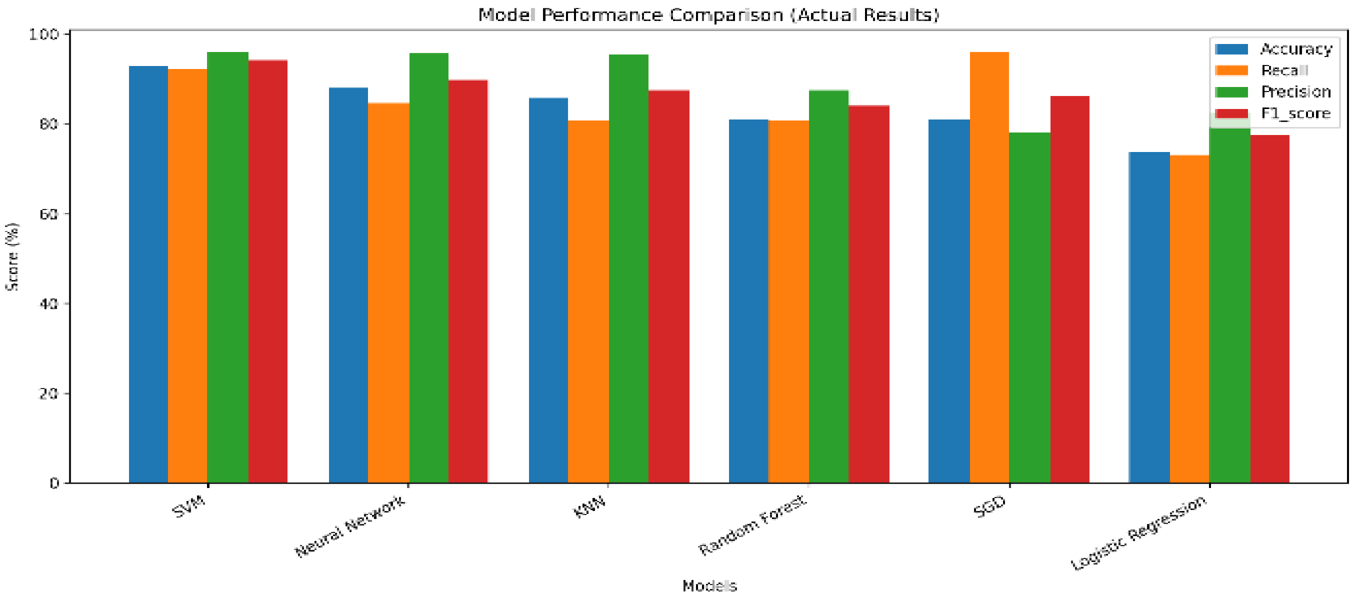

Fig. 7. Model Performance Comparison Bar Graph.

VIII. CONCLUSION OF ANALYSIS

In this project, a machine learning model was developed to classify sonar signals as either rock or mine. The dataset was processed using preprocessing techniques, and multiple models such as SVM, Logistic Regression, Random Forest, KNN, Neural Network, and SGD were trained and compared.
Among all the models, the Support Vector Machine (SVM) with RBF kernel gave the best performance. It achieved high accuracy and showed good balance between precision and recall. Cross-validation also proved that the model is stable and performs well on different data samples.
The model was further deployed as a web application using Flask, where users can enter 60 sonar values and get real-time predictions. The system also shows probability values and a signal graph, which makes the prediction more understandable.
Overall, this project shows that machine learning can effectively classify sonar signals and can be used for real-world applications like underwater object detection and marine safety.

IX. FUTURE SCOPE

The future scope of this project involves improving the model by using larger and real-time sonar datasets to enhance accuracy and reliability. Advanced techniques such as deep learning can be applied to capture more complex patterns in the data. The system can also be developed into a real-time application integrated with actual sonar devices for practical use in marine environments. Additionally, deploying the application on cloud or mobile platforms can increase accessibility. Further improvements can include better visualization dashboards and extending the model to classify multiple types of underwater objects beyond just rocks and mines.

REFERENCES

[1]UCI Machine Learning Repository. Sonar Dataset. Available at: https://archive.ics.uci.edu/ml/datasets/sonar
[2] S. Sahoo, et,al., “A Multi-Faceted Approach for Underwater Sonar Rock Vs Mine Classification Using Machine Techniques,”.
[3] Sri Ramya Yaramasu, Uppada Sai Gayatri, Vadlamani Manjusha,Vaishna C Bhanu, Koda Indu, “Underwater mine & rock prediction byevaluation of machine learning algorithms”.
[4] Siddharth, Jetty Bangaru, T. Jaya, and V. Rajendran. "RNN for classification and prediction of Rock/Mine in nderwater acoustics."
[5] Nikitha, ., B. Hadhvika, and K. Vaidehi. "Prediction of Underwater.
[6] Padmaja, Venkataraman, V. Rajendran, and P. Vijayalakshmi. "Study On metal mine detection from underwater sonar images using data mining and machine learning techniques."

[7] Sreenivasa R L et.al., “Analysis of hidden units in Layered Network Trained to Classify Sonar Targets” in Neural Networks.
[8]S. Pathak, V. Raul, P. Desale, D. Bhandari, and P. Sanjekar, "Sonar Rock vs Mine Prediction,"