ELECTRCICAL FAULT CLASSIFICATION AND DETECTION

INTRODUCTION

DISTRIBUTION OF FEATURES

DATA PREPOCESSING

EXPLORATORY DATA ANALYSIS

DATA VISUALIZATION

FEATURE ENGINEERING

MODEL SELECTION

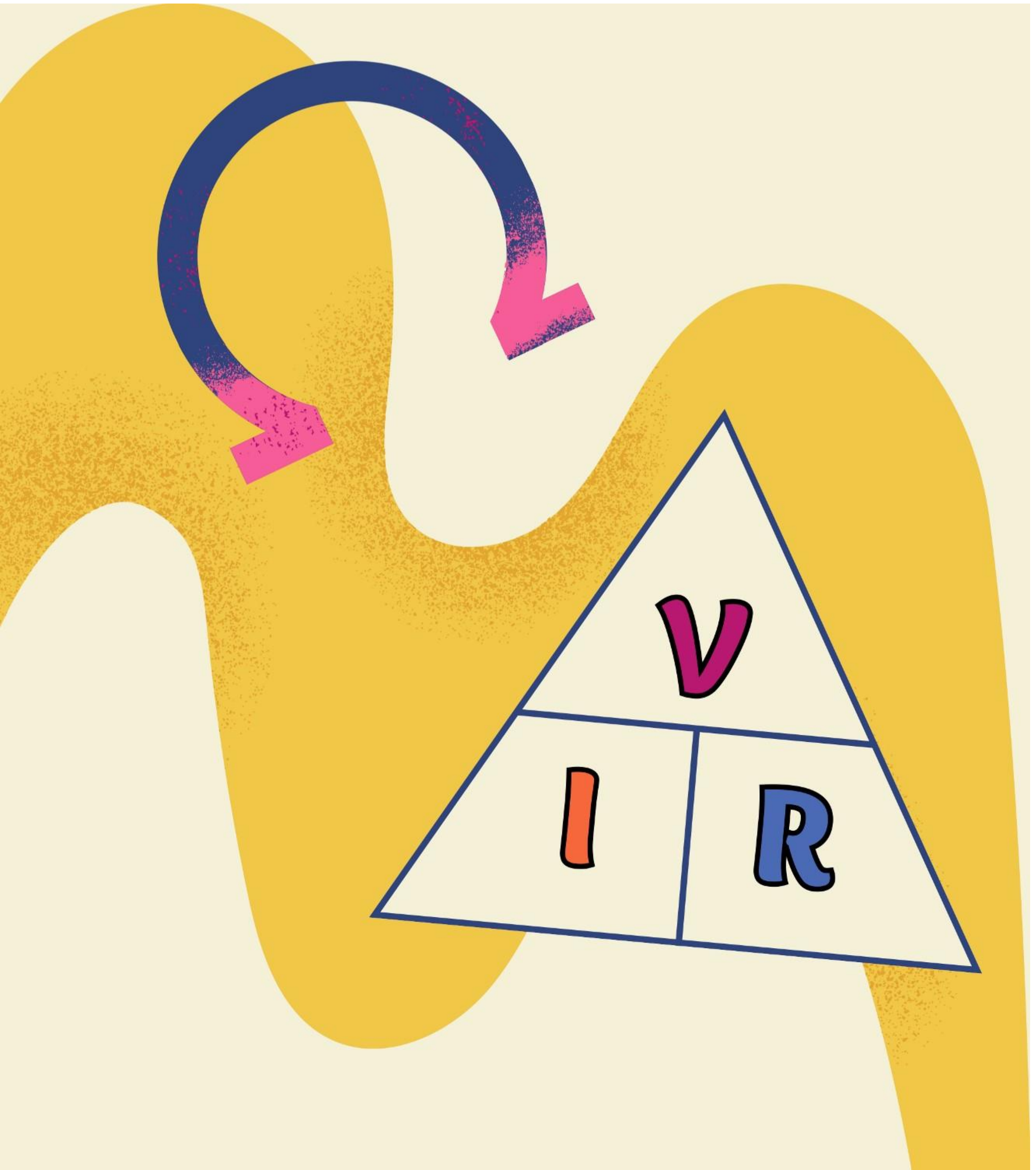
MODEL TRAINING

CONCLUSION

10

RECOMMENDATIONS

INTRODUCTION

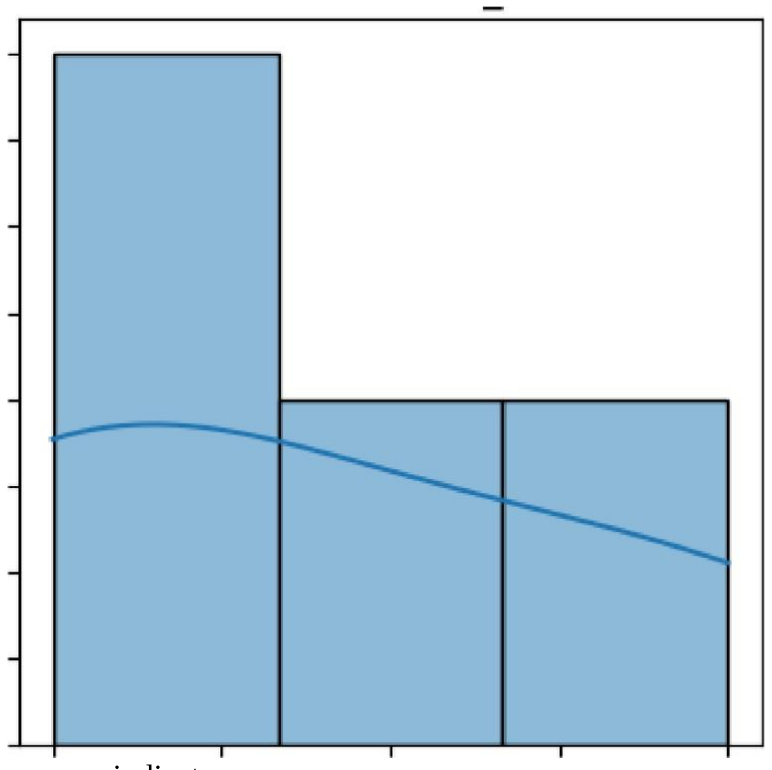


Our aims to develop a machine learning model for classifying and detecting various types of electrical faults in transmission lines.

Will use advanced algorithms and data analysis to enhance the efficiency of power distribution and reduce the risks associated with electrical faults, such as power outages and wildfires.

DISTRIBUTION OF FEATURES

Distribution offault



indicator

2.002.00

1.751.75

1.501.50

1.251.25

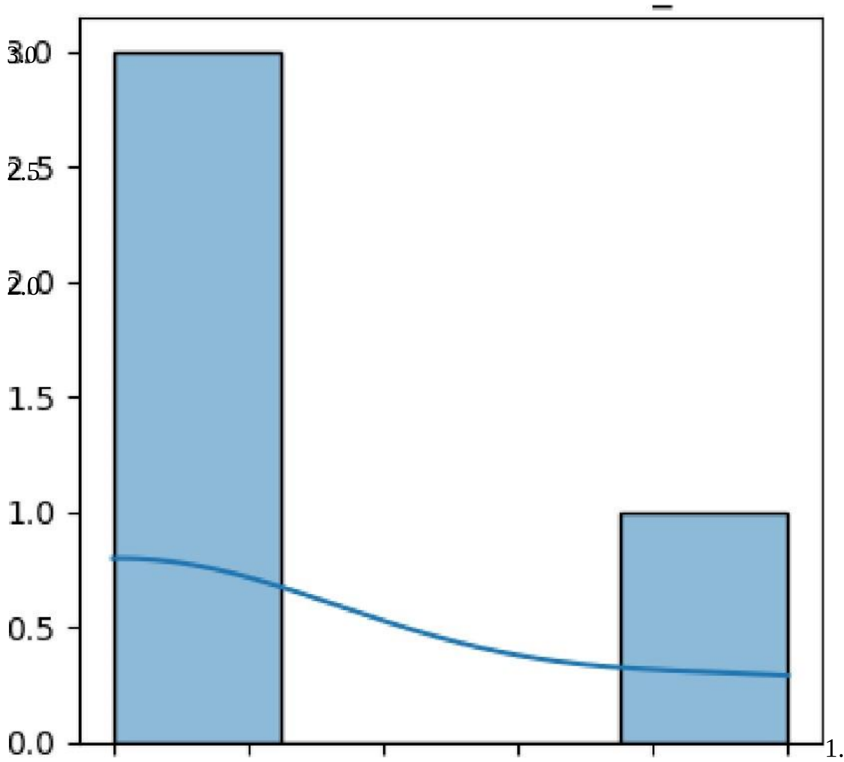
1.001.00

0.750.75

0.500.50

0.0 0.5 1.0 1.5 2.0

0.250.25 S



1.5

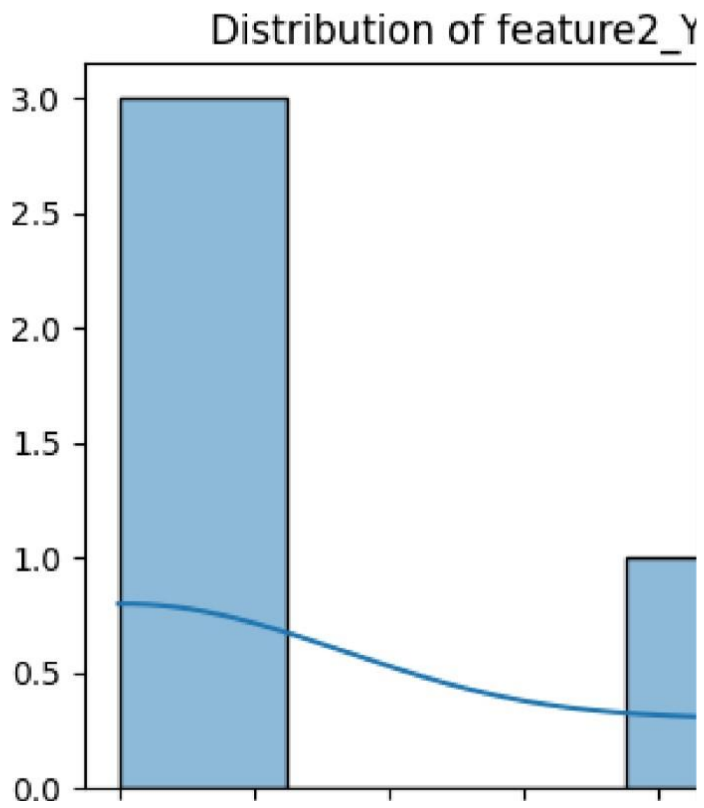
0.000.00

10

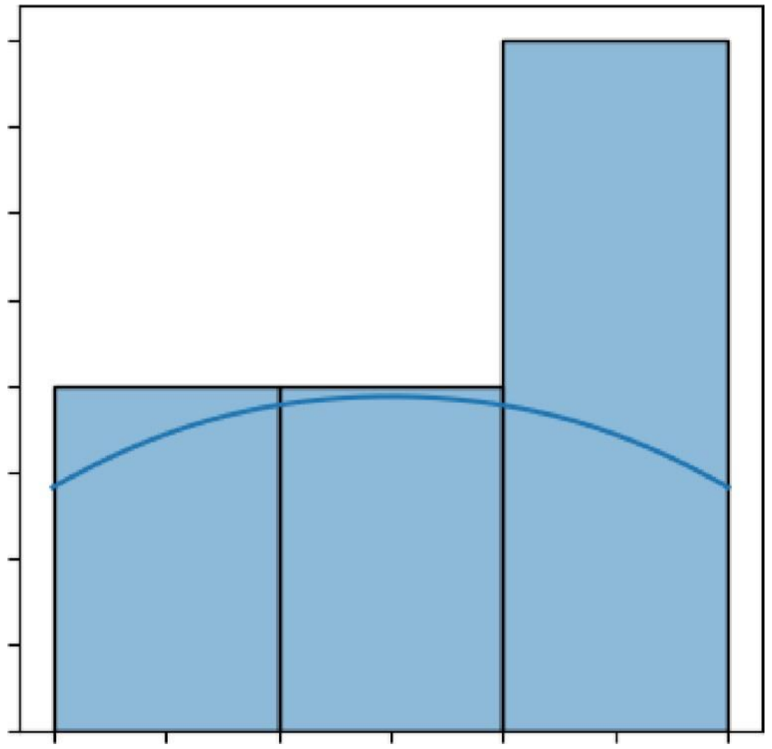
fault indicator

Distribution of feature2 z

0.0 0.2 0.4 0.6 0.8 1.0 feature2 Z



15 20 25 30 35 40 0.0 0.2 0.4 0.6 0.8 featurel feature2 Y



Distribution of featurel

DATA SCALLING

1. Data Prepocessing is done by filling missing values using the mean strategy and encoding categorical variables using label encoding and one-hot encoding.
2. Then Data Visualization using histograms with kernel density estimates to understand the distribution of features.

3. A correlation matrix is also generated using a heatmap to visualize the relationships between different features

DATA SCALLING

4. Finally, the data is split into features and target variables using the train-test-split function from scikit-learn.

The features are then scaled using the StandardScaler class from scikit-learn.

EXPLORATORY DATA ANALYSIS

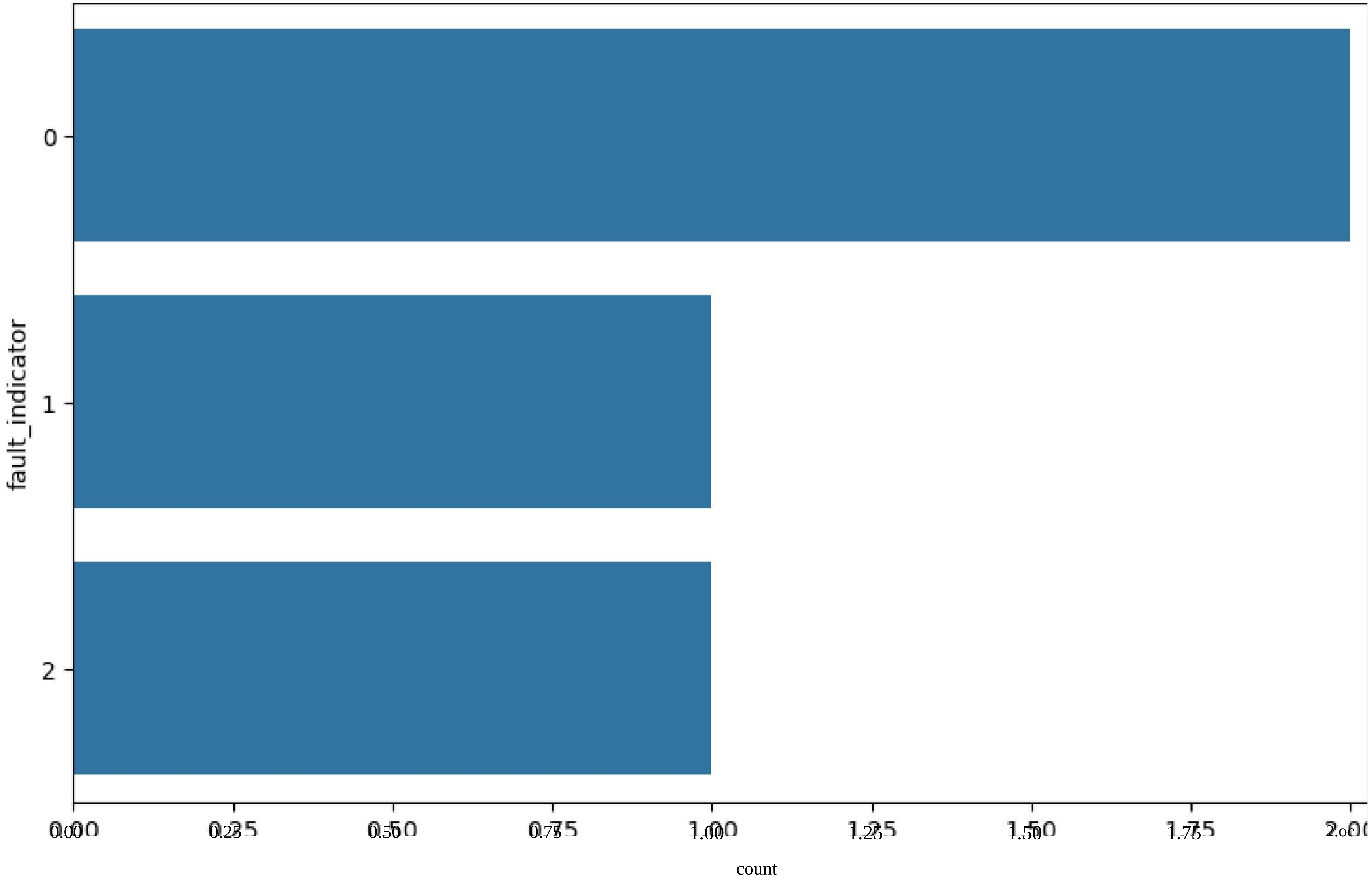
The dataset is performed by printing the unique values in each column.

The output shows the unique values in the 'fault-indicator' column.

In the 'featurel' column as, in the 'feature2-Y' column as [0.0, 1.0], and in the 'feature2-Z' column as [0.0, 1.0].

DATA VISUALIZATION

Distribution of fault types



FEATURE ENGINEERING

iew

f r fe

FEATURE ENGINEERING

The feature engineering process involves creating a new feature by combining existing features. In this case, a new feature 'new_feature' is created by multiplying 'feature1', 'feature2_Y', and 'feature2_Z'. Additionally, a numerical feature 'feature1' is binned into categories 'Low', 'Medium', and 'High' based on predefined bins. The resulting dataset includes the original features, the new feature, the binned feature, and one-hot encoded columns for the 'fault_indicator' feature. The dataset is then scaled using the StandardScaler to ensure all features have the same scale for model training. The final dataset includes the scaled features, the new feature, the binned feature, and the one-hot encoded columns for the 'fault_indicator' feature.

MODELThe accuracy is 0.5,

1

indicating that the model is not performing well

2

The classification are 0.0, indicating that the model is not able to accurately predict the classes.

3

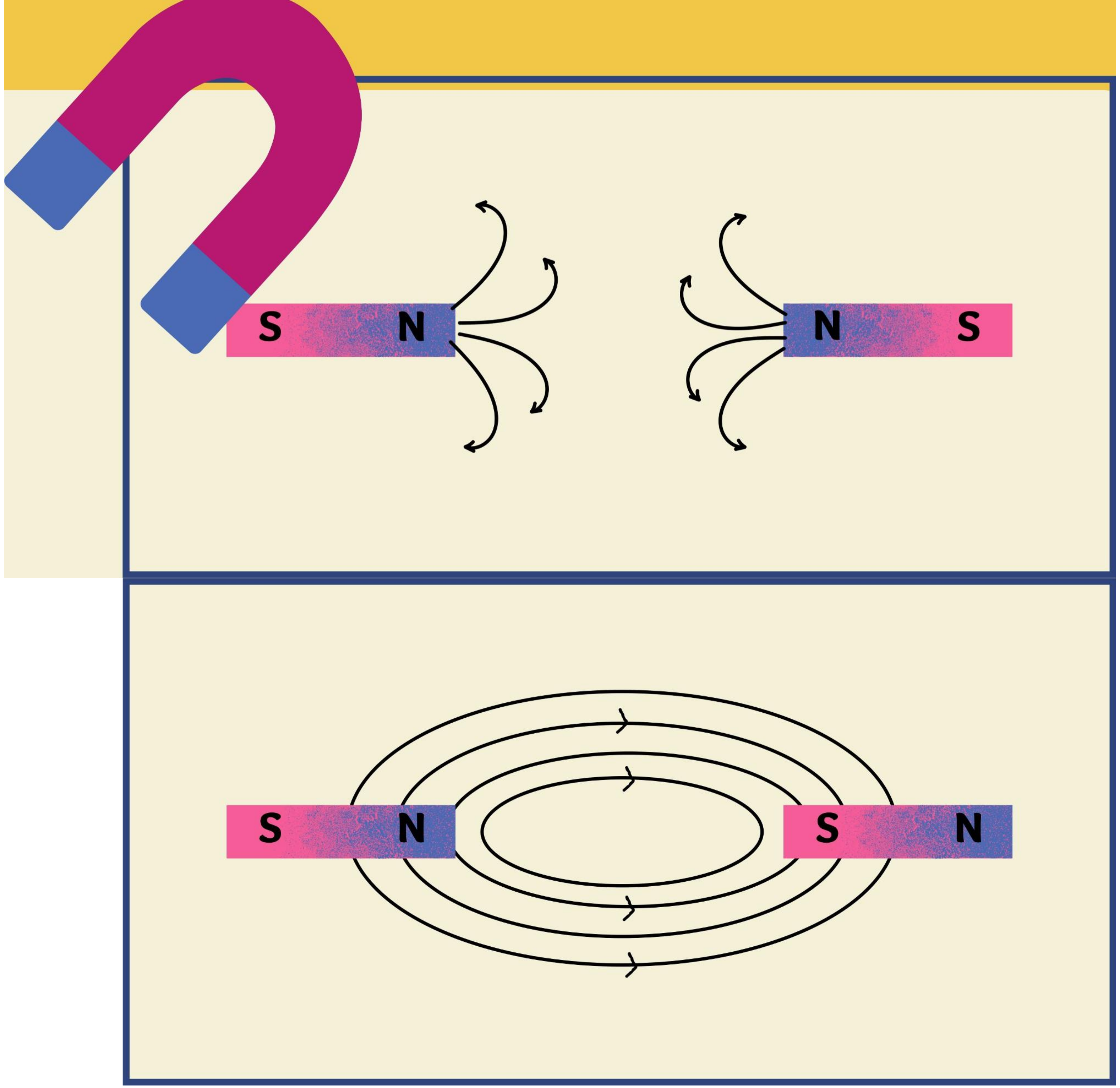
The AUC is 0.5, indicating that the model is not able to distinguish between the classes. cross-validation indicates that the model is not performing well. The average cross-validation score is 0.5.

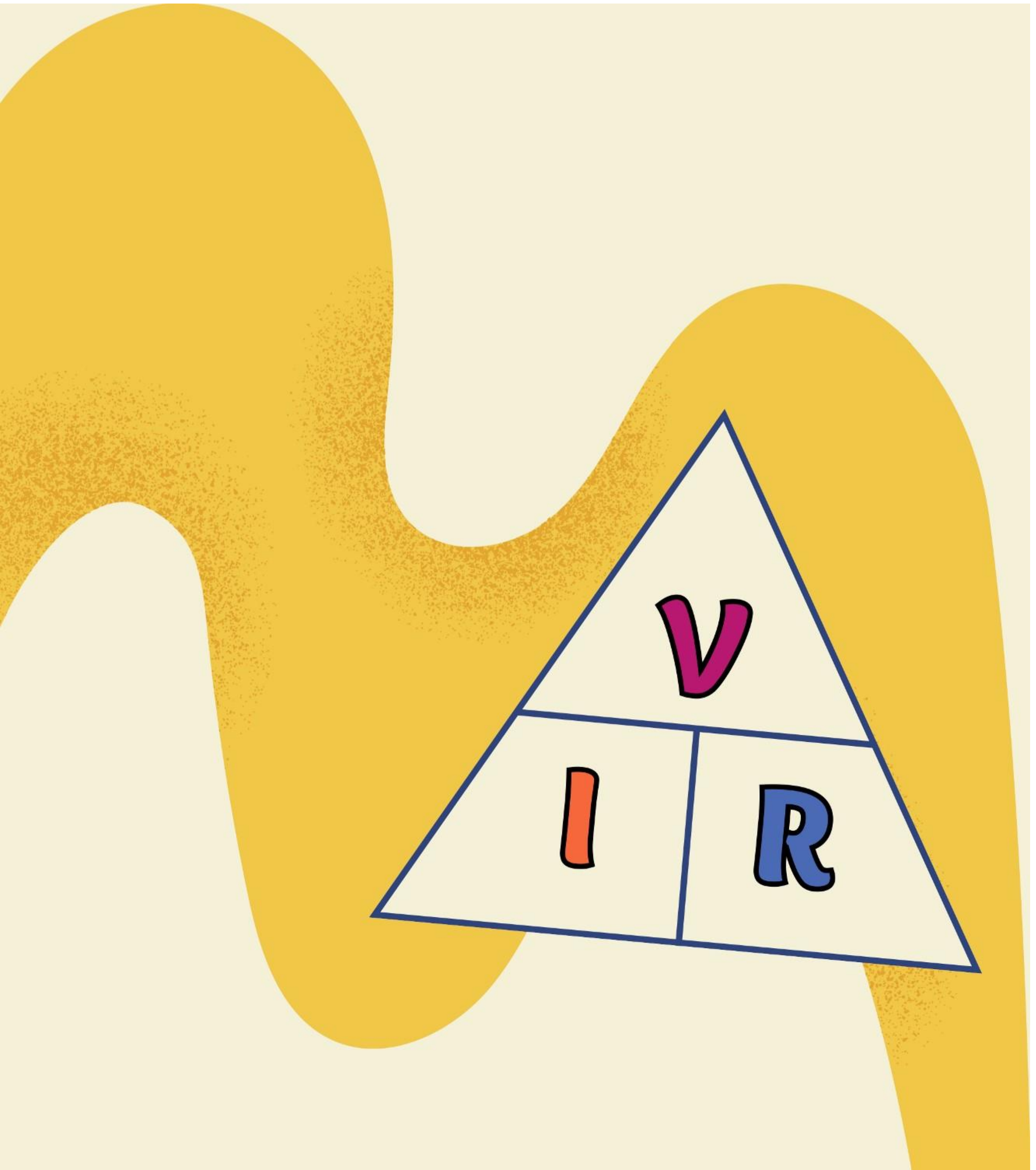
4

CONCLUSION

This dataset is a small but important tool for classifying and detecting faults in a system. The EDA and data visualization steps provide valuable insights into the data, while feature scaling ensures that the da a is prepar d for machine le rning algorith s.

RECOMMENDATIONS1. Feature Importance: Analyze to find key features contributing to fault detection.





2. Model Selection: Experiment with various algorithms like SVM or GBM for fault classification.
3. Hyperparameter Tuning: Optimize model parameters to enhance performance.
4. Cross-Validation: Apply crossvalidation to validate model robustness.
5. Ensemble Methods: Use ensemble techniques like Random Forest or XGBoost for improved accuracy.
6. Threshold Adjustment: Set classification thresholds based on business needs.
7. Monitoring System: Establish a real-time monitoring system for ongoing model evaluation.

THAN K you