Rock vs. Mine Prediction

Aditya Chauhan

Bachelors *of Technology in*
Computer Science Engineering
Graphic Era Hill University
Dehradun, India

rkshadityachauhan@gmail.com

ABSTRACT Usually, mines are mistaken as rocks during their identification, as mines can have the same shape, length, and width as rocks. To avoid this confusion, it is better to use a more accurate input to receive an accurate output. One of the methods in detecting the mines is SONAR. 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. To have a great accuracy we need accurate data to generate accurate results. I worked on the data set which is provided by Gorman, R. P., and Sejnowski, T. J. (1988). The data is used to train the machine. This paper presents a method for the prediction of underwater mines and rocks using Sonar signals. Sonar signals are used to record the various frequencies of underwater objects at 60 different angles. We constructed three binary classifier models according to their accuracy. Then, prediction models are used to predict the mine and rock categories. Python and Supervised Machine Learning Classification algorithms are used to construct these prediction models.

Keywords: Underwater Mines, SONAR, Rock vs. Mine, Supervised Machine Learning, Classification Algorithms, Prediction Model.

I. INTRODUCTION:

The vast expanse of the Earth's oceans holds a wealth of natural resources, including valuable rocks and minerals. Traditionally, the identification and classification of these underwater resources relied heavily on the expertise of geologists and manual interpretation of geological data. However, with the advent of advanced technology, particularly SONAR (Sound Navigation And Ranging), and the availability of large-scale geospatial data, data-driven approaches and machine learning techniques have emerged as powerful tools for automating and enhancing the accuracy of rock vs. mine prediction.

This paper delves into the exploration of methods, challenges, and outcomes associated with employing machine learning and geospatial data for predicting geological features. The primary objectives of this research are as follows:

- Developing Predictive Models: The aim is to create machine learning models capable of accurately classifying rock formations and mines.
- Evaluating Machine Learning Algorithms: The performance of various machine learning algorithms will be assessed in the context of rock vs. mine prediction.
- Exploring Real-World Applications: The potential applications of these predictive models in real-world scenarios will be discussed.

Data-Driven Approaches to Rock vs. Mine Prediction

The integration of machine learning and geospatial data offers several advantages over traditional methods of rock vs. mine prediction:

- Automation: Machine learning algorithms can automate the classification process, reducing reliance on manual interpretation and human expertise.

- Accuracy: Machine learning models can achieve higher levels of accuracy compared to traditional methods, particularly when trained on large datasets.

- Efficiency: Machine learning algorithms can process large amounts of data efficiently, making them suitable for real-time applications.

.

II. LITERATURE SURVEY:

To accurately classify objects as either rocks or mines, a predictive system was developed utilizing machine learning techniques. The system employed a dataset from a study by R. Paul Gorman and Terrence J. Sejnowski, which involved SONAR trials in a simulated region with metal cylinders representing mines. The objects were struck with sonar signals from various angles, and the results were recorded. This dataset was used to train three binary classifier models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression.

A. KNN Algorithm

The KNN algorithm works in a similar way. It classifies new data points based on the majority class of their k nearest neighbors in the training set. In other words, it looks for the k data points in the training set that are most similar to the new data point and assigns the new data point to the class that is most common among those k neighbors.
The value of k, known as the "k-nearest neighbors" parameter, is a hyper parameter that needs to be determined before using the KNN algorithm. A higher value of k will result in a smoother decision boundary, but it may also make the algorithm more prone to overfitting. A lower value of k will result in a more jagged decision boundary, but it may also make the algorithm more sensitive to noise in the data.
The KNN algorithm is a simple and versatile algorithm that can be used for a variety of classification tasks. It is particularly well-suited for tasks where the data is high-dimensional and there is no clear linear relationship between the features and the classes.
Using the train test split() method, we split the data into training and testing data. We go for the most appropriate distance measure. The k value, on the other hand, must be calculated.
The k value represents the number of nearest neighbors considered for classification. Here I used k value 3.

B. SVM Algorithm

The SVM algorithm finds the best hyperplane that separates the two categories of data points with the widest margin. The margin is the distance between the hyperplane and the closest data points from each category. A wider margin means that the hyperplane is more likely to correctly classify new data points. To find the best hyperplane, the SVM algorithm focuses on the data points that are closest to the hyperplane, called support vectors. The SVM algorithm only considers these support vectors when calculating the hyperplane, which makes the algorithm more efficient and less prone to making mistakes. Once the best hyperplane has been found, the SVM algorithm can be used to classify new data points. If a new data point falls on one side of the hyperplane, it is classified as one category. If it falls on the other side, it is classified as the other category.
The optimal hyper parameters for the SVM model were determined using a grid search approach. The optimal value for the c parameter was found to be 1.5.

C. Logistic Regression

Logistic regression is a statistical method that finds the best decision boundary for separating the two categories of data points. It does this by estimating the probability that each data point belongs to one category or the other. The data points with a higher probability of belonging to one category are classified as that category.

The logistic regression model is based on the logistic function, which is a sigmoid function that squashes its input to a value between 0 and 1. The logistic function represents the probability of an object belonging to one category. The probability of belonging to the other category is simply 1 minus the probability of belonging to the first category.

The logistic regression model learns the relationship between the input features and the target variable (the class label) by estimating the coefficients of the logistic function. These coefficients represent the strength of the association between each input feature and the probability of belonging to one category.

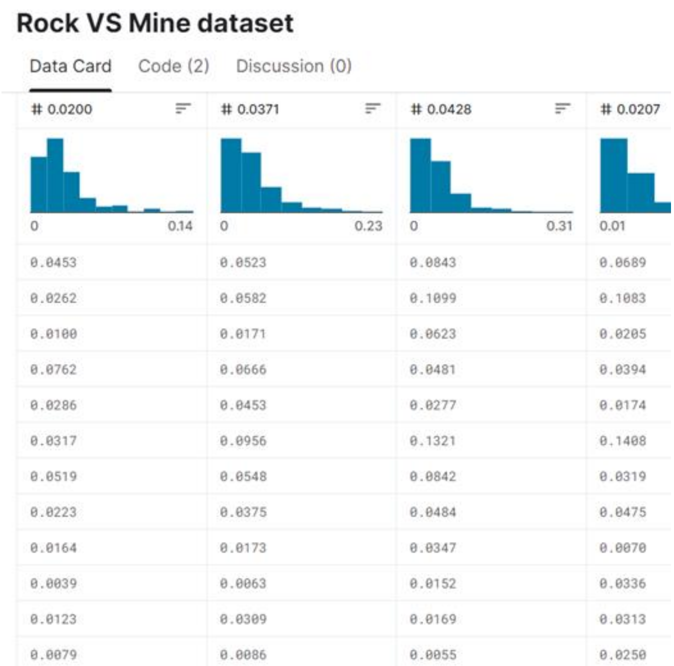
Logistic regression is a statistical method that predicts the probability of an object belonging to a particular class. The optimal solver for the logistic regression model was found to be the bilinear solver.

D. Evaluation and Prediction

The performance of the three classifiers was evaluated using a confusion matrix, classification error, and precision scores. The KNN model exhibited the highest accuracy, followed by the SVM and logistic regression models. The developed prediction system can be used to classify new objects based on their sonar signal frequencies.

III. DATASET:

To have a great accuracy we need accurate data to generate accurate results. I worked on the data set which is taken from kaggle website(https://www.kaggle.com/datasets/deepikaarikesavan/rock-vs-mine-dataset).



Rock VS Mine dataset

Data Card Code (2) Discussion (0)

# 0.0200	# 0.0371	# 0.0428	# 0.0207
0 – 0.14	0 – 0.23	0 – 0.31	0.01
0.0453	0.0523	0.0843	0.0689
0.0262	0.0582	0.1099	0.1083
0.0100	0.0171	0.0623	0.0205
0.0762	0.0666	0.0481	0.0394
0.0286	0.0453	0.0277	0.0174
0.0317	0.0956	0.1321	0.1408
0.0519	0.0548	0.0842	0.0319
0.0223	0.0375	0.0484	0.0475
0.0164	0.0173	0.0347	0.0070
0.0039	0.0063	0.0152	0.0336
0.0123	0.0309	0.0169	0.0313
0.0079	0.0086	0.0055	0.0250

IV. METHODOLGY:

Step 1: Data Preparation and Exploratory Data Analysis

Before diving into the machine learning process, it's crucial to ensure the data is clean and ready for analysis. This involves handling missing values, outliers, and inconsistencies in the dataset. Exploratory Data Analysis (EDA) helps understand the data's characteristics, patterns, and relationships between variables. EDA techniques like data visualization and statistical analysis provide insights into the data distribution, trends, and potential issues.

Step 2: Splitting Data into Train and Test Sets

To evaluate the performance of machine learning models, we split the data into two sets: training and testing. The training set is used to train the models, while the testing set is used to assess their generalization ability on unseen data. This split ensures that the models are not simply memorizing the training data and can perform well on new data.

Step 3: Selecting and Evaluating Top-Performing Models

Based on the evaluation results, we identify the top three performing models: KNN, SVM, and Logistic Regression. These models demonstrate superior accuracy and effectiveness in classifying rocks and mines.

Step 4: Accuracy Evaluation and Classification Report

We further evaluate the accuracy of the selected models by calculating their classification metrics, such as precision, recall, and F1-score. These metrics provide a comprehensive understanding of the models' ability to correctly classify rocks and mines, considering both true positives and false positives/negatives. A classification report summarizes the performance of each model, providing detailed insights into their classification accuracy.

Step 5: Model Fitting for Prediction System

We fit the selected models using the training data, refining their parameters to optimize their performance. This step allows us to create a prediction system that can accurately classify new objects as rocks or mines based on their sonar signal frequencies.

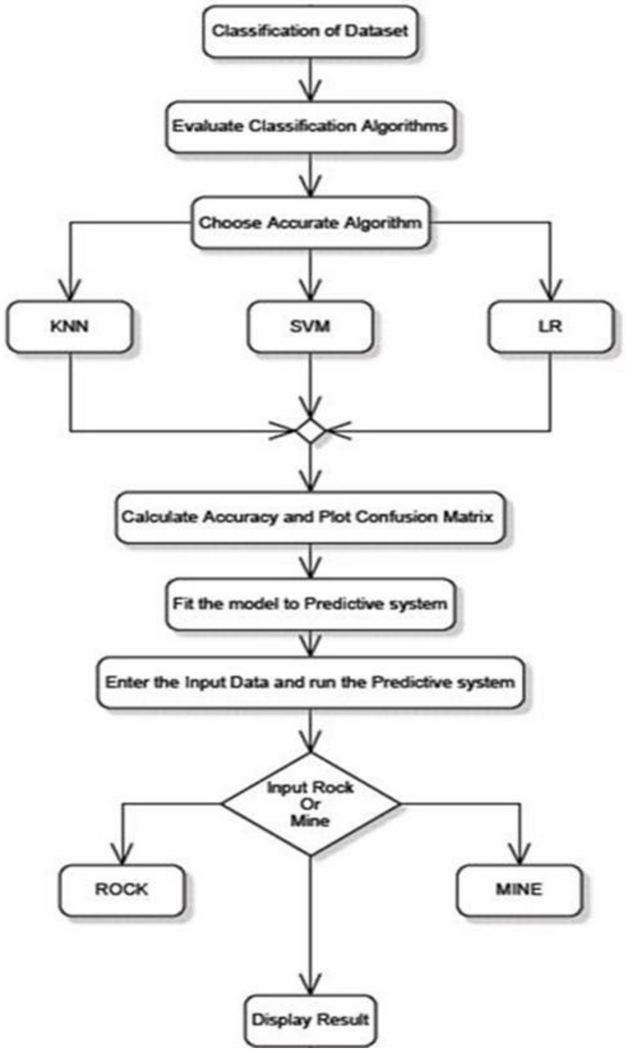
Step 6: Object Classification and Prediction System Usage

With the prediction system in place, we can now feed new sonar signal frequency data into the system. The system will utilize the trained models to classify the object as either a rock or a mine. This allows us to make accurate predictions without the need for manual inspection or further analysis.

Choosing an ML Model: Considerations and Performance Evaluation

Selecting the most appropriate machine learning model for a given task requires careful consideration of various factors, including performance, understandability, complexity, dataset size and dimensionality, and interface time. Evaluating model performance is crucial before making a final decision. Model assessment techniques and evaluation measures, such as classification metrics, help us compare the performance of different models and choose the one that best suits our needs.

In this case, we have chosen KNN, SVM, and Logistic Regression as the top-performing models based on their accuracy and classification metrics. These models provide a balance of performance, understandability, and complexity, making them suitable for our rock vs. mine classification problem.



V. RESULTS:

https://colab.research.google.com/drive/1uobbCAKzyZM9AZkBz0VtMGobwwYRkej9?usp=sharing

```
Applying logistics regression:
Accuracy on training data:
0.8342245989304813 or 83.4%

Accuracy on test data:
0.7619047619047619 or 76.19%
```

```
SVM accuracy on test data: 0.8528125
or 85.28%

KNN accuracy on test data: 0.8125 or
81.25%
```

VI. FUTURE SCOPE OR IMPLICATION:

The successful implementation of machine learning for rock vs. mine prediction can have significant implications:

- Improved resource exploration: Accurately identifying rock formations and mines can streamline underwater resource exploration and extraction.
- Enhanced environmental protection: By reducing the need for extensive physical surveys, machine learning can minimize environmental disruption.
- Cost-effective operations: Automated classification can reduce operational costs and improve efficiency in underwater resource exploration and management.

VII. CONCLUSION:

Naval mines pose a significant threat to underwater navigation, hindering maritime operations and causing substantial economic and environmental damage. Conventional mine detection methods, often relying on sonar signals or manual inspection, can be time-consuming, expensive, and risky for personnel.
Our project, titled "Underwater mine and rock prediction by the evaluation of machine learning algorithms," aims to address this challenge by developing an advanced prediction system utilizing machine learning techniques. The system employs sonar signal data to accurately distinguish between rocks and mines on the ocean floor.
The project leverages the power of Python, an open-source programming language, to implement machine learning algorithms and analyze sonar signal data. Python's computational efficiency and cost-effectiveness make it an ideal choice for this application.
By evaluating various machine learning algorithms, we can identify and compare their performance metrics, such as accuracy, precision, and recall. This evaluation process enables us to select the best-performing algorithm for our prediction system, ensuring optimal detection accuracy and minimizing false positives.
Our project aims to simplify and streamline the

underwater mine detection process, enhancing safety and efficiency for maritime operations. By leveraging machine learning, we can significantly reduce the reliance on risky manual inspections and improve the overall effectiveness of mine detection efforts.

REFERENCES

1) Dura, Esther, et al. "Active learning for detection of mine-like objects in side-scan sonar imagery." IEEE Journal of Oceanic Engineering 30.2: 360-371 (2005).

2) 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).

3) 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).

4) N. Hooda et al. “B 2 FSE framework for high deimensional imbalanced data: A case

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

6) 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).

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

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

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

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

10) 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).