Land Mine Type Prediction

# Iffat Tasnim(2020-2-60-004),Rakibul Islam Joy(2020-2-60-035), Zin Din Ziden(2020-2-60-132), Pulok Islam(2020-2-60-193)

*Department of Computer Science and Engineering, East West University Aftabnagar, Dhaka-1212, Bangladesh*

# [2020-2-60-004@std.ewubd.edu](mailto:2020-2-60-004@std.ewubd.edu) [2020-2-60-035@std.ewubd.edu](mailto:2020-2-60-035@std.ewubd.edu) [2020-2-60-132@std.ewubd.edu](mailto:2020-2-60-132@std.ewubd.edu) [2020-2-60-193@std.ewubd.edu](mailto:2020-2-60-193@std.ewubd.edu)

***Abstract*—*Abstract-* There are many categories of land mine. There are many types of land mines such as Anti-personnel and Anti-vehicle mines. Anti- personnel landmines are designed to explode when as little as two kilograms of pressure is applied - or when a person steps on them or disturbs them. Anti- vehicle landmines are designed to explode when at least 200 kilograms of pressure is applied - or when a car, jeep, truck, or tank drives over them. There are many types of land mines: Fragmentation and stake mines, Directional mines, Blast mines, Bounding mines, Flame mines, Chemical mines, Blast mines, Full-width mines, Side attack mines, Wide area mines, Anti-helicopter mines, Nuclear land mines. In this study, we propose to apply different machine learning techniques to our dataset, including Lin- earRegression, Logistic Regression(LR), Ada Boost, Support Vector Machines(SVM), Decision Tree(DT), Random Forest(RF), SVC, confusion matrix, MLP, PCA achieved through machine learning approaches are contrasted with one another. In our study, we see that in the methods Support Vector Machine(SVM) gives the best accuracy result than other methods. We also use Minimum redundancy maximum rele- vance(MRMR) for feature selection and from this we test and train the data and split the data. By using confusion matrix, we can get a clear result of our predicted result.**

***Index Terms*—SVM, SVC, LogisticRegression, De- cisionTree, RandomForest, AdaBoost, GaussianNB, confusion matrix, Naive Bayes, MLP, PCA**

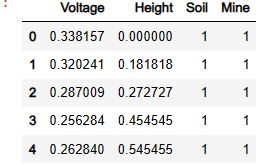
1. INTRODUCTION

**T**

HE advancement of technology has improved our quality of life while also posing certain challenges. The advantage of our technology is

that it makes it simple day by day to get a variety of facts that we need. It is necessary to process the easily accessible raw data from online sources for it to be useful. In this context, feature engineering techniques and machine learning techniques are crucial to the process.

The object of this study is to use machine learning and feature engineering techniques to obtain the most accurate results from the raw and missing data. The land mine dataset is used here for the purpose. It is acknowledged that the dataset had no missing values. This dataset has 338 instances and 4 attributes. This dataset is taken from the military defense area. The soil category here is dry sandy, dry humus, dry limy, humid sandy, humid humus, humid limy, and voltage(0-0.6v) and height 0 to 20 cm from the FLC sensor. Land Mine types are 1. Not-Mine, 2. Anti-tank, 3. Anti-personnel, 4. Booby-Trapped-Anti-personel, 5. M14-Anti-personnel



1. RELATED WORKS

The study of machine Mine detection using Ma- chine Learning has generated significant attention from researchers, resulting in numerous. Research has been conducted in this field. Here are reviews of some of them:

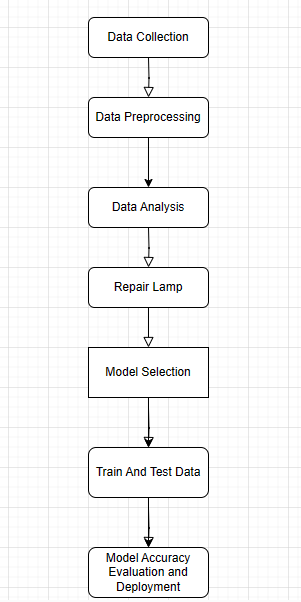
Safatly[1] used multiple logistic regression meth- ods. He provided performance figures for several cases. In his research Bagging algorithm showed the best-performing algorithm with the highest number of accuracy results which was 100% per class.

Silvia[2] presented a pipeline for landmine de- tection based on his analysis. They proposed an approach which is based on the use of convolution neural networks(CNN). Validation showed that the system provided 95% accuracy.

Achkar[3] used The Black Propagation algorithm of multilayer perception(MLP) in the Artificial Neural Network in his dataset. The system proved to be able to identify and classify different types of

landmines under various conditions with a success rate of up to 90%.

1. METHODOLOGY



Data Collection: the military dataset was collected from Kaggle which was donated by authentic resource so the data distribution was good enough for ML models.

Data Preprocessing: We checked the data to find if there was any missing values and outlier in any row or column. Then we see the distribution of data via data analysis to find which techniques will be good for the data. Then we took 65% data for training and finally we scaled the data using StandardScaler to set the mean of each data close to 0 and std deviation close to 1 for good accuracy.

Data Analysis: here we have find the correlation of each feature to find if there is dependency between them to create a model. Then we checked the distribution of each feature to find if they are normally distributed using seaborn and finally we drew the boxplots also to check outliers so that we can use scaling. In the end we just drew how best fitted model works to get decisions.

Dat

### Logistic Regression:

A supervised learning classification approach called logistic regression is used to forecast the likelihood of a target variable. One of the simplest Ml algorithms, it can be applied to a variety of classification issues, including spam filtering, cancer diagnosis, and other issues.

### Random Forest Classifier:

A popular supervised machine learning ap- proach for classification and regression issues is the Random Forest Classifier. On various samples, it constructs decision trees and uses their average for classification. This classifier can both handle continuous variables and categorical variables.

### Support Vector Machine(SVM):

A machine learning algorithm called Support Vector Machine examines data for categorization and regression analysis. It examines data and clas- sifies it into one or two groups.

### Naive Bayes:

Naive Bayes is a supervised machine learning algorithm primarily used for classification tasks. It is based on Bayes’ theorem and assumes inde- pendence among features, which simplifies calcu- lations. There are different types of Naive Bayes classifiers, including Gaussian, Multinomial, and Bernoulli, suitable for various types of data. The algorithm estimates probabilities from training data and is commonly applied to text classifi- cation, spam filtering, and other categorization problems. Naive Bayes often performs well with high-dimensional datasets and requires minimal training data.

### Decision Tree:

Among the most popular classifiers is the deci- sion tree because of its relatively easy-to-construct structure. A decision tree is a model compromis- ing prediction and decision nodes that are tree- structured.

### Ada Boost:

One of the most popular and successful ensem- ble learning techniques is AdaBoost. The funda- mental idea behind AdaBoost is that by linearly adding several weak classifiers, a strong classifier can be produced. AdaBoost increases the weights of the incorrectly categorized data points and de- creases the weights of successfully classified data points during the training phase.

1. EXPERIMENTS

### Dataset:

Kaggle contributed to the Land Mine dataset. Detection of mines buried in the ground is very important in terms of the safety of life and prop- erty. Many different methods have been used in this regard; however, it has not yet been possible to achieve 100% success. Mine detection process consists of sensor design, data analysis, and de- cision algorithm phases. The magnetic anomaly method works according to the principle of mea- suring the anomalies resulting from the object in the magnetic field that disturbs the structure of it, the magnetic field, and the data obtained at this point are used to determine the conditions such as motion and position. The determination of parameters such as position, depth, or direction of motion using magnetic anomaly has been carried out since 1970. This dataset has 338 instances and 4 attributes. This dataset is taken from the military defense area. The soil category here is dry sandy, dry humus, dry limy, humid sandy, humid humus, humid limy, and voltage(0-0.6v) and height 0 to 20 cm from the FLC sensor.

Table 1

Number of features of the dataset

|  |  |  |
| --- | --- | --- |
|  | | |
| Feature | Value of Feature | Feature Characteristic |
| Voltage | 0 to 1 | Float |
| Height | 0 to 1 | Float |
| Soil | 0 to 6 | Integer |
| Mine | 0 to 5 | Integer |

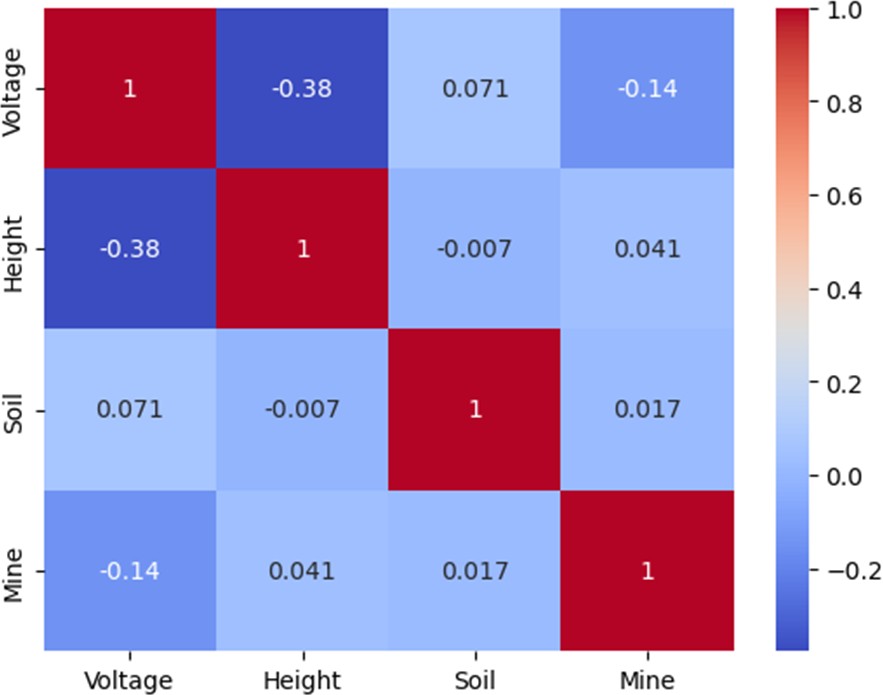
Data Analysis:

After reading the dataset ”Land mines” we calculated the summary where the dataset contains 338 rows and 4 columns.

Correlation between data:

Correlation, which is frequently used as a preliminary strategy to uncover correlations

between variables in machine learning classification tasks and it may be a key to increasing the accuracy of a prediction model. To determine how independent features affect intuitive predicting, classification algorithms might employ the positive or negative correlation between feature values. In figure 1, the correlation coefficient ranges from -1 to 1, indicating the strength and direction of the linear relationship between the two variables. Here’s a brief interpretation of the correlation we’ve provided: Voltage and Voltage (Self-correlation): The



correlation (0.017346) between Soil and Mine. Mine and Voltage: Same as the correlation between Voltage and Mine (-0.144569).

Mine and Height: Same as the correlation between Height and Mine (0.041326).

Mine and Soil: Same as the correlation between Soil and Mine (0.017346).

Mine and Mine (Self-correlation): The correlation of a variable with itself is always 1.0.

The correlation matrix provides information about the linear relationships between pairs of variables. Positive correlations indicate a positive linear relationship, negative correlations indicate a negative linear relationship and values close to zero indicate a weak or no linear relationship. Our correlation analysis does not imply causation, and the strength of the relationship may not capture non-linear associations or other complex patterns in the data.

**Voltage:**

This plot shows the distribution of values in the ’Voltage’ column of your ’mine’ dataset. It helps visualize the spread and central tendency of voltage values.

Fig. 1. Correlation between data

correlation of a variable with itself is always 1.0. Voltage and Height: There is a negative correlation (-0.377523) between Voltage and Height. As Voltage increases, Height tends to decrease, and vice versa.

Voltage and Soil: There is a weak positive correlation (0.070673) between Voltage and Soil. Voltage and Mine: There is a negative correlation (-0.144569) between Voltage and Mine.

Height and Voltage: Same as the correlation between Voltage and Height (-0.377523).

Height and Height (Self-correlation): The correlation of a variable with itself is always 1.0. Height and Soil: There is a very weak negative correlation (-0.006957) between Height and Soil. Height and Mine: There is a positive correlation (0.041326) between Height and Mine.

Soil and Voltage: Same as the correlation between Voltage and Soil (0.070673).

Soil and Height: Same as the correlation between Height and Soil (-0.006957).

Soil and Soil (Self-correlation): The correlation of a variable with itself is always 1.0.

Soil and Mine: There is a very weak positive

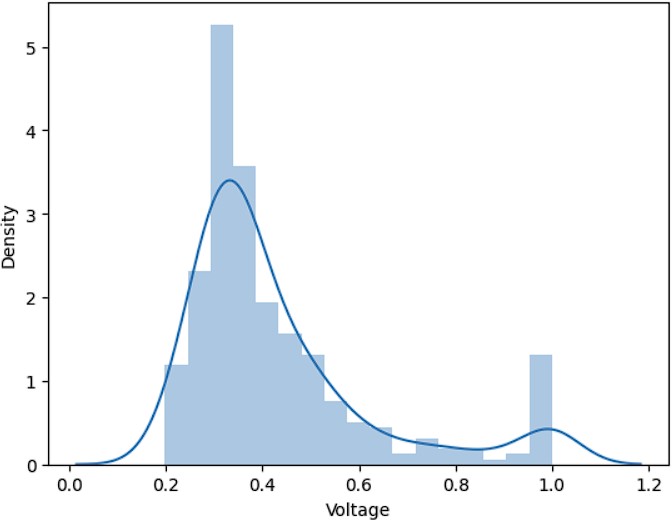


Fig. 2. Distribution of Voltage feature

## Height:

This plot represents the distribution of values in the ’Height’ column. It provides insights into the variability and central tendency of height values in your dataset. Ref. Fig. 3.

## Soil:

This plot displays the distribution of values in the ’Soil’ column. It helps you understand the distribution pattern of soil-related data in your dataset. Ref. Fig. 4.

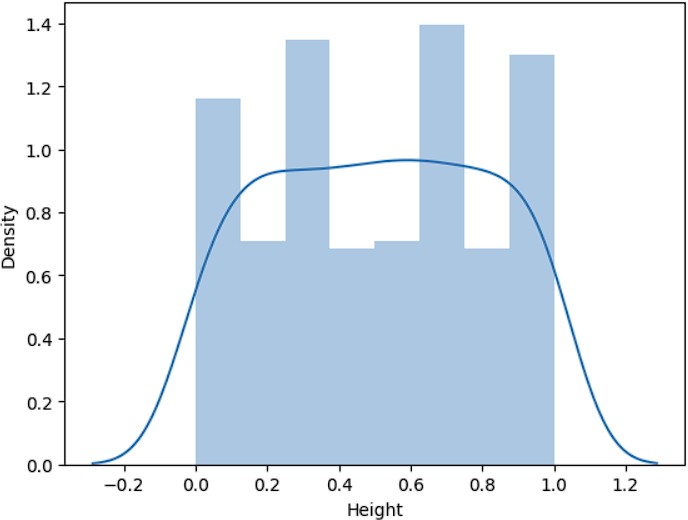


Fig. 3. Distribution of Height feature

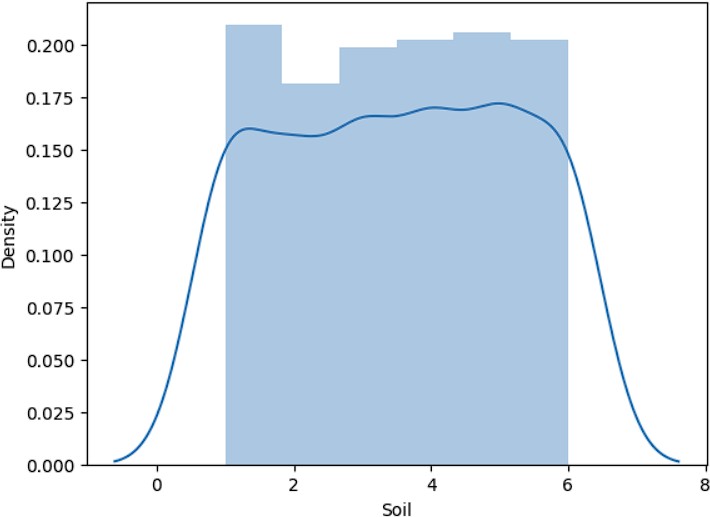


Fig. 4. Distribution of Soil feature

## Mine:

This plot shows the distribution of values in the ’Mine’ column. It visualizes the spread and central tendency of data related to mines. Ref. Fig. 5.

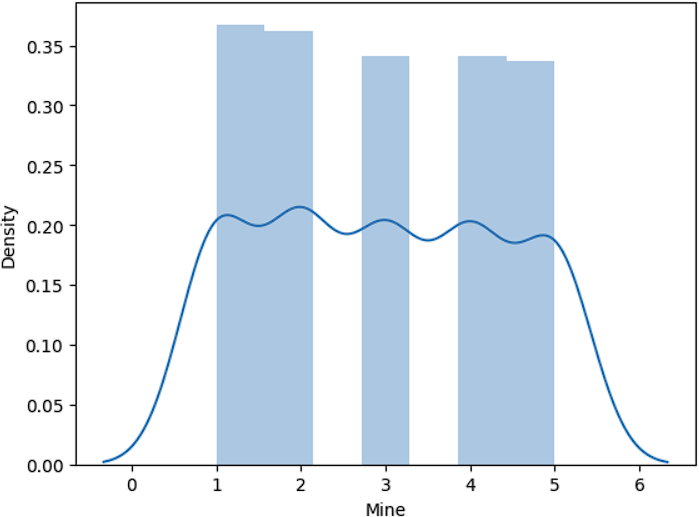


Fig. 5. Distribution of Mine feature

## Boxplot:

The boxplot represents the interquartile range (IQR) of the data, the line inside the box

represents the median, and the ”whiskers” extend to show the range of the data. Outliers may be plotted as individual points beyond the whiskers. We used Seaborn to create boxplots for different variables ’Voltage’, ’Height’, and ’Soil’ based on the categories defined by the ’Mine’ column in our dataset.

Boxplot for Voltage vs Mine: In this boxplot, the y-axis represents the ’Voltage’ values, and the x-axis represents the different categories defined by the ’Mine’ column. This visualizes the distribution of voltage values across different mines and identifies any variations or outliers.

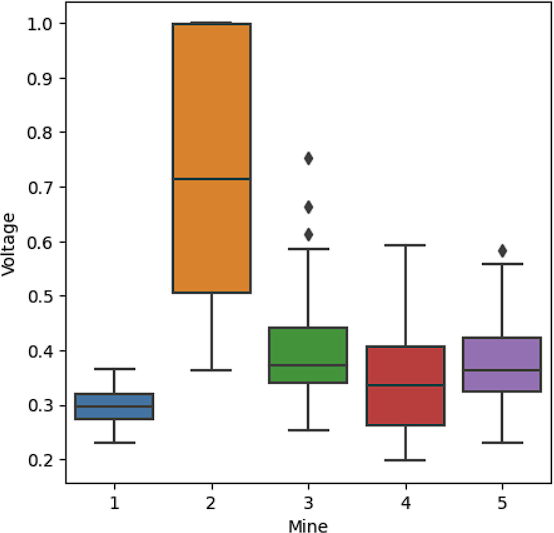


Fig. 6. Distribution of Voltage based on Mine feature

Boxplot for Height vs Mine: Similarly, this boxplot for the ’Height’ variable, shows how height values vary across different mines.

Boxplot for Soil vs Mine: This boxplot for the ’Soil’ variable, illustrates the distribution of soil-related data across different mines.

These boxplots help us compare the central tendency, spread, and presence of outliers for each variable across different mines in your dataset. This gives us a proper understanding of the distributional characteristics of the data and identifies potential differences or patterns among the mines.

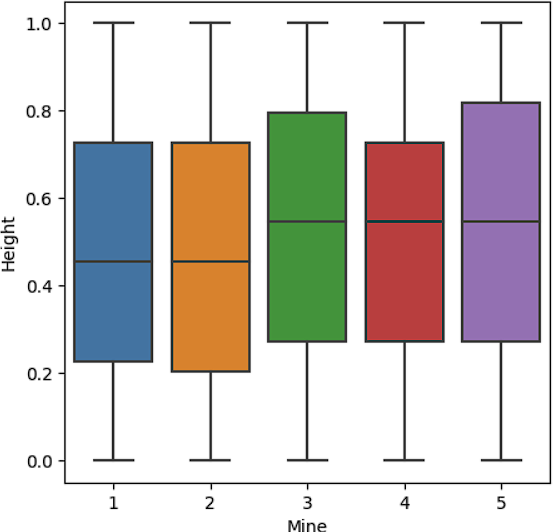


Fig. 7. Distribution of Height based on Mine feature

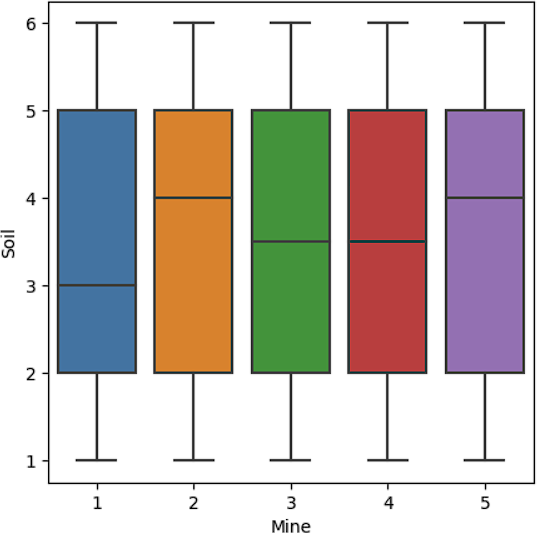


Fig. 8. Distribution of Soil based on Mine feature

### Prepocessing Steps:

Data collection, Data normalization, Data standardization, Feature selection, and Data representation are used as preparation processes in this study. We had no missing values so we did not have to fill in the missing values. We also had no unnecessary data so we did not have to reduce our data.

After importing libraries we read our data. We checked if we had a null value or not. We checked the mean and standard deviation. We used correlation to check if our dataset has dependencies or relationships among them. The target was whether we could apply any prediction model or not like is there any linear relationship. Then we used histograms to check

if the distributions were normal. Then we made a boxplot, checked outliers, and checked how was the distribution, and features. As our data was normal distributional data, we used standard scalar for preprocessing with the help of standard deviation. We did not use mean, max as our data had normal distribution. Then we put our first 225 values in the X train and the other 110 values in the X test. The same goes for the Y train and the Y test. After scaling we trained the model. We used the model SVC, logistic regression, etc. We got the best performance from SVC because of the rbf radial basis. Because of the kernels, gamma, and c values, SVC’s hyperplane’s value got the best fit. As it is non-linear that’s why rbf is used. Then we checked the accuracy in test and train cases. We also used Adaboost, a neural network but they did not give any good results. So those were not acceptable. SVM gave us the best result. Then we predicted again. We scaled the data with standard deviation and then trained it. We showed how the SVM SVC rbf kernel worked and drew a decision boundary. Then we saved it so that it could be deployed on the website.

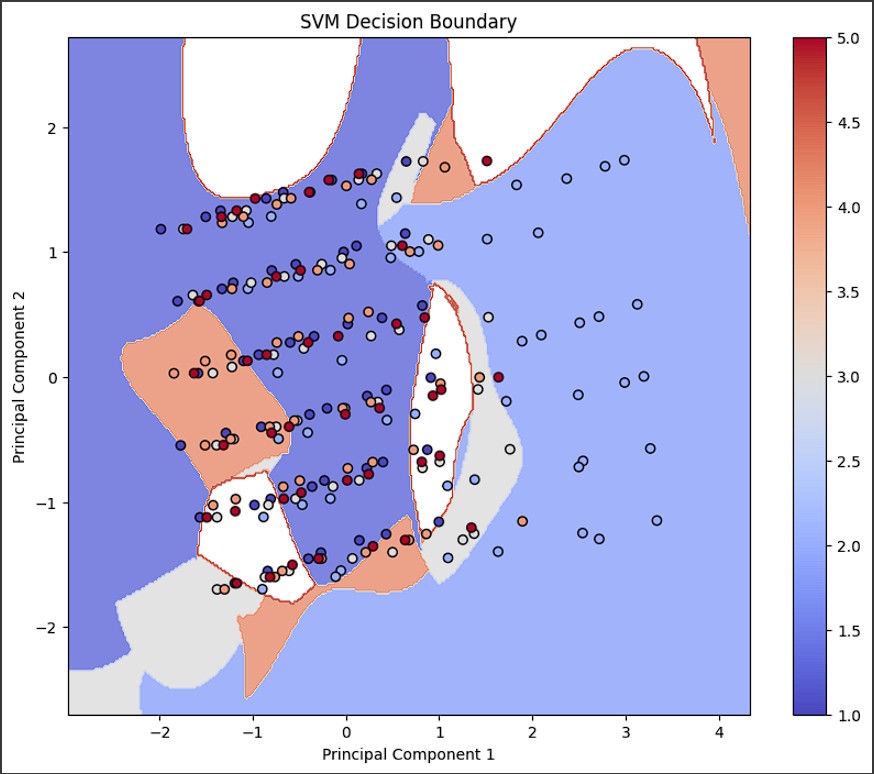


Fig. 9.

### Experimental Results

All algorithms are used to examine the mine characteristics. We discovered that further tweaks to several model parameters are needed to make the algorithm accurate when applying methods to the Land mines dataset.

Table 1

the Land Mine dataset to determine which features influence categorization outcomes.

REFERENCES

1. Safatly, L. Baydoun, Mohammed Alipour, M. Al Takach, Ali Atab, K. Husseini, Mohammed El-Hajj, A. Ghaziri, H.. (2020). Detection and classification of land- mines using machine learning applied to metal detector data. Journal of Experimental Theoretical Artificial Intel- ligence. 33. 1-24. 10.1080/0952813X.2020.1735529.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy of all model | | | | |
| Model | Train  Accuracy | Test  Accuracy | Preci-  sion | Recall |
| SVM | 84 | 79 | 80 | 79 |
| LogisticReg | 54 | 54 | 54 | 54 |
| Decision Tree | 100 | 62 | 63 | 62 |
| Random Forest | 99 | 53 | 53 | 53 |
| Adaboost | 43 | 43 | 51 | 45 |
| Naive Bayes | 47 | 49 | 54 | 49 |

1. S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi and S. Tubaro, ”Landmine detection from GPR data using con- volutional neural networks,” 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 2017, pp. 508-512, doi: 10.23919/EUSIPCO.2017.8081259.
2. Achkar, R., Owayjan, M., Mrad, C. (2011). Landmine Detection and Classification Using MLP. 2011 Third International Conference on Computational Intelligence, Modelling Simulation. https://doi.org/10.1109/CIMSIM.2011.10

Sample output:



1. DISCUSSION

In our dataset, we have three inputs and one output. The three inputs are Voltage, Height, and Soil type and the output is Mine type. Here Voltage, Height, and Soil type are independent variables and The mine type is our dependent variable. This means Mine type is dependent on the voltage, height, and soil type. Here the voltage is the output voltage value of the FLC sensor due to magnetic distortion. The height is the height of the sensor from the ground. The soil type is 6 different types of soil depending on the moist conditions: Dry and sandy, Dry and humus, dry and Limy, Humid and Sandy, Humid and Humus, and Humid and Limy. The Mine types are null(No mine), Anti-Tank, Anti-personnel, Booby-Trapped Anti-personnel and M14 Anti-personnel. Through this Analysis, we were able to pinpoint numerous crucial elements that had a big impact on our Mine dataset.

1. CONCLUSION

In a knowledge-based society, it is crucial to get useful results from incomplete and unprocessed data utilizing machine learning and feature en- gineering techniques. Here, we present models for determining land mine types from the data of sensors depending on inputs to help soldiers or miners. Our dataset created in the preparation step is classified using different machine learning algorithms in the classification step. This research compares machine learning methods to evaluate