Project Report: Sonar Rock vs Mine Detection

Introduction

The aim of this project is to develop a machine learning model that can accurately classify sonar signals as either rocks (R) or mines (M). This classification is crucial for submarine navigation, ensuring safe passage by detecting potential underwater hazards.

Dataset

The dataset contains 208 samples, each with 60 features representing sonar readings. The labels, located in the 61st column, indicate whether the sample is a rock or a mine.

Dataset Characteristics

- **Total Samples:** 208
- **Features per Sample:** 60
- **Labels:** 111 Mines (M), 97 Rocks (R)

Sample Data

The first few rows of the dataset are as follows:

	0	1	2	...	58	59	60
0	0.0200	0.0371	0.0428	...	0.0090	0.0032	R
1	0.0453	0.0523	0.0843	...	0.0052	0.0044	R
2	0.0262	0.0582	0.1099	...	0.0095	0.0078	R
3	0.0100	0.0171	0.0623	...	0.0040	0.0117	R
4	0.0762	0.0666	0.0481	...	0.0107	0.0094	R

Statistical Summary

The statistical summary of the dataset provides insights into the distribution of values across the features.

	0	1	2	...	58	59	60
count	208.000000	208.000000	208.000000	...	208.000000	208.000000	208
mean	0.029164	0.038043	0.043370	...	0.013165	0.006940	
std	0.023119	0.032130	0.038676	...	0.018222	0.012409	
min	0.000600	0.000200	0.000300	...	0.000600	0.000200	
25%	0.013700	0.014925	0.017400	...	0.004500	0.001500	
50%	0.021800	0.030850	0.035300	...	0.008300	0.003900	
75%	0.035825	0.048025	0.056050	...	0.015000	0.007600	
max	0.137100	0.233900	0.305900	...	0.120500	0.137100	

Class Distribution

- **Mines (M):** 111
- **Rocks (R):** 97

The dataset is relatively balanced, which is advantageous for model training as it avoids bias towards one class.

Mean Values Analysis

Mean values of the features for each class (R and M) provide insight into the general differences between the sonar readings for rocks and mines. The first few features are shown below:

Feature	Mean (Rocks)	Mean (Mines)
0	0.031778	0.026455
1	0.045368	0.032684
2	0.048702	0.039195
...	...	...
58	0.016847	0.010128
59	0.011501	0.002669

Data Preparation

1. **Feature and Label Separation:**
    - Features (`X`): All columns except the last one.
    - Labels (`Y`): The last column.

python

```
X = sonar_data.drop(columns=60, axis=1)
Y = sonar_data[60]
```

2.
3. **Training and Testing Split:**
    - **Training Set:** 90% of the data (187 samples).
    - **Testing Set:** 10% of the data (21 samples).
    - **Stratification:** Ensures an equal distribution of labels in both sets.

python

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y,
test_size=0.1, stratify=Y, random_state=1)
```

Model Development

Logistic Regression

Model Initialization:
python

```
model = LogisticRegression()
```

1.

Model Training:
python

```
model.fit(X_train, Y_train)
```

2.

Model Evaluation

Training Data Accuracy:
python

```
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```

1.
    - **Accuracy:** ~83.4%

Testing Data Accuracy:
python

```
testing_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

2.
    - **Accuracy:** ~76.2%

The model shows strong performance on both training and testing datasets, with slightly lower accuracy on the testing set, which is typical as it contains unseen data.

Predictive System

The logistic regression model is used to make predictions on new sonar readings. For example, given the following input data:

python

```
input_data =
(0.0307,0.0523,0.0653,0.0521,0.0611,0.0577,0.0665,0.0664,0.1460,0.2792
,0.3877,0.4992,0.4981,0.4972,0.5607,0.7339,0.8230,0.9173,0.9975,0.9911
,0.8240,0.6498,0.5980,0.4862,0.3150,0.1543,0.0989,0.0284,0.1008,0.2636
,0.2694,0.2930,0.2925,0.3998,0.3660,0.3172,0.4609,0.4374,0.1820,0.3376
,0.6202,0.4448,0.1863,0.1420,0.0589,0.0576,0.0672,0.0269,0.0245,0.0190
,0.0063,0.0321,0.0189,0.0137,0.0277,0.0152,0.0052,0.0121,0.0124,0.0055
)
input_data_as_numpy_array = np.asarray(input_data)
input_data_reshaped = input_data_as_numpy_array.reshape(1, -1)
prediction = model.predict(input_data_reshaped)
```

- **Output:** `'M'`
- **Conclusion:** The object is a mine.

Conclusion

This project successfully demonstrates the application of logistic regression for classifying sonar signals into rocks and mines with good accuracy. The balanced dataset, proper data splitting, and evaluation methods ensure the robustness of the model. Future improvements could involve experimenting with other algorithms and tuning model parameters for potentially better performance.