ANALYZING SEISMIC DATA FOR EARLY WARNING DETECTION SYSTEMS

Dipti Aswal, gusaswdi@student.gu.se
Ian Rhys Jenkins, gusjenia@student.gu.se

A need for - Volcanic early warning systems

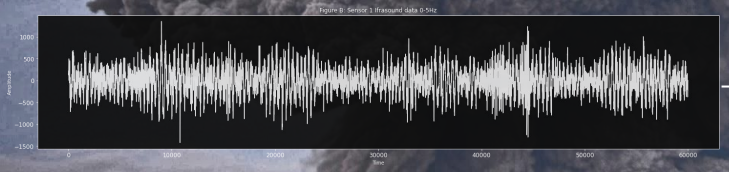
Recent Eruptions:

- Iceland 2010
- Italy - Mt Etna 2013
- New Zealand 2019

Related works:

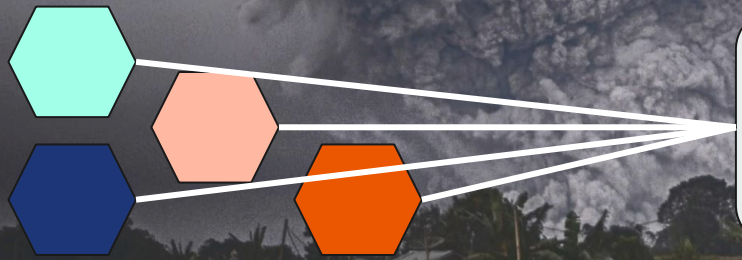
Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives - Marielle Malfante et al.(2018)
SVM : 92% accuracy classification of 6 sonic classes.

Prototype - Input 10 min long sequence



Preprocesses data (FFT)

Extract Input Features



MODEL

Predict n-classes : related to time to eruption

Topics

Terminology

Raw data analysis

Data preprocessing

Data description

Feature selection

Model evaluation

Next steps

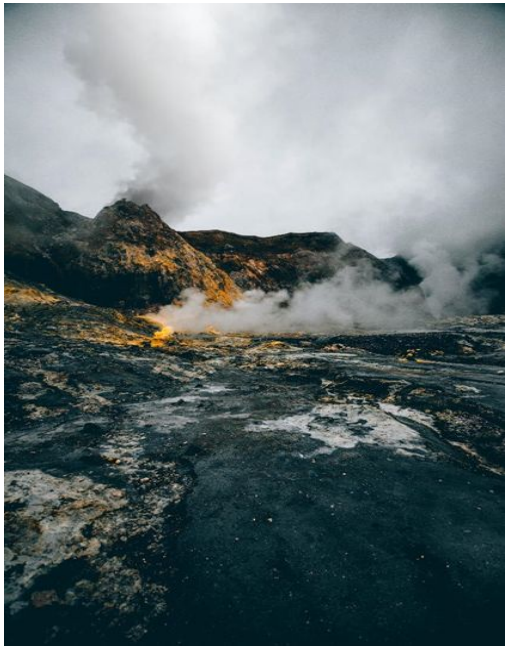
Terminology

Seismic waves : Energy travelling through the earth.

LP events: Volcanic-Earthquake related events that produce very low frequency waves, called infrasound.

Frequency range: Frequency events are 0-9Hz, completely inaudible to the human ear.

Time Series data: Data that occur in sequence.



Raw Data Analysis

Listening to: Long Period events (volcanic related earthquakes)

1-min-5 days before eruption

Dataset includes: many eruptions, captured by several seismic sensors.

60,000 timesteps,10 sensors, 4000 samples

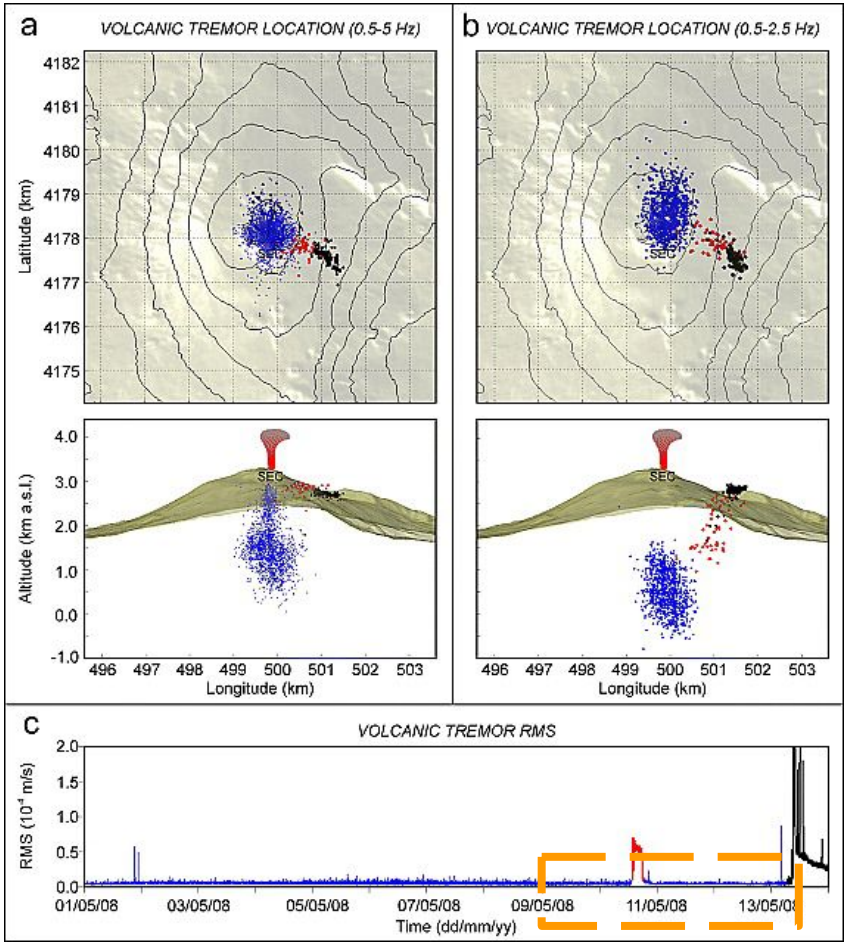
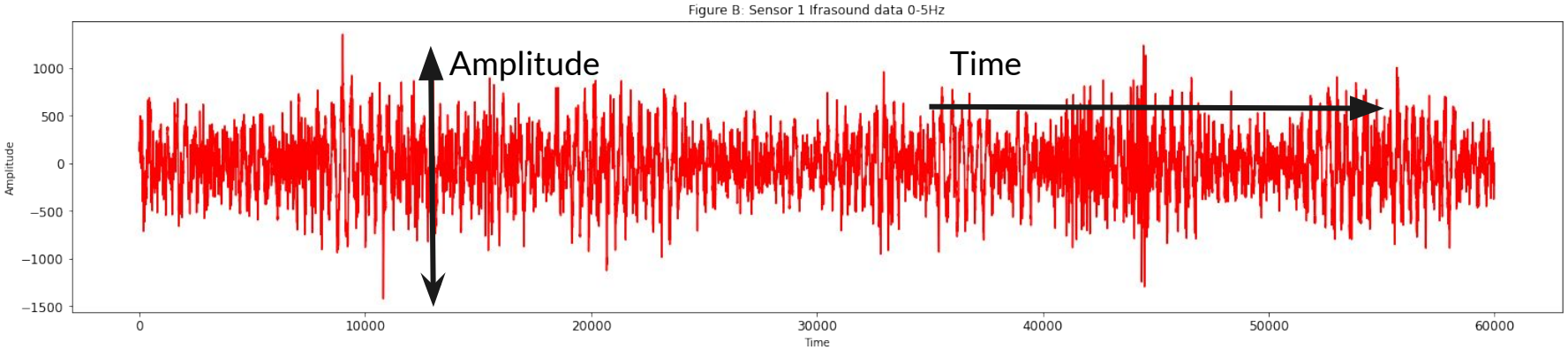


Image: Figure 8 Multidisciplinary investigation on a lava fountain preceding a flank eruption: The 10 May 2008 Etna case, A. Bonaccorso, 2008

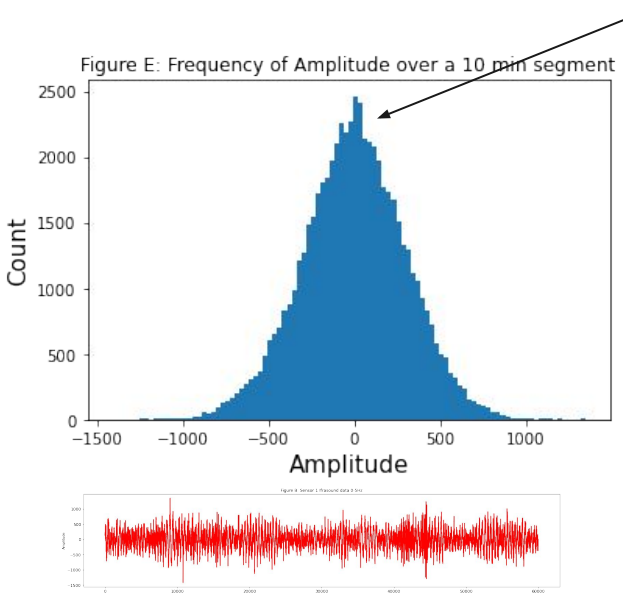
Raw Data Analysis

60,000 timesteps,10 sensors,

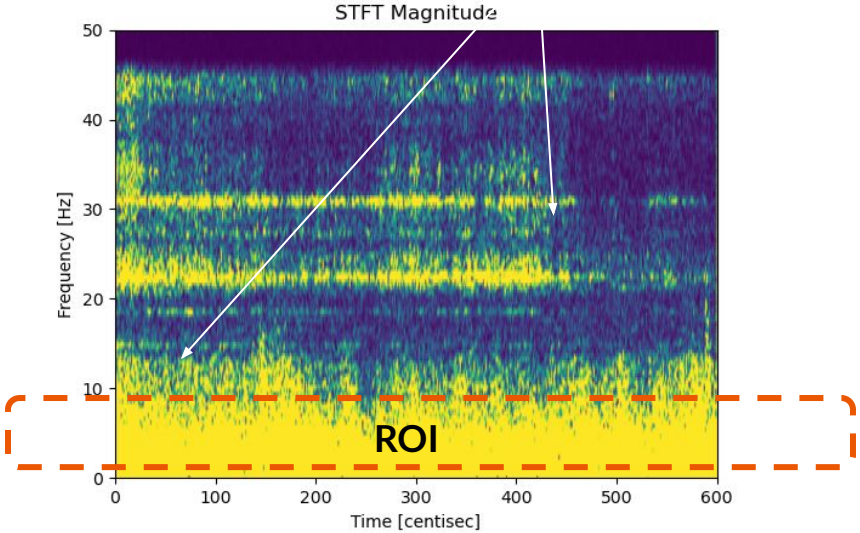


Data Preprocessing

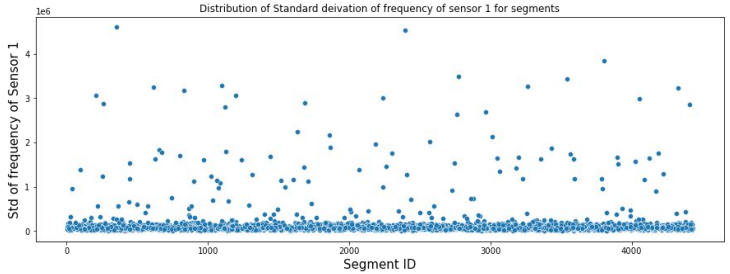
Ambiguous data representation

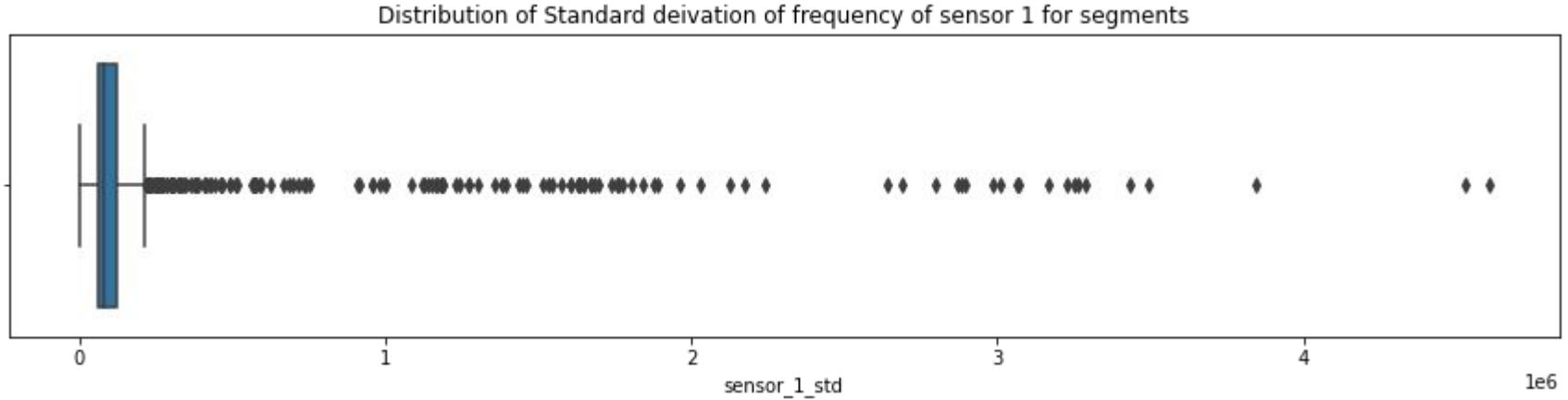


Frequency over Time



Data Description





Topics

Business Case

Terminology

Dataset description

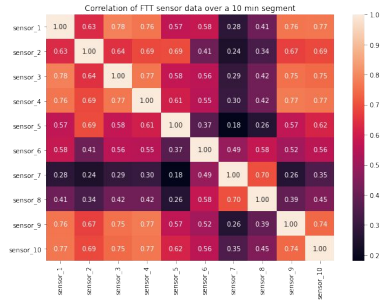
Pre-processing

Feature selection

Performance

Next Steps

Feature Selection




```
# Testing H0 for sensor 1 and 9
stats.ttest_ind(X_train[['sensor_1_mean','sensor_1_std']],X_train[['sensor_9_mean','sensor_9_std']])

# As per the test statistic p_value for both mean and std deviation is > 0.05 hence we "Do Not Reject the hypothesis"
```

```
Ttest_indResult(statistic=array([-0.91673247,  0.0068563 ]), pvalue=array([0.35932185, 0.99452975]))
```

Models

Gaussian Naive Bayes - Base

SVC

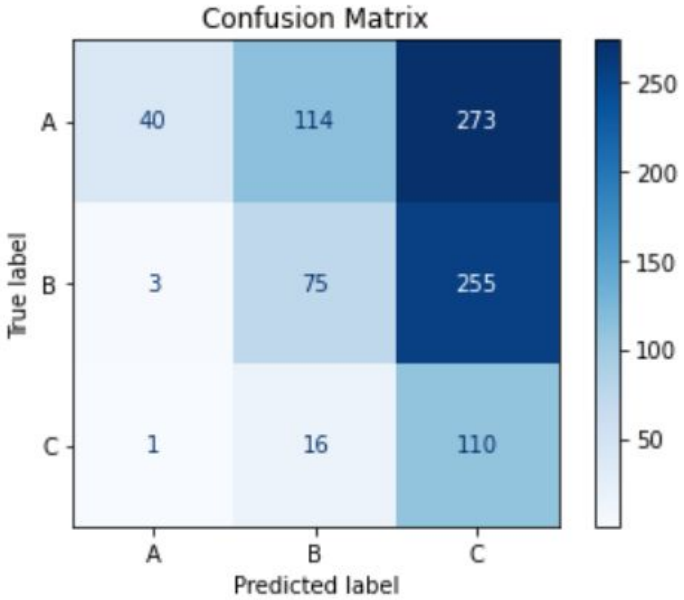
Decision Tree Classifier

Gaussian Naive Bayes

- Features using normal distribution
- Supports multi-class classification
- Assumes independent features
- Supports continually valued features
- Simple and fast algorithm

Equation:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$




```
Accuracy: 0.25366403607666294
Precision: 0.25366403607666294
Recall: 0.25366403607666294
F1 score: 0.25366403607666294
```

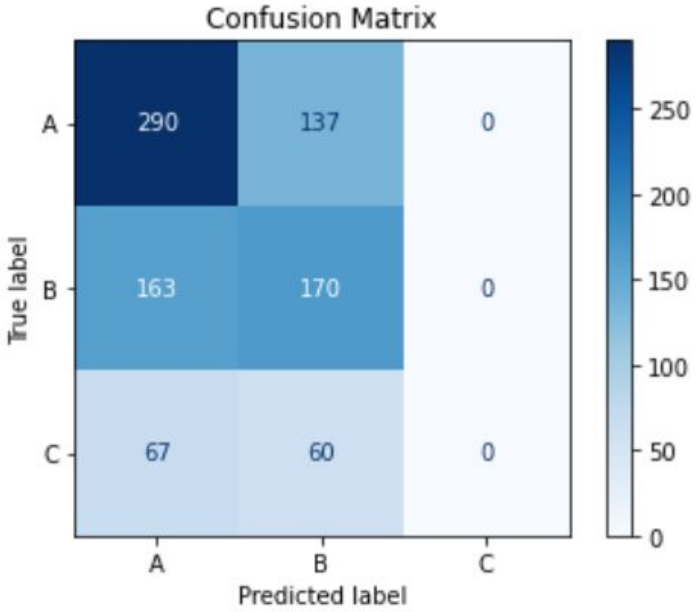
SVC

- Non-parametric algorithm
- Used for multi-class classification
- Effective in high dimensional spaces
- Versatile: Uses kernel trick to model decision boundaries

Optimisation function:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$

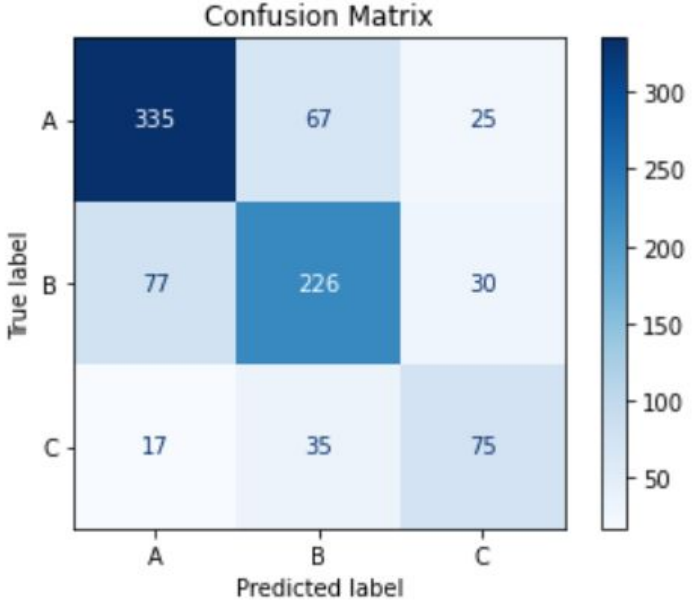
$$\text{subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$
$$\zeta_i \geq 0, i = 1, \ldots, n$$




```
Accuracy: 0.5186020293122886
Precision: 0.5186020293122886
Recall: 0.5186020293122886
F1 score: 0.5186020293122886
```

Decision Tree Classifier

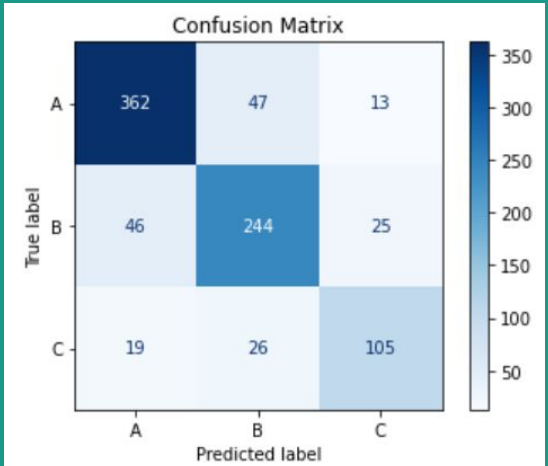
- Non-parametric algorithm
- Supports multi-class classification
- Handles non-linear data sets effectively
- Feature selection happens automatically
- Correlated features do not affect quality




```
Accuracy: 0.7170236753100339
Precision: 0.7170236753100339
Recall: 0.7170236753100339
F1 score: 0.7170236753100339
```

Next Steps

- Feature selection - Increased Accuracy by 10%
- Investigate outliers - Filter
- Regression models - LSTM time series data,




```
Accuracy: 0.8015783540022547
Precision: 0.8015783540022547
Recall: 0.8015783540022547
F1 score: 0.8015783540022547
```

Thank you- Hear and see sounds of the earth.

https://www.youtube.com/watch?v=A2BTqaRhRF0&feature=youtu.be