MINE SURVEY AND GEO INFORMATICS MINI PROJECT

Analysis of Seismic Data for predicting the occurrence of rockbursts in underground mines

TEAM MEMBERS:

Aaryan Awadh- 21MI10001
Archana Satapathy- 21MI10012
Gaurav Kumar- 21MI10020
Krish Alok Jain- 21MI10031
Tanya Jain- 21MI10048

INTRODUCTION:

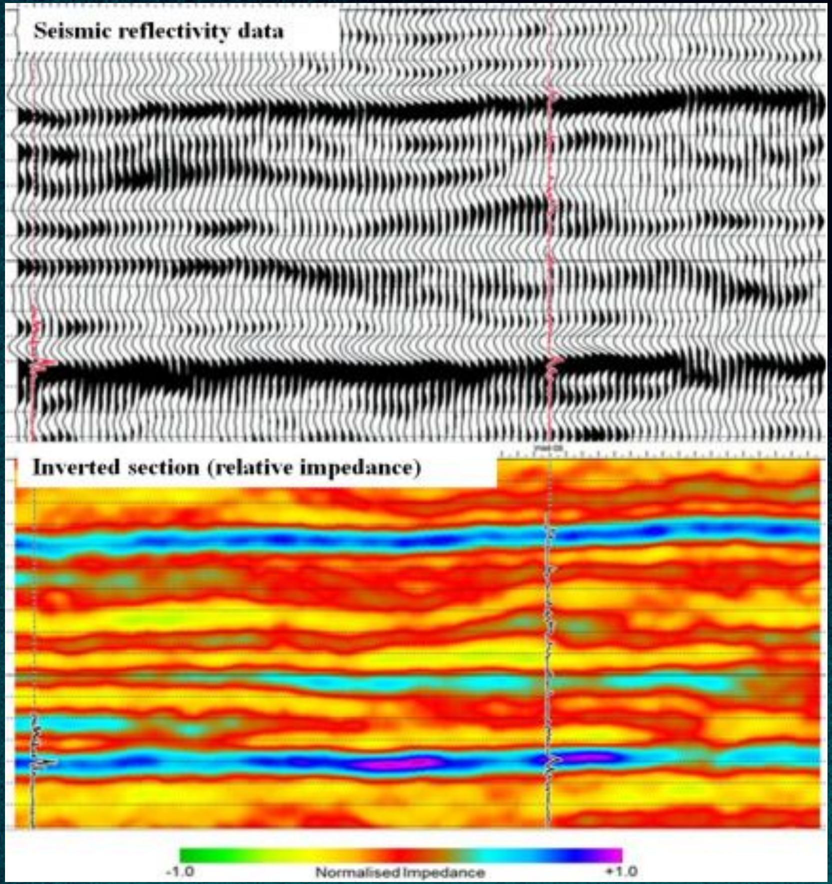
Seismic data analysis is an important tool in predicting the occurrence of rockbursts in underground mines.
Seismic data analysis involves the following steps:

- **Data acquisition:** Seismic data is acquired through the use of seismometers, which are placed strategically in the mine.

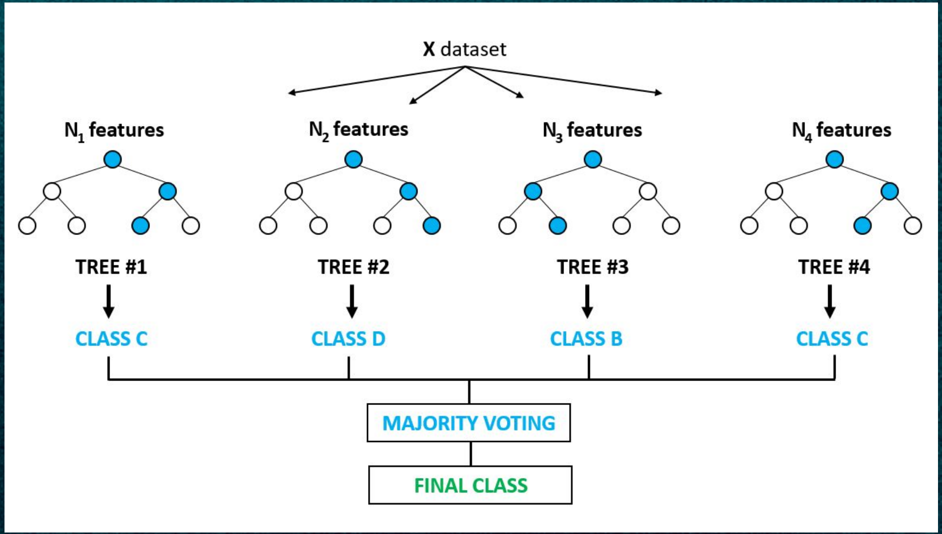
- **Data processing:** Seismic data is processed to remove noise and enhance the signal. This involves filtering, deconvolution, and normalization.

- **Data interpretation:** Seismic data is interpreted to identify the location, intensity, and direction of rock mass movement. This involves analyzing the amplitude, frequency, and polarization of the waves.

- **Risk assessment:** Seismic data is used to assess the risk of rockbursts in the mine. This involves analyzing the seismicity of the mine, identifying areas of high risk, and developing mitigation strategies.



MAJOR PROCESS:



- Feature Engineering
- Split the Data
- Train the Model
- Hyperparameter Tuning
- Assess Model Performance
- Trying New Models for Higher Accuracy
- Making the final Predictions

DATA SET:

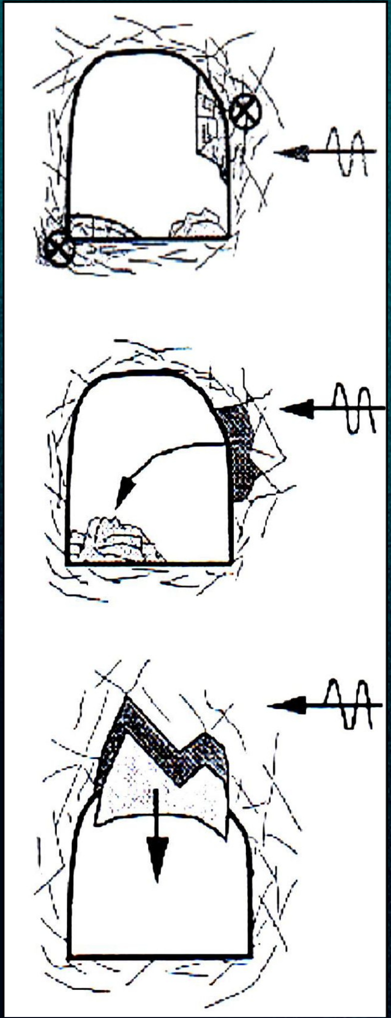
Seismic Hazard is the probability that an earthquake will occur in a given geographic area, within a given window of time, with ground motion intensity exceeding a particular threshold.

The data present in this data set is taken from Zabrze-Bielszowice coal mine, in Poland.

Independent Variables

- **seismic** : Result of shift seismic hazard assessment obtained by seismic method a(lack of hazard), b(low hazard), c(high hazard), d(danger state)
- **seismoacoustic** : Result of shift seismic hazard assessment obtained through seismoacoustic method.
- **shift** : Type of shifts w(coal getting), n(preparation shift)
- **genergy** : Seismic energy recorded within previous shift by the most active geophone out of all geophones monitoring the long wall
- **gpuls** : Number of pulses recorded within previous shift by the most active geophone
- **gdenergy** : A deviation of energy recorded within the previous shift by the most active geophone**gdpuls** : A deviation of a number of pulses recorded within previous shift by GMax from average number
- **energy** : total energy of seismic bumps
- **energy** : total energy of seismic bumps registered within previous shift

maxenergy : the maximum energy of the seismic bumps registered within previous shift



DATA SET:

- **energy** : total energy of seismic bumps registered within previous shift
- **maxenergy** : the maximum energy of the seismic bumps registered within previous shift
- **nbumps** : the number of seismic bumps recorded within previous shift
- **nbumps2** : the number of seismic bumps (in energy range [10^2,10^3)) registered within previous shift
- **nbumps3** : the number of seismic bumps (in energy range [10^3,10^4)) registered within previous shift
- **nbumps4** : the number of seismic bumps (in energy range [10^4,10^5)) registered within previous shift
- **nbumps5** : the number of seismic bumps (in energy range [10^5,10^6)) registered within previous shift
- **nbumps6** : the number of seismic bumps (in energy range [10^6,10^7)) registered within previous shift
- **nbumps7** : the number of seismic bumps (in energy range [10^7,10^8)) registered within previous shift
- **nbumps89** : the number of seismic bumps (in energy range [10^8,10^10)) registered within previous shift

DATA SET:

Dependent Variable

- **class** : the decision attribute - "1" means that high energy seismic bump occurred in the next shift ("hazardous state"), "0" means that no high energy seismic bumps occurred in the next shift ("non-hazardous state")

Python Libraries used

NumPy

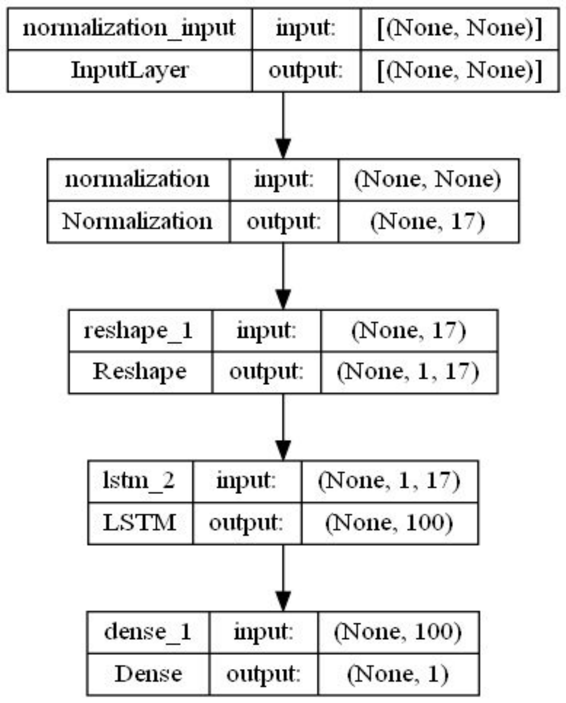
Pandas

TensorFlow

Scikit Learn

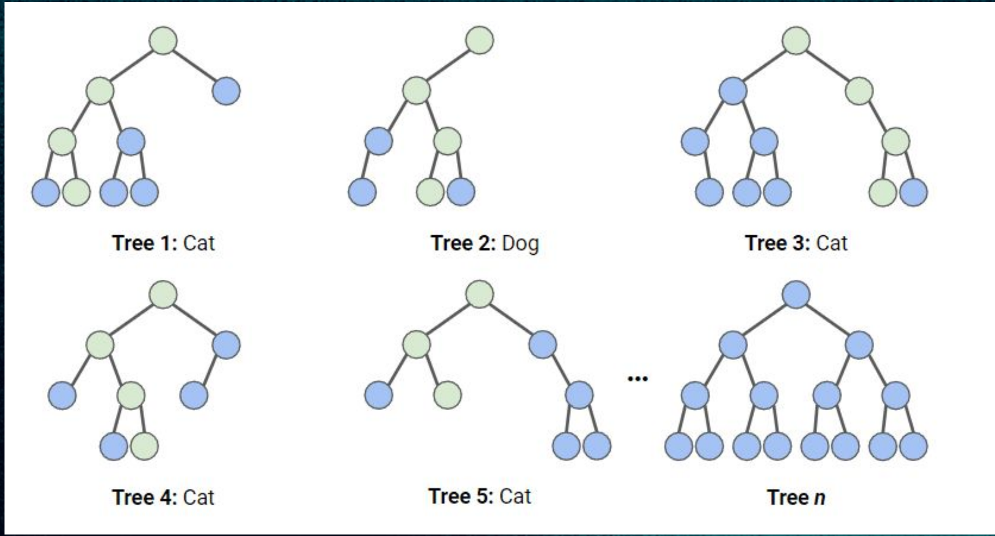
ABOUT ALGORITHMS:

An LSTM layer is an RNN layer that learns long-term dependencies between time steps in time series and sequence data. The state of the layer consists of the hidden state (also known as the output state) and the cell state. The hidden state at time step t contains the output of the LSTM layer for this time step.

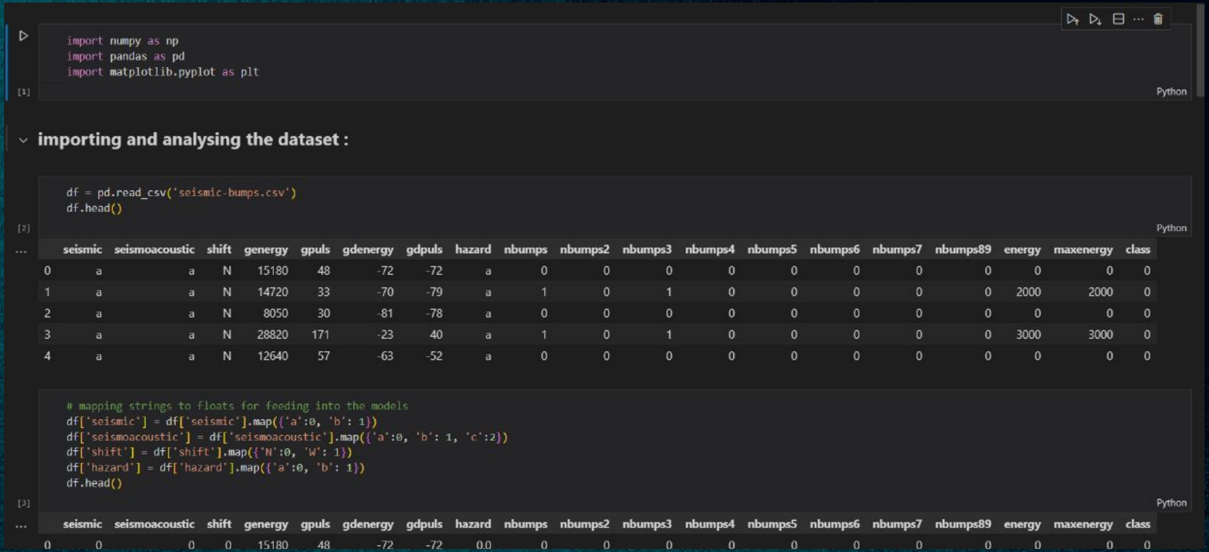


ABOUT ALGORITHMS:

The algorithm used is Random Forest Classification. It can be used for both classification and regression. In a random forest classification, multiple decision trees are created using different random subsets of the data and features. Each decision tree is like an expert, providing its opinion on how to classify the data. Predictions are made by calculating the prediction for each decision tree, then taking the most popular result.
In the diagram below, we have a random forest with n decision trees, and we've shown the first 5, along with their predictions (either "Dog" or "Cat").



PYTHON CODE



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[1] Python

importing and analysing the dataset :

```
df = pd.read_csv('seismic-bumps.csv')
df.head()
```

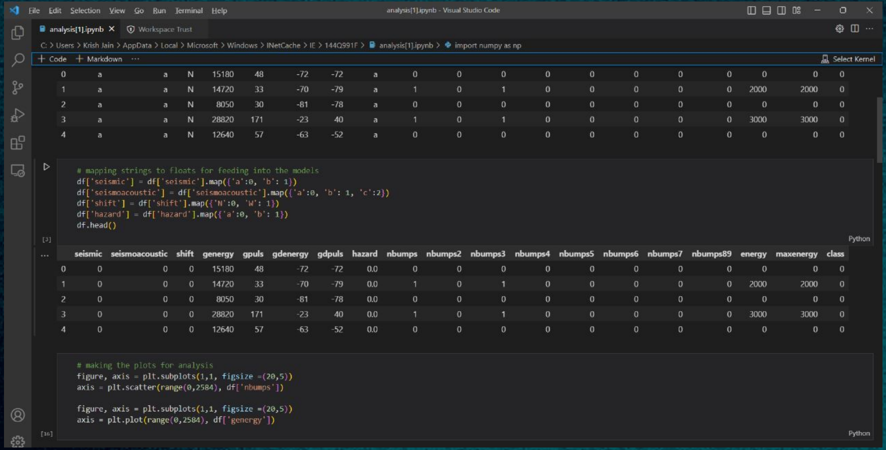
[2] Python

	seismic	seismoacoustic	shift	genergy	gpuls	gdenergy	gdpuls	hazard	nbumps	nbumps2	nbumps3	nbumps4	nbumps5	nbumps6	nbumps7	nbumps89	energy	maxenergy	class
0	a	a	N	15180	48	-72	-72	a	0	0	0	0	0	0	0	0	0	0	0
1	a	a	N	14720	33	-70	-79	a	1	0	1	0	0	0	0	0	2000	2000	0
2	a	a	N	8050	30	-81	-78	a	0	0	0	0	0	0	0	0	0	0	0
3	a	a	N	28820	171	-23	40	a	1	0	1	0	0	0	0	0	3000	3000	0
4	a	a	N	12640	57	-63	-52	a	0	0	0	0	0	0	0	0	0	0	0

```
# mapping strings to floats for feeding into the models
df['seismic'] = df['seismic'].map({'a':0, 'b': 1})
df['seismoacoustic'] = df['seismoacoustic'].map({'a':0, 'b': 1, 'c':2})
df['shift'] = df['shift'].map({'N':0, 'W': 1})
df['hazard'] = df['hazard'].map({'a':0, 'b': 1})
df.head()
```

[3] Python

	seismic	seismoacoustic	shift	genergy	gpuls	gdenergy	gdpuls	hazard	nbumps	nbumps2	nbumps3	nbumps4	nbumps5	nbumps6	nbumps7	nbumps89	energy	maxenergy	class
0	0	0	0	15180	48	-72	-72	0.0	0	0	0	0	0	0	0	0	0	0	0



0	a	a	N	15180	48	-72	-72	a	0	0	0	0	0	0	0	0	0	0	0
1	a	a	N	14720	33	-70	-79	a	1	0	1	0	0	0	0	0	2000	2000	0
2	a	a	N	8050	30	-81	-78	a	0	0	0	0	0	0	0	0	0	0	0
3	a	a	N	28820	171	-23	40	a	1	0	1	0	0	0	0	0	3000	3000	0
4	a	a	N	12640	57	-63	-52	a	0	0	0	0	0	0	0	0	0	0	0

```
# mapping strings to floats for feeding into the models
df['seismic'] = df['seismic'].map({'a':0, 'b': 1})
df['seismoacoustic'] = df['seismoacoustic'].map({'a':0, 'b': 1, 'c':2})
df['shift'] = df['shift'].map({'N':0, 'W': 1})
df['hazard'] = df['hazard'].map({'a':0, 'b': 1})
df.head()
```

[3] Python

	seismic	seismoacoustic	shift	genergy	gpuls	gdenergy	gdpuls	hazard	nbumps	nbumps2	nbumps3	nbumps4	nbumps5	nbumps6	nbumps7	nbumps89	energy	maxenergy	class
0	0	0	0	15180	48	-72	-72	0.0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	14720	33	-70	-79	0.0	1	0	1	0	0	0	0	0	2000	2000	0
2	0	0	0	8050	30	-81	-78	0.0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	28820	171	-23	40	0.0	1	0	1	0	0	0	0	0	3000	3000	0
4	0	0	0	12640	57	-63	-52	0.0	0	0	0	0	0	0	0	0	0	0	0

```
# making the plots for analysis
figure, axis = plt.subplots(1,1, figsize =(20,5))
axis = plt.scatter(range(0,2584), df['nbumps'])

figure, axis = plt.subplots(1,1, figsize =(20,5))
axis = plt.plot(range(0,2584), df['genergy'])
```

[16] Python

analysis[1].ipynb - Visual Studio Code

C: > Users > Krish Jain > AppData > Local > Microsoft > Windows > INetCache > IE > 144Q991F > analysis[1].ipynb > import numpy as np

Splitting into training and testing part:

80:20 split

```
x_train = df.drop(['class','hazard'], axis=1, inplace=False).iloc[0:2067]
x_test = df.drop(['class','hazard'], axis=1, inplace=False).iloc[2067:2584]

y_train = df['class'].iloc[0:2067]
y_test = df['class'].iloc[2067:2584]
```

[11] Python

Selecting the model and training it:

Random Forest is a very good classification model so let's try that.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import RandomizedSearchCV
from scipy.stats import randint
```

[6] Python

```
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[12] Python

```
RandomForestClassifier
RandomForestClassifier()
```

Ln 1, Col 1 Spaces: 4 CRLF Cell 1 of 16

analysis[1].ipynb - Visual Studio Code

C: > Users > Krish Jain > AppData > Local > Microsoft > Windows > INetCache > IE > 144Q991F > analysis[1].ipynb > import numpy as np

Checking the accuracy:

```python
y_pred = rf.predict(x_test)
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

[13]

```
Accuracy: 0.965183752417795
```

As we can see the accuracy is very high and thus this model can very relaibly predict the occurence of earthquakes from the sensor inputs. Now let us try and increase its accuracy by tuning the hyperparameters.

```python
param_dist = {'n_estimators': randint(50,500), 'max_depth': randint(1,20)}

# Use random search to find the best hyperparameters
rand_search = RandomizedSearchCV(rf, param_distributions = param_dist, n_iter=5, cv=5)

# Fit the random search object to the data
rand_search.fit(x_train, y_train)
```

[9]

```
RandomizedSearchCV
estimator: RandomForestClassifier
RandomForestClassifier
```

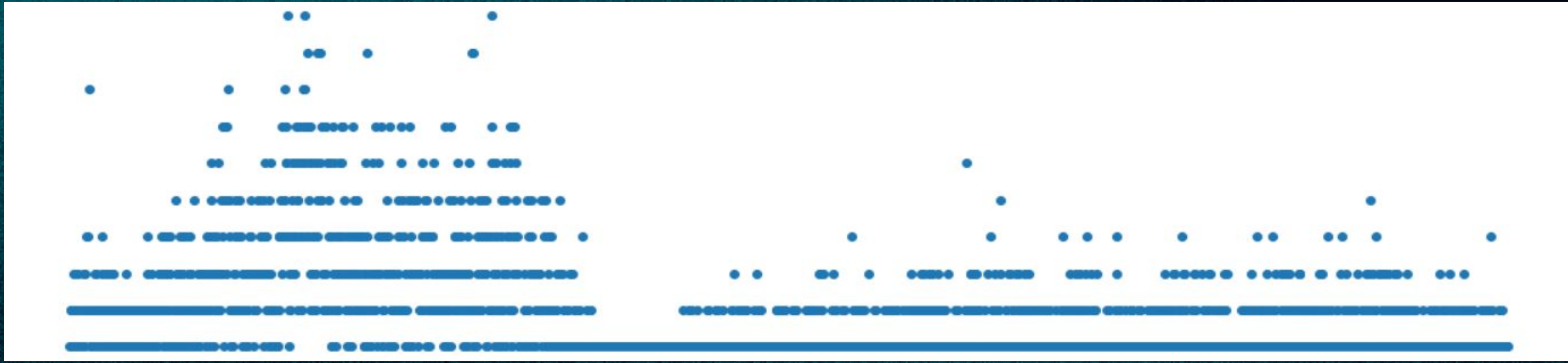
```python
# Create a variable for the best model
best_rf = rand_search.best_estimator_

y_pred = best_rf.predict(x_test)
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

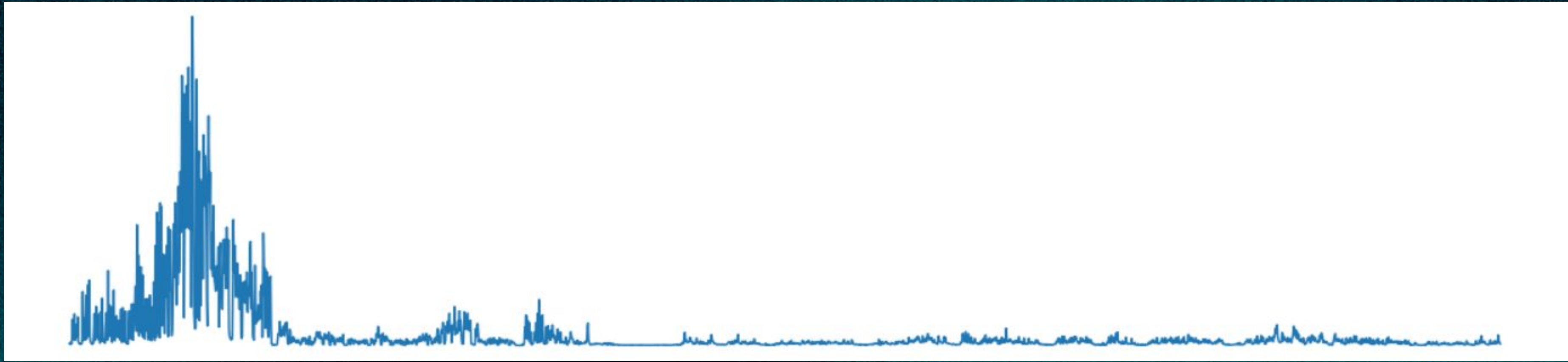
[14]

```
Accuracy: 0.965183752417795
```

The accuracy remains almost the same so our final accuracy is **96.51%** which is very reliable for prediction purposes.

Graphs Generated:





Advantages:

There are several benefits of conducting a project on the analysis of seismic data for predicting the occurrence of rockbursts in underground mines:

Improved safety: The main benefit of conducting such a project is to improve the safety of miners and mining operations. By predicting and mitigating the risk of rockbursts, the project can help prevent injury and loss of life, as well as damage to equipment and infrastructure.

Cost savings: By reducing the risk of rockbursts, mining operations can be conducted more smoothly and with greater productivity. This can result in cost savings for the mining company, as well as improved efficiency and profitability.

Environmental protection: By reducing the likelihood of rockbursts, the project can also help to protect the environment by reducing the risk of spills and leaks of hazardous substances.

Improved mining practices: The project can also lead to the development of improved mining practices and techniques that can be applied in other mines and industries.

Scientific advancement: By conducting research on the analysis of seismic data, the project can contribute to the scientific understanding of rock mass behavior and earthquake prediction.

References:

Kaggle.com
Wikipedia
Stackoverflow.com
docs.opencv.org

Thank You!