KAUNAS UNIVERSITY OF TECHNOLOGY

FACULTY OF MATHEMATICS AND NATURAL SCIENCES

SRIHARI JOTHIMANI

Study module

P170M121 DATA SCIENCE PROJECT

Sonar Sensor Performance Analysis for Underwater Mine Detection

Data Science Project

KAUNAS, 2026

CONTENT

Sonar Sensor Performance Analysis for Underwater Mine Detection

Abstract

Underwater mine detection serves as a critical security mission which military organizations perform by using sonar sensor technology. The process of using sonar to identify dangerous objects like mines against harmless objects like rocks becomes difficult because of two main issues which are environmental interference and reduced signal strength and similar acoustic characteristics between objects. The research investigates how data science methods perform when they use sonar frequency-band characteristics to identify underwater mines from rocks. Using a publicly available benchmark dataset, the full data science lifecycle is applied, including data acquisition, preprocessing, exploratory analysis, modeling, evaluation, and interpretation. The research evaluates different machine learning approaches which Support Vector Machines (SVM) produce the highest detection results. The research examines prediction confidence and detection trade-offs and feature stability to establish if classification errors stem from model weaknesses or they result from natural sonar sensor uncertainty. The results show that machine learning technology improves detection accuracy but sonar signal interference makes it impossible to identify all targets correctly.

1. Introduction

Sonar technology performs vital duties in underwater operations because it allows navigation and seabed mapping and defends against threats. Sonar sensors serve their most vital purpose during mine countermeasure operations because they help detect underwater objects which could pose threats. A key challenge in this context is distinguishing man-made mines from natural seabed objects such as rocks. Underwater environments create multiple types of disturbances which produce noise and reverberation and multipath effects that modify acoustic signals and make different objects generate identical sonar responses. The existing conditions make it more probable for systems to produce incorrect alerts while failing to identify actual events which would result in major operational threats.

The present state of data science and machine learning technology allows researchers to successfully analyze complex sensor data. The research project evaluates how well sonar technology identifies objects while it identifies the factors which lead to incorrect classification results. The research uses machine learning and interpretive analysis to study actual sonar data which enables the detection of system detection accuracy and specific sensor performance restrictions.

2. Problem Statement

The main objective of this research involves testing sonar frequency-band characteristics for their ability to identify underwater mines from rocks and to determine if machine learning model weaknesses or sonar sensor measurement uncertainties lead to detection mistakes.

3. Research Hypotheses

The research investigates two hypotheses which aim to evaluate both the informative value of sonar features and the basic limitations which affect sonar detection systems.

Hypothesis 1 (H1): Feature Dominance Hypothesis

A limited subset of sonar frequency-band features carries the majority of the discriminative information needed to separate mines from rocks.

Sonar sensors measure energy across multiple frequency bands but physical principles show that different frequencies provide different amounts of information. The interaction between specific frequency ranges and object geometry and material composition and surface characteristics will be more pronounced. The evaluation of this hypothesis required multiple training sessions of Random Forest models which used various random starting points. The analysis of feature importance rankings throughout these runs revealed which frequency bands appeared most often in the classification process. The analysis showed that few specific frequency bands appeared as the most important features but most bands failed to produce significant results. The results confirm our prediction that discrimination depends on only a few specific spectral characteristics.

\

Hypothesis 2 (H2): Intrinsic Sensor Ambiguity Hypothesis

The research examines system performance behavior when built-in sensor uncertainty exists through Hypothesis 2 (H2).

The main cause of classification errors stems from the natural overlap between sonar sensor responses instead of having to do with limited modeling ability.

The research hypothesis aims to identify the reasons which cause machine learning models with their sophisticated capabilities to produce ongoing classification mistakes. The research included multiple supporting studies which used Principal Component Analysis (PCA) and two additional methods to analyze average frequency-band energy profiles and prediction confidence scores and nonlinear classification performance. The PCA visualizations showed that mine and rock samples contained many identical characteristics which made it hard to identify their differences. The energy profile analysis showed that the system generated equal responses when it tested both objects using different frequency bands. The SVM model produced excellent prediction results but it still made some mistakes which mostly occurred in areas where the model showed weak prediction certainty. The research findings show that model errors result from similar sonar data patterns instead of insufficient model parameters which confirms Hypothesis

4. Dataset Description

The analysis uses the Sonar Mines vs Rocks dataset obtained from the UCI Machine Learning Repository. The dataset includes 208 sonar observations which are described by 60 numerical features that represent energy measurements across various frequency bands. Each observation is labeled as either a mine or a rock. The dataset is relatively balanced between classes and contains no missing values. Because the features are already derived from sonar signal processing, the dataset is well suited for machine learning-based evaluation without additional signal transformation.

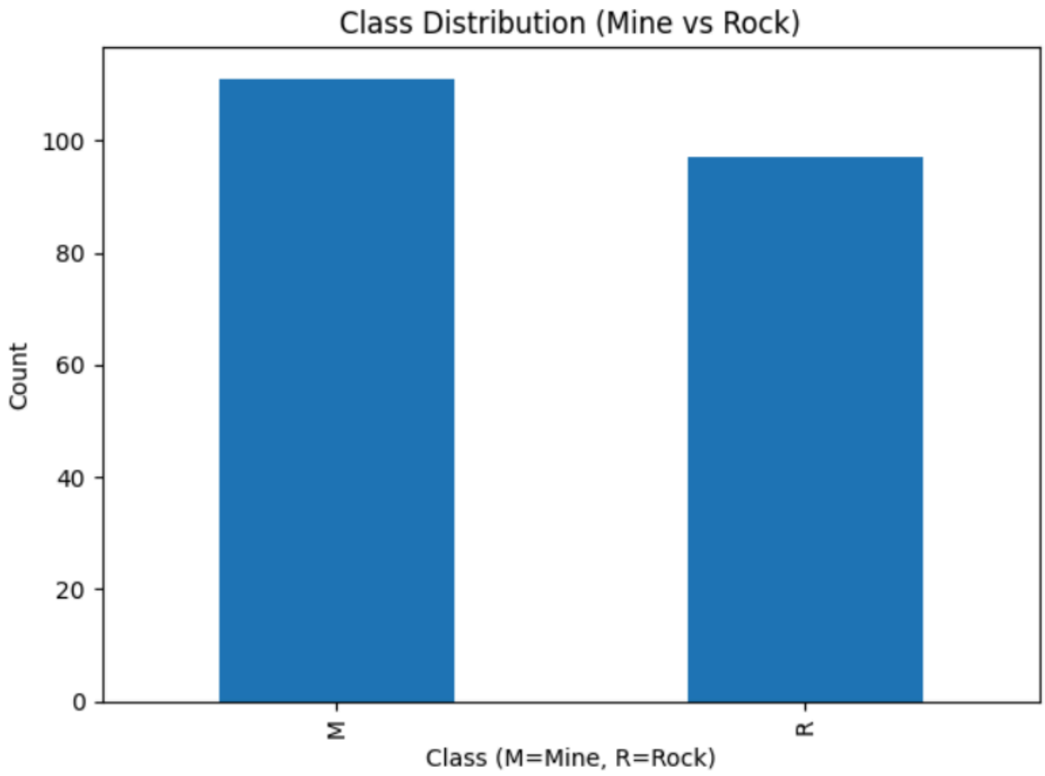

Fig 1 : Class distribution of mine and rock samples in the sonar dataset.

5. Data Processing and Feature Engineering

The data processing required a controlled process to add data to the system while checking its accuracy for achieving both reliable results and duplicate-free outcomes. All features were converted into numerical form, and checks confirmed the absence of missing or invalid values. Feature standardization was applied so that all frequency bands were placed on a common scale, which is particularly important for distance-based learning algorithms. The evaluation process employed a stratified train–test split method which preserved equal class distributions during the assessment.

The process of feature engineering focused on creating features which scientists could understand instead of generating new artificial characteristics. PCA was employed strictly for exploratory purposes to assess underlying structure and separability in the data, while Random Forest-based feature importance analysis was used to identify dominant frequency bands relevant to discrimination.

6. Exploratory Data Analysis

The exploratory analysis revealed that mine and rock samples contained many identical characteristics which spanned across different frequency bands. PCA using the first two principal components captured less than half of the total variance and showed no clear separation between classes. The analysis of mean frequency-band energy revealed that only a few bands displayed significant variations but the majority of bands maintained equivalent energy levels between the two object categories. The observed data contains complex elements which make it impossible to achieve complete separation through spectral data analysis for the classification task.

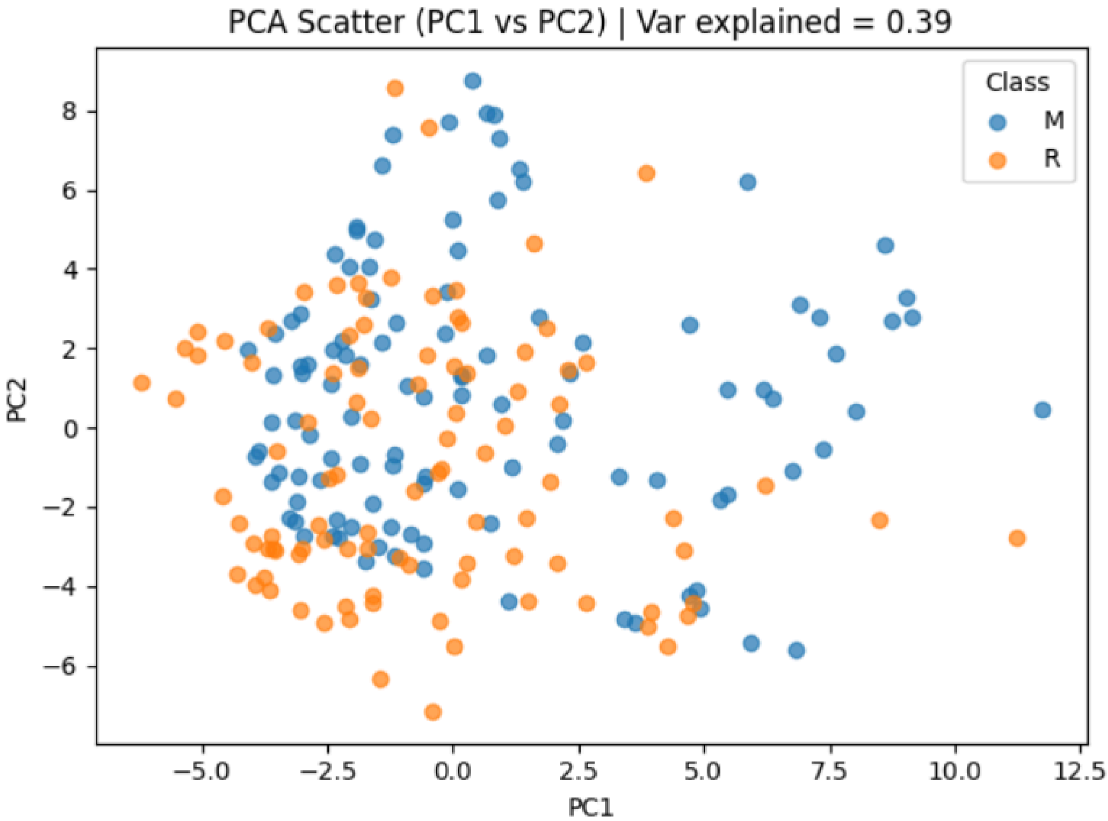

Fig 2 : The PCA scatter plot between PC1 and PC2 dimensions shows that mine and rock samples contain similar characteristics.

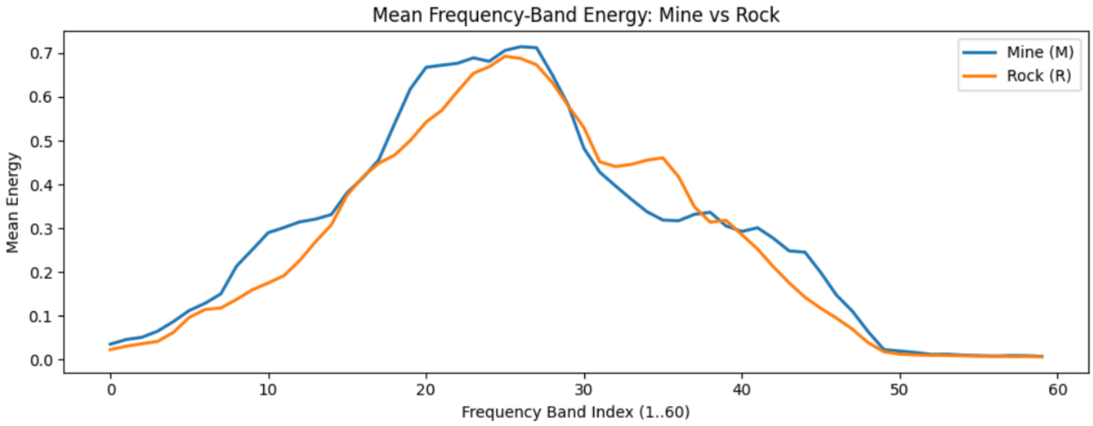

Fig 2.1 : Mean frequency-band energy comparison between mines and rocks.

7. Modeling Methodology

Three machine learning approaches were evaluated. Logistic Regression was used as a baseline linear model to assess whether the data could be separated using linear decision boundaries. A Support Vector Machine with a radial basis function kernel was selected as the primary nonlinear model due to its flexibility in modeling complex boundaries. Random Forest classifiers were employed both as an alternative nonlinear model and as a tool for feature importance analysis. The model robustness evaluation through repeated stratified cross-validation demonstrated that SVM produced superior results which were also more dependable during each testing iteration.

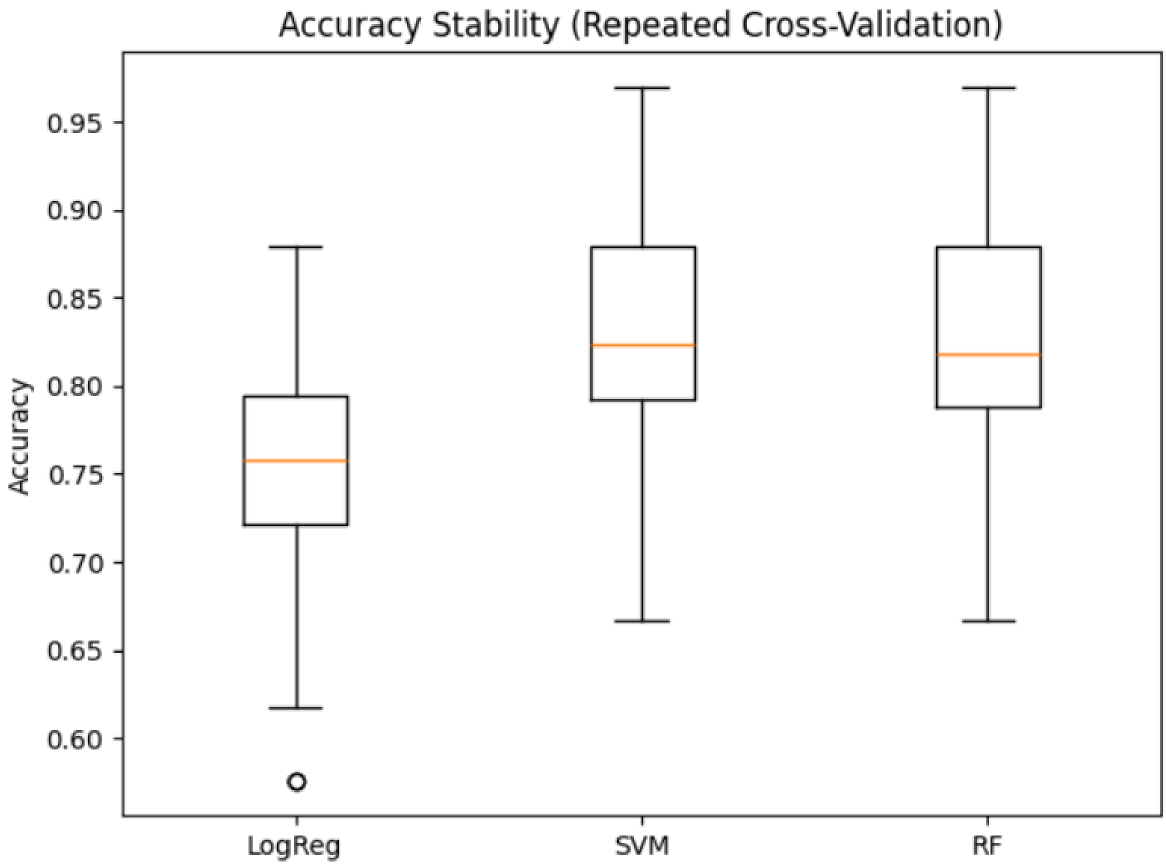

Fig 3 : Accuracy stability comparison across models using repeated cross-validation.

8. Experimental Results

The SVM model achieved 85-86% classification accuracy when it was tested on new data while generating a ROC–AUC value of 0.92. The confusion matrix showed that the system detected mines with high accuracy but it also produced some incorrect results by identifying rocks as mines. The ROC curve analysis showed that the system produced an excellent relationship between its ability to identify correct events and its production of wrong alerts. The prediction probability analysis showed that most prediction errors occurred when the system showed moderate confidence because it failed to select the correct answer when the outcome was ambiguous.

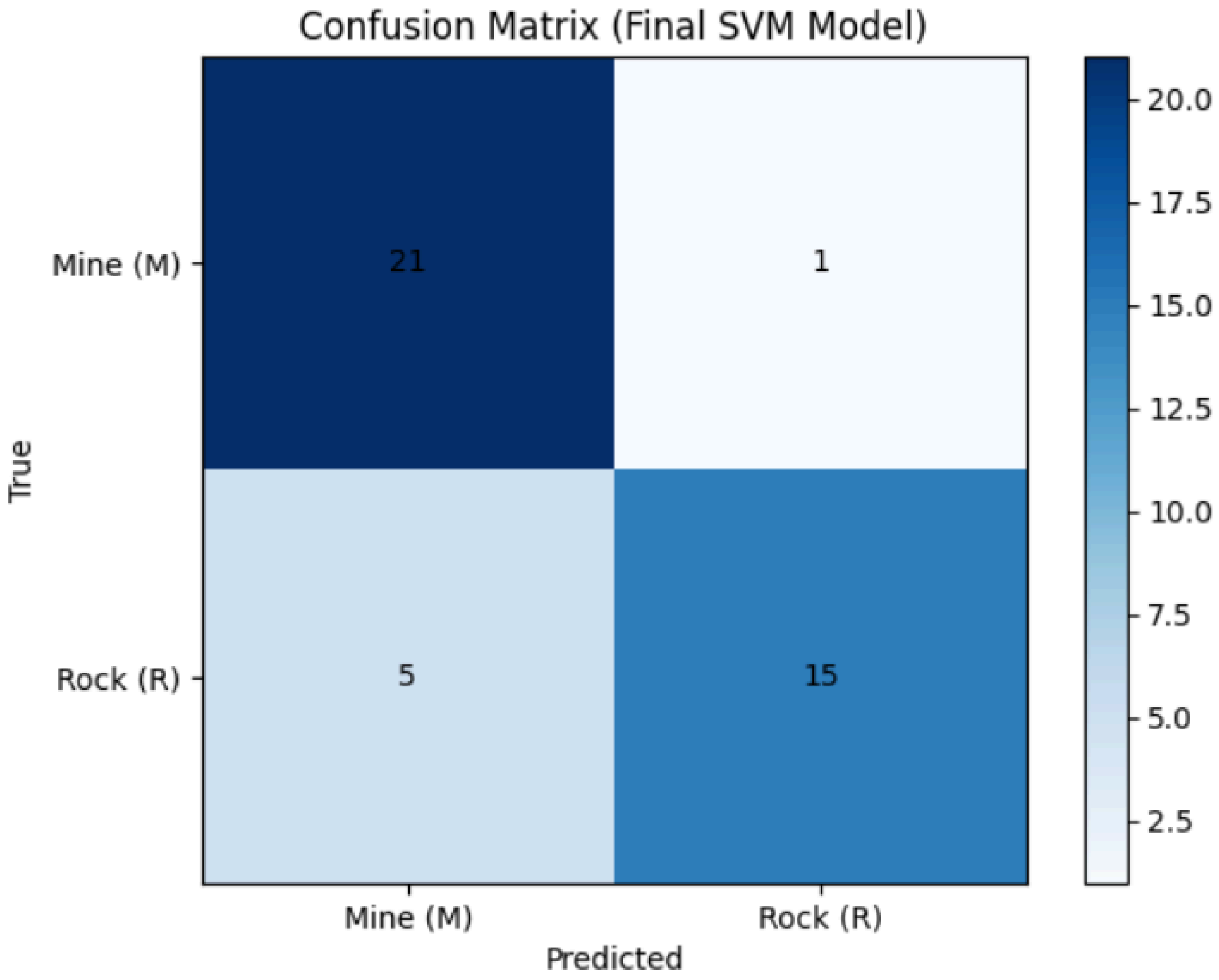

Fig : 4 : Confusion matrix for the final SVM model on unseen test data.

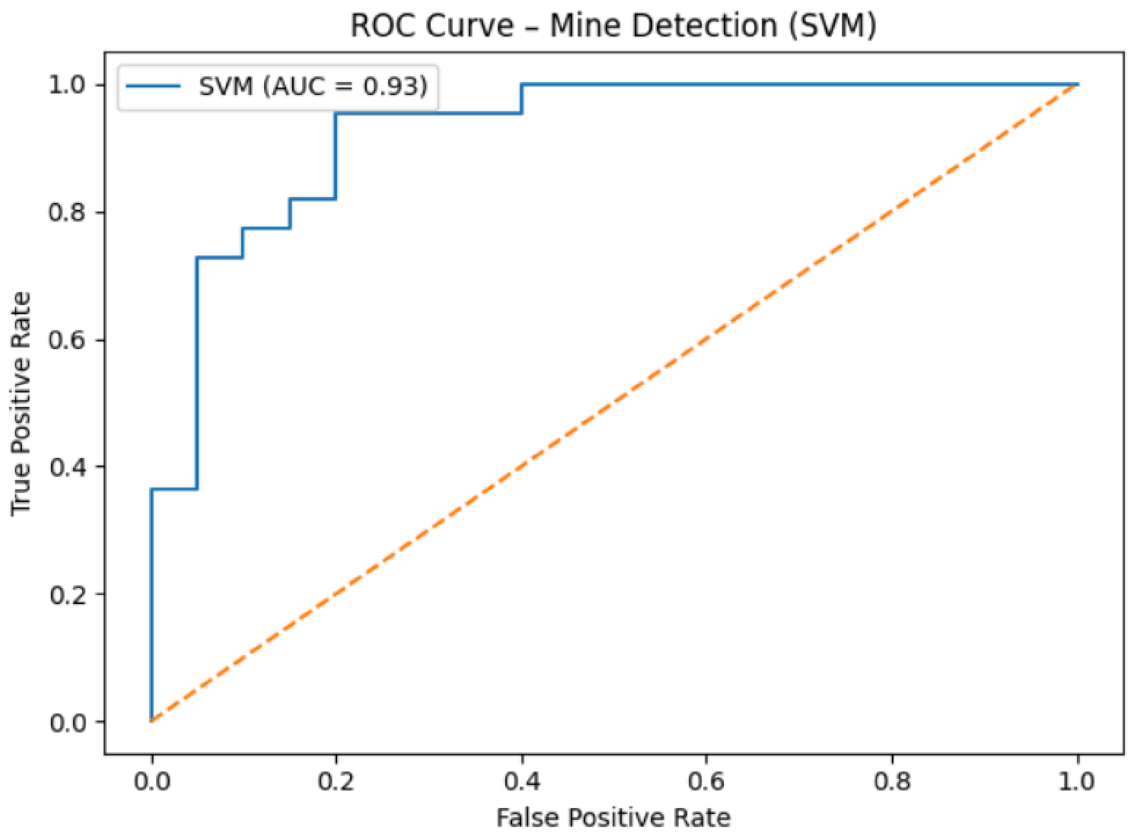

Fig 4.1 : ROC curve illustrating mine detection performance of the SVM model.

9. Hypothesis Evaluation

The experimental results confirmed all research hypotheses. The stability analysis of feature importance showed that only a few specific frequency bands maintained their ability to help with classification which supports Hypothesis 1. The results from PCA visualizations and confidence distributions and misclassified samples confirm Hypothesis 2 because they show that built-in sonar sensor ambiguity limits how well the system can perform.

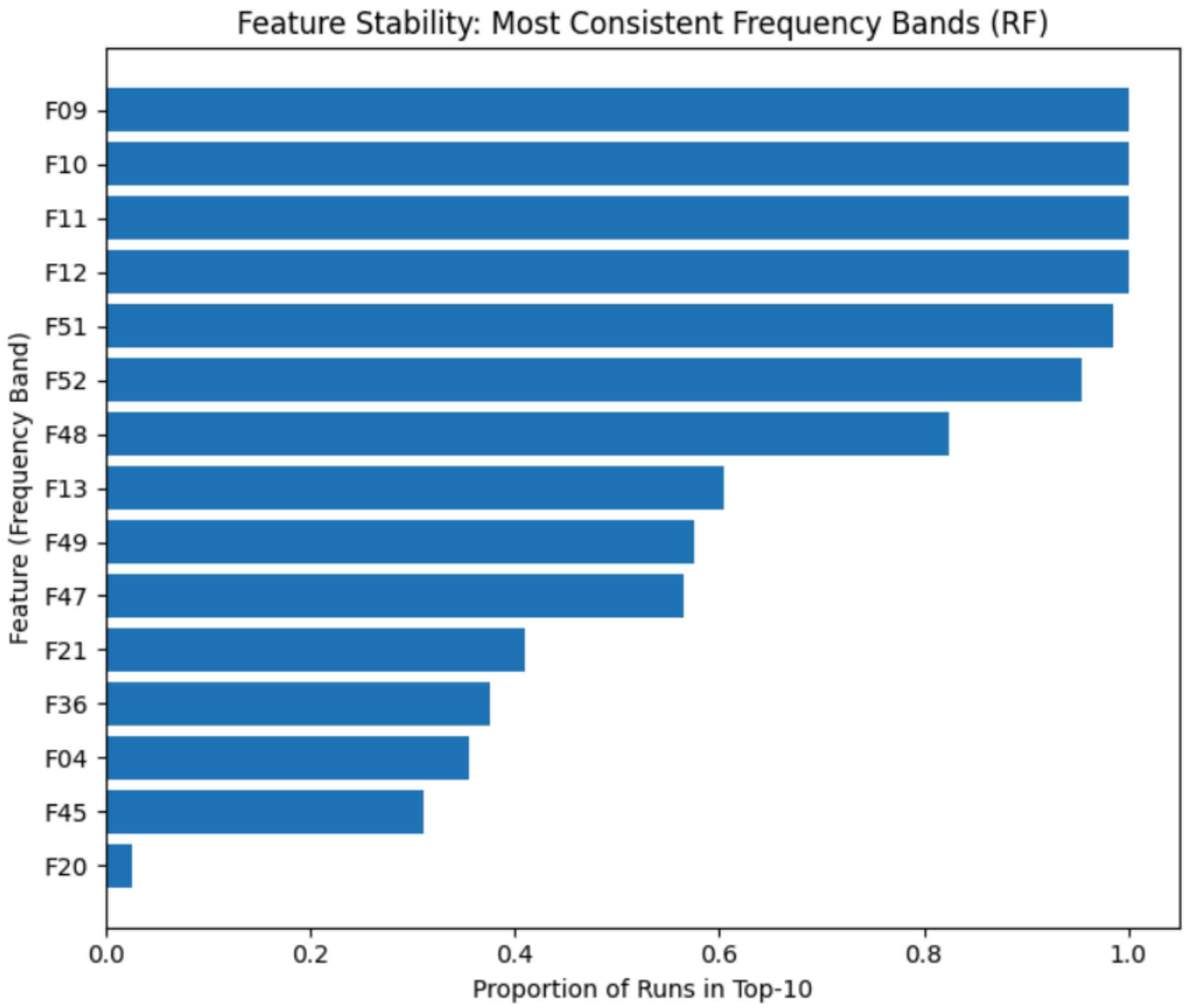

Fig 5 : Feature stability analysis showing dominant sonar frequency bands.

10. Discussion

The results show that data science methods improve sonar-based mine detection but they do not solve all the problems which physical sensors encounter. The SVM model successfully learns nonlinear decision boundaries but the overlapping acoustic signatures make it impossible to achieve complete separation between object classes. The system shows a preference for detecting mines on the conservative side because its main goal is to prevent missing any mines even though it generates some incorrect alerts.

11. Limitations

The research faces two main limitations because it works with a small data set and uses preprocessed frequency-band features instead of analyzing the original sonar signals. The research lacks detailed information about environmental conditions and sensor setup parameters which could affect the results' applicability to different situations.

12. Future Work

Research in the future needs to study extensive datasets which need direct signal measurement to determine how time variables and environmental conditions affect experimental results. The combination of adaptive decision thresholds with multi-sensor fusion systems will result in improved real-world applications because it will enhance detection reliability.

13. Conclusion

The research used a full data science workflow to assess sonar sensor capabilities for detecting underwater mines. The predictive results of machine learning models remained strong but the built-in similarity between sonar responses made it impossible to achieve better than chance accuracy in classification. The research shows that sensor-based detection systems require predictive modeling to function properly with interpretive analysis for their correct evaluation.

Appendix A: Prediction Confidence Analysis

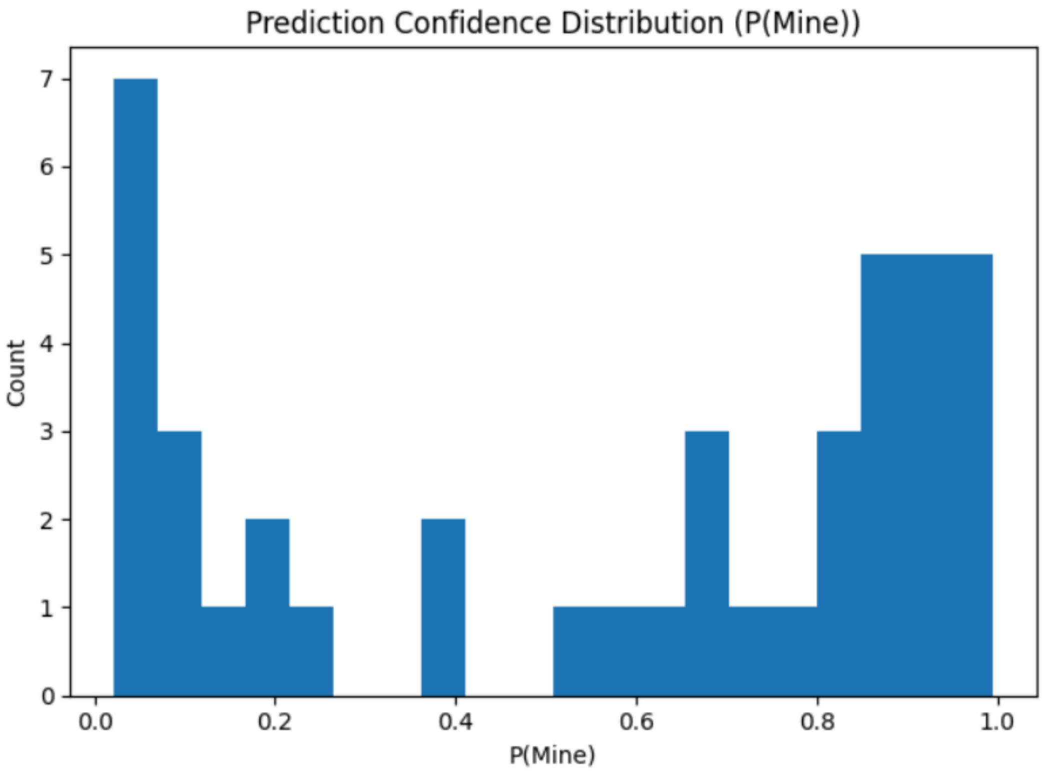

Fig A1 : Prediction confidence distribution showing regions of high certainty and uncertainty in mine detection.

Appendix B: SVM Decision Boundary Visualization in PCA Space

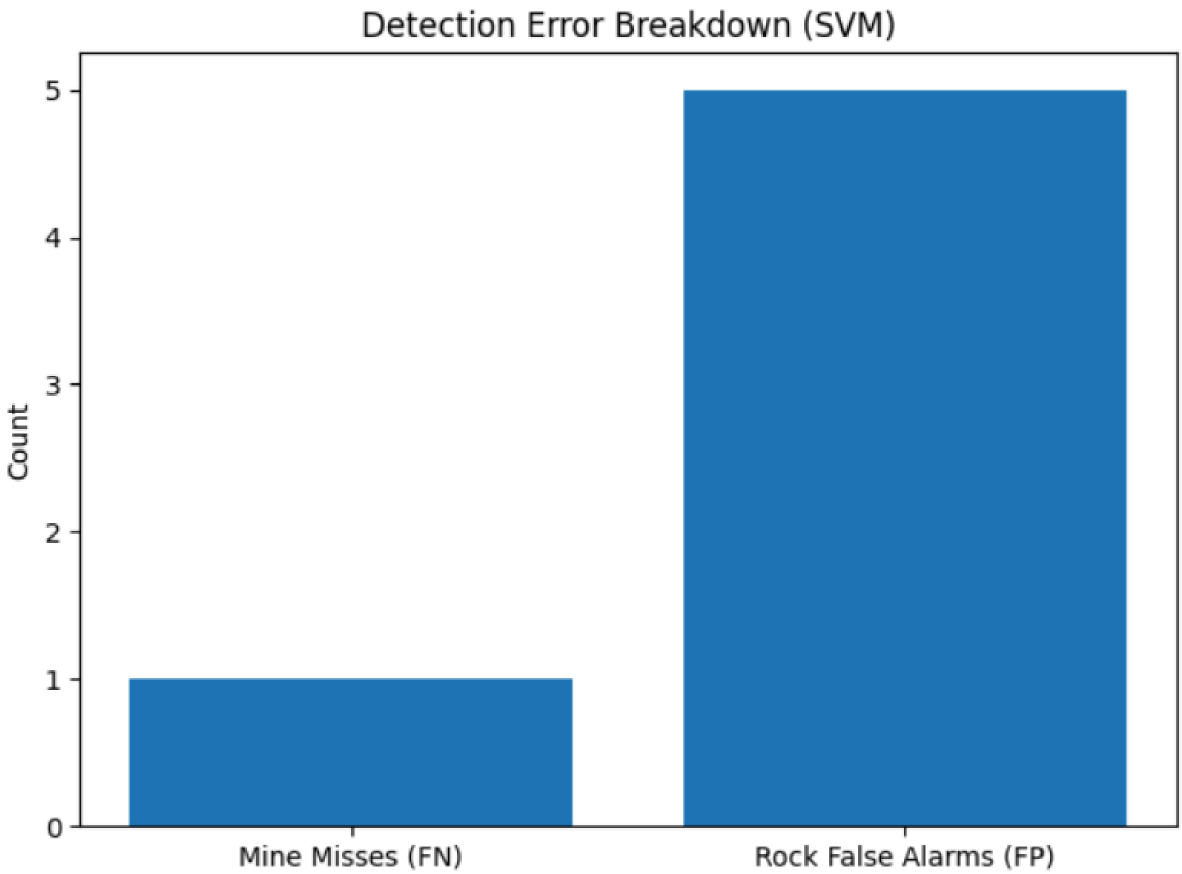

Fig A2 : SVM decision boundary in PCA space illustrating partial separation between mine and rock samples.

GITHUB REPOSITORY LINK

https://github.com/SRIHARIJTNB/Data-Science-Project-Sonar-Sensor-Performance-Analysis-for-Underwater-Mine-Detection-.git