Earthquake Prediction using XAI

Rishabh Jain

IIT Delhi
Research Intern
National University of Singapore

Guide: Vishal Srivastava

National University of Singapore

Group Meeting

19th July 2024

NUS
National University
of Singapore

Outline

1. Problem Statement
2. Key Highlights
3. Next Steps

Problem Statement

Objective

Implement and study various post-hoc interpretability methods on machine learning models for earthquake detection

Data

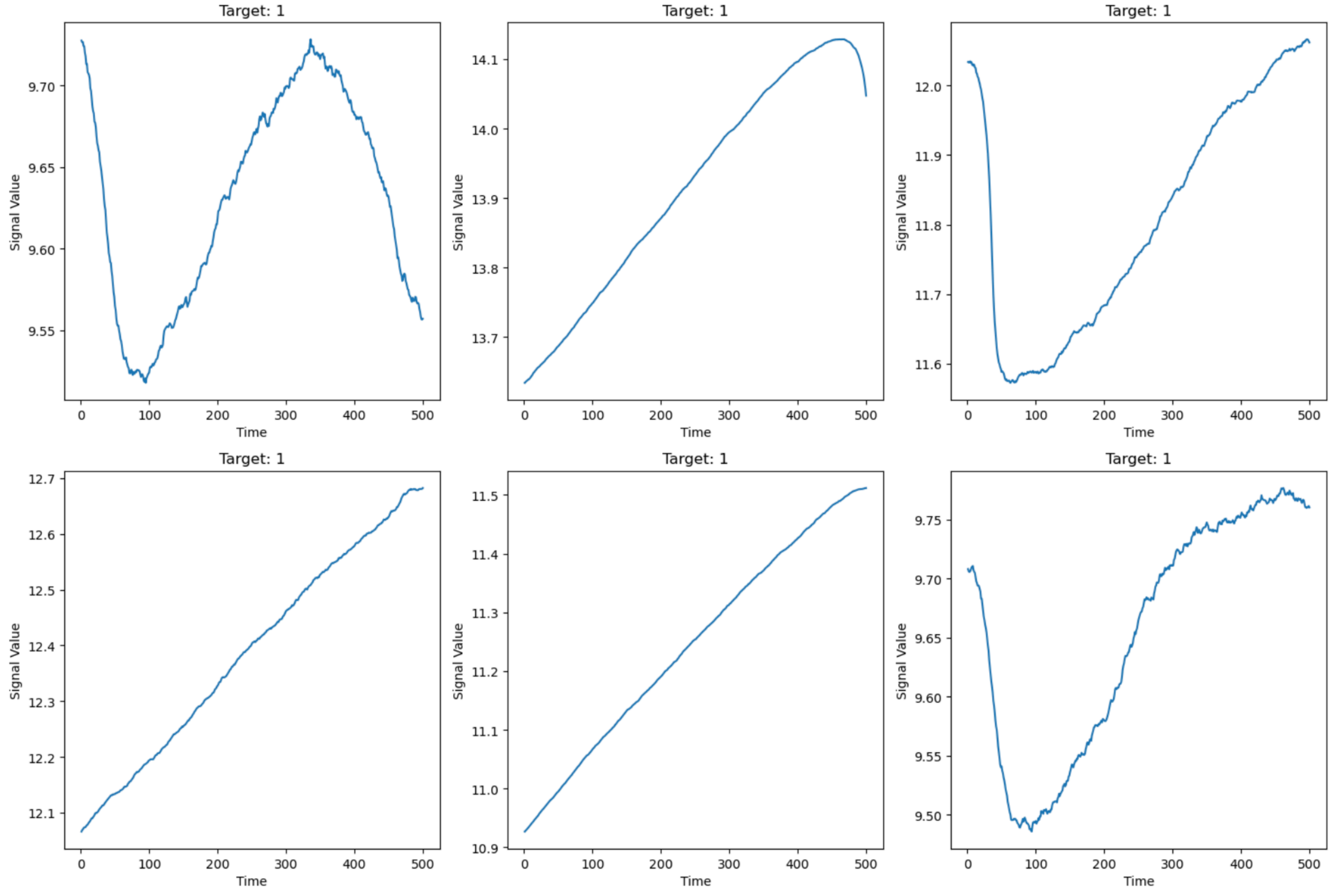
- Seismic electric signal (SES) values for 500 time steps
- Artificially generated using computer simulations
- Roughly 30,000 data points

Target

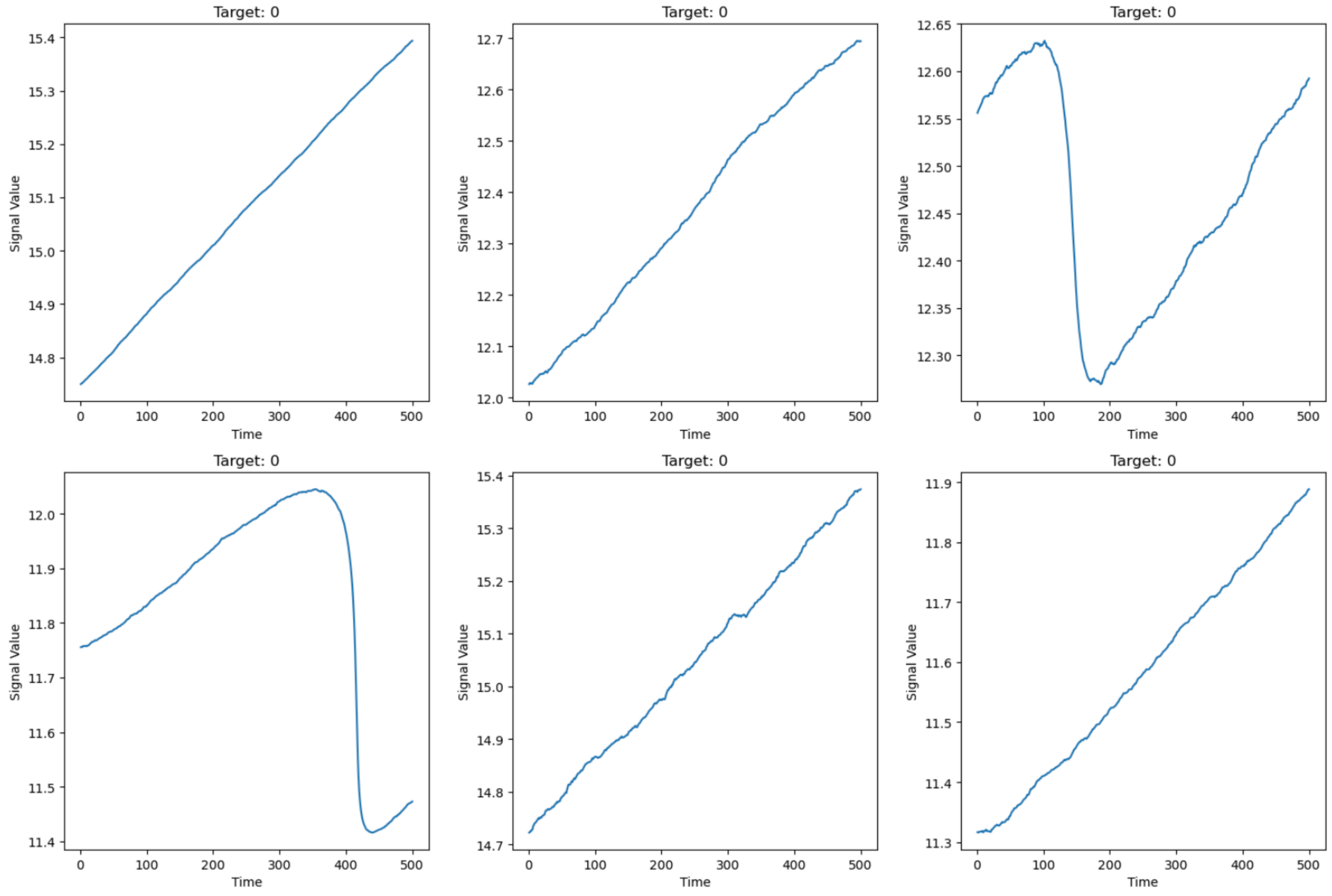
Detect earthquake in next 45 time steps

Signal Value vs Time step

Target = 1



Target = 0



Key Highlights

- Trained 1D CNN and LSTM Network
- Binary Classifiers (Class 1: Earthquake Occurs)
- Extremely poor precision and recall for Class 1
- Models lack sophistication

1D CNN

```
119/119 [==============================] - 2s 13ms/step
              precision    recall  f1-score   support

     class 0       0.80      0.99      0.88      3010
     class 1       0.37      0.02      0.04       784

    accuracy                           0.79      3794
   macro avg       0.58      0.51      0.46      3794
weighted avg       0.71      0.79      0.71      3794
```

LSTM

```
119/119 [==============================] - 24s 182ms/step
              precision    recall  f1-score   support

     class 0       0.79      1.00      0.88      3010
     class 1       0.00      0.00      0.00       784

    accuracy                           0.79      3794
   macro avg       0.40      0.50      0.44      3794
weighted avg       0.63      0.79      0.70      3794
```

Next Steps

Hydra-based framework

1. Fine-tune a model that yields better precision and recall
2. Refer to Prof's paper on 'Evaluation of post-hoc interpretability methods in time-series classification'
3. Implement below mentioned interpretability methods

- Integrated Gradient
- DeepLIFT
- DeepLIFTSHAP
- GradSHAP
- KernelSHAP
- Shapley Sampling

Next Steps

Interpretability Results

1. Compare the results of different interpretability methods
2. Look for patterns and verify them

Transfer Learning

1. Use the refined model to train a new model on laboratory generated data
2. Finally, train and validate the model on real world data

Next Steps

Questions

1. What is considered as an acceptable precision and recall for this problem?

2. How can I validate the results I will obtain from interpretability analysis?

Thank You!