Analysing Seismicity to Forecast Volcanic Eruptions: A Machine Learning Approach

Hilal Mustafa Khan
khanhilalmustafa@gmail.com

Mohd Mazin Jameel
mazinjameel987@gmail.com

Nabeel Ali
nabe63552@gmail.com

1 Abstract

Volcanic eruptions are a fascinating and complex phenomenon, and anticipating their timing and location is of paramount importance to protect the surrounding communities. While volcanic unrest is often preceded by a range of observable signals such as increased earthquake rates, ground inflation, and gas emissions, their implications for eruption timing remain unclear.

In this project report, we applied various machine learning algorithms, including logistic regression, random forests, k-means clustering, and neural networks, to explore the relationship between earthquakes and volcanic eruptions at Kilauea, a highly active volcano in Hawaii. We first used individual earthquake characteristics to predict the contemporaneous eruption of the volcano, a binary classification task. We then attempted to predict the time until an eruption begins using regression.

Our findings suggest that the random forest algorithm demonstrated some success in forecasting eruptions from the earthquake catalog. However, it is essential to note that additional datasets or extracted features beyond the historical earthquake catalog may be required for future progress in eruption forecasting.

2 Introduction

It is a well-established fact that over 600 million people live in close proximity to a volcano and are therefore exposed to their potentially disastrous hazards. Consequently, it is crucial to develop accurate and useful forecasts of volcanic eruptions, a central goal of volcanology.

While most eruptions are preceded by significant unrest, accurate predictions of volcanic activity, even at well-monitored volcanoes, remain elusive. One such well-monitored volcano is Kilauea, a large basaltic volcano located on the Big Island of Hawaii. Kilauea frequently erupts, posing a significant threat to the residents of the island.

In this work, we explore the use of earthquake records to predict volcanic activity at Kilauea. Physically, volcanic activity can alter the stress distribution in the crust, which can trigger earthquakes. Our models aim to predict Kilauea's eruptive status, with historical eruption data obtained from the compiled Kilauea eruption history provided by the Hawaii Center for Volcanology.

Our project effort involves two distinct tasks. First, we use the characteristics of each earthquake to predict whether Kilauea is erupting at the time of the earthquake, which we refer to as the "contemporaneous eruption classification task." Second, we perform a forward-looking forecast by utilizing the earthquake characteristics to predict the time until an eruption commences, which we refer to as the "time to eruption regression task." For the latter, the time is zero if an eruption is already underway.

3 Related Work

As with many scientific endeavors, previous research has laid the foundation for our work. Despite 50 years of concerted effort, accurate forecasts of volcanic eruptions remain a challenge [2, 3]. This point was tragically emphasized by the unexpected eruption of White Island in New Zealand in December 2019, which resulted in the loss of over a dozen lives despite constant monitoring of volcanic signals [4]. In rare cases where accurate forecasts have been made, success has often been achieved through the empirical identification of characteristic patterns in pre-eruptive seismicity and ground deformation. For instance, the 2000 eruption of Hekla was predicted with remarkable accuracy down to the minute by Icelandic scientists who observed a pattern of unrest that closely resembled the 1991 eruption of the same volcano [5]. Despite some successes in forecasting volcanic eruptions using statistical approaches [6], machine learning techniques have only been applied in one previous study [7], which showed that the eruptive state of Piton de la Fournaise could be determined by building a gradient boosted decision tree around the data from a single seismic station.

4 Dataset and Features

We obtained the earthquake catalog from the Advanced National Seismic System (ANSS), which is hosted by the World Volcano Database (WOVOdat). The earthquake catalog contains information on the time, longitude, latitude, depth, and magnitude of each earthquake. To extract earthquakes that are relevant to our study, we filtered the earthquakes by location along the East Rift Zone, where the eruptions occur at Pu‘u O‘o crater. This gave us a total of 3764 earthquakes in the catalog, as depicted in Figure 1.

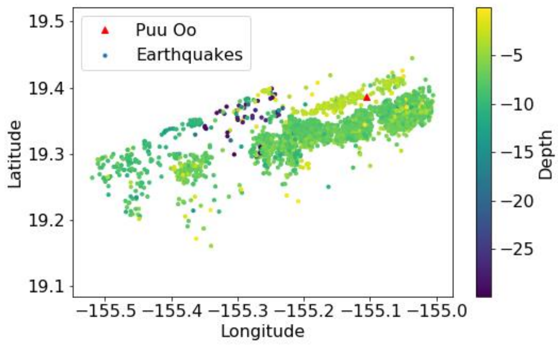

Figure 1: Map of earthquakes along the East Rift Zone of Kilauea. Colors indicate earthquake depths and the red triangle is Pu‘u O‘o.

In addition to the standard earthquake features, we derived an additional feature of earthquake rate to capture the physical interpretation that elevated volcanic activity should alter the stress in the crust and trigger earthquakes. Specifically, we calculated earthquake rates based on counts in the last day, last 7 days, and last 30 days.

The features used in our analysis include longitude, latitude, depth, magnitude, and earthquake rate. For labels, we either use erupting (1) or repose (0) in the classification part or time to eruption in the eruption forecasting part. To ensure accurate distance calculations for algorithms that employ distances (e.g. K-means), we scaled all the features and time to eruption to ensure they are distributed around mean 0 and standard deviation 1.

We divided the data into a training set (70% of the data), a development set (20%), and a test set (10%), while maintaining constant proportions of eruptions and repose across the three sets.

5 Methods

5.1 Logistic Regression

Logistic regression was only used in the contemporaneous eruption classification task using the features listed above.

5.2 K-Means Clustering

After dividing the training set into two subsets, one consisting of earthquakes during repose periods and the other during eruptive periods, we utilized the K-means clustering algorithm to perform our analysis. Through experimentation with different numbers of clusters, we found that the optimal number of clusters for both sets was 8 (as seen in Figure 2). In the time to eruption regression task, we utilized K-means clustering on the features and time to eruption, and determined that the ideal number of clusters was 10. To predict the time until an eruption in the development and test sets, we identified the closest cluster centroid to the features (excluding the time to eruption) and assigned the time to eruption as the centroid of the nearest cluster. This methodology proved to be a highly effective approach to our classification and regression tasks.

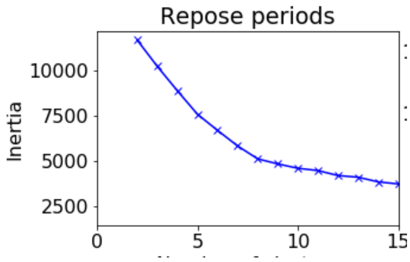

Figure 2: Inertia for different numbers of clusters for K-means clustering of earthquakes during repose periods. There is a clear elbow in the curve at 8 clusters.

5.3 Random Forest

We tried the Random Forest for both the classification and forecasting problems. This is due to its ability to account for non-linear relationships, thereby reducing bias, and also because averaging decorrelated trees helps to decrease variance [8]. To ensure optimal performance, we set the number of parameters considered at each split to int(√p) = 3. Hyperparameter tuning was then carried out to determine the optimal values for tree depth and the number of trees used for averaging, based on the AUROC scores for the classification problem (as shown in Figure 3) and R2 values for the regression problem.

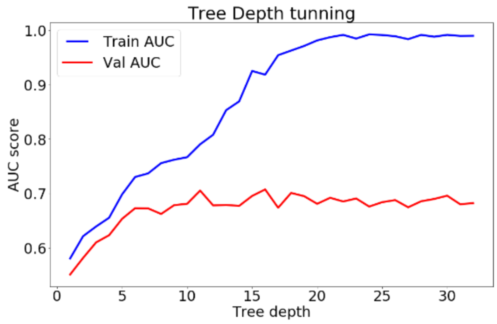

Figure 3: Hyperparameter tuning example for the random forest classification. The maximum tree depth in this case chosen to be 15 on the basis of the development set plateau.

5.4 Neural Network

As part of our analysis, we trained a four-layer fully-connected neural network for both classification and forecasting tasks. Each hidden layer of the network consisted of precisely one thousand nodes, resulting in a model containing roughly four million trainable parameters. Our models were built with the aid of PyTorch and implemented within the Google compute cloud platform to ensure optimal performance.

For the eruption classification task, we incorporated a final node with sigmoid activation and binary cross-entropy loss,

$$\ell_{\text{BCE}} = \frac{1}{N}\sum_i y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})$$

where N denotes the number of samples within a batch, and the hat symbol signifies model predictions. Conversely, for the forecasting task, we employed a linear activation function for the final node and mean squared error for loss.

$$\ell_{\text{MSE}} = \frac{1}{N}\sum_i \left(y^{(i)} - \hat{y}^{(i)}\right)^2 .$$

Both models were regulated via weight decay and trained utilizing stochastic gradient descent with a batch size of 50. We conducted numerous tests to fine-tune hyperparameters, including learning rate and weight decay parameter, by utilizing the validation set to ensure optimal model performance.

5.5 Metrices

Our dataset, as we have astutely observed, exhibits a significant class imbalance, with a staggering 80% of labels being negative (no eruption). Naturally, this presents a challenge in evaluating our models using classification accuracy alone. Therefore, we have chosen to utilize alternative metrics for the contemporaneous eruption classification task, namely Cohen's Kappa Coefficient and the Area Under the Receiver Operating Characteristic Curve (AUROC). The latter measures the probability of our model assigning a higher score to a random positive sample in comparison to a random negative sample, where a score of 0.5 denotes a meaningless model and a score of 1 signifies a perfect model. Cohen's Kappa, on the other hand, is a modified version of accuracy that considers the likelihood of an accurate test result occurring by random chance alone. Hence, a score of 0 represents a model performance that is expected purely by random chance, whereas a score of 1 indicates a perfect model.

As for the time to eruption regression task, we have opted to use root mean squared error and the R2 value for the plot of observed versus predicted times. We have determined that these metrics are the most suitable for accurately assessing the performance of our models in this particular task.

6 Experiments/Results/Discussion

6.1 Contemporaneous Eruption Classification Task

After careful analysis and comparison of different machine learning methods, we have determined that the random forest algorithm outperforms the others in classifying earthquake data to determine if a volcano is erupting or not. As shown in Figure 4, the random forest classifier achieved a Cohen's kappa score of 0.42, indicating moderate agreement [9]. Although the other methods showed some discriminatory capability in classifying earthquakes, they were not as successful on the test set as the random forest.

Surprisingly, the neural network achieved significantly lower performance than the random forest on this task. This could be attributed to several factors, such as the choice of loss functions that may not account for the imbalanced nature of the dataset, model architecture, or hyperparameters. Future research should explore these factors and potentially explore alternative approaches.

Algorithm	Dataset	Kappa	AUROC	Confusion matrix
Logistic Regression	Train	0.37	0.64	[[2165 45] [295 128]]
	Dev	0.25	0.59	[[630 17] [83 22]]
	Test	0.28	0.60	[[315 9] [42 13]]
K-means	Train	0.31	0.67	[[1895 315] [219 204]]
	Dev	0.31	0.67	[[559 88] [54 51]]
	Test	0.24	0.63	[[275 49] [32 23]]
Random Forest	Train	0.84	0.89	[[2043 0] [0 590]]
	Dev	0.46	0.66	[[575 9] [102 66]]
	Test	**0.52**	**0.71**	[[288 5] [50 36]]
Neural Network	Train	0.35	0.63	[[2025 18] [429 161]]
	Dev	0.27	0.60	[[572 12] [131 37]]
	Test	0.29	0.61	[[290 3] [67 19]]

Figure 4: Comparison of different machine learning methods to classify Kilauea earthquakes as occurring during eruption or repose.

6.2 Time to Eruption Regression Task

After thorough analysis, it is evident that the random forest outperforms other methods in predicting the time to eruption. However, it is worth noting that the correlation between observed and predicted times to eruption is still weak. The random forest model identifies earthquake rate and latitude as the most important features for predicting eruption time, as seen in Figure 6.

On the other hand, the neural network's performance is relatively poor in this task. To address this issue, we recommend exploring alternative loss functions that better account for the imbalanced nature of the dataset, alternative model architectures, and a more exhaustive search of hyperparameters.

In conclusion, while the random forest is the best-performing model for both eruption classification and forecasting, there is still room for improvement in predicting the time to eruption. Further research is needed to optimize the performance of machine learning models in predicting volcanic eruptions.

6.3 Summary

Our empirical findings demonstrate the inherent strength of random forest models in obtaining commendable performance on disparate tasks, namely regression and classification. On the other hand, our modified k-means technique exhibits superior classification prowess, but lags behind in the regression domain, which is not altogether unexpected given the method's inherently categorical outputs. The neural network, although moderately successful on both tasks, necessitates further refinement to achieve consistent and reliable outcomes. As such, our results underscore the importance of continued model optimization and exploration to achieve optimal performance.

Algorithm	RMSE	R^2
K-means (train/dev/ test)	0.95/0.95/1.02	0.097/0.059/0.037
Random Forest (train/dev/ test)	**0.16/0.68/0.64**	**0.82/0.37/0.38**
Neural Net (train, dev, test)	0.86/0.92/0.90	0.26/0.15/0.19

Figure 5: Comparison of different machine learning methods to forecast the time to eruption based on Kilauea earthquakes.

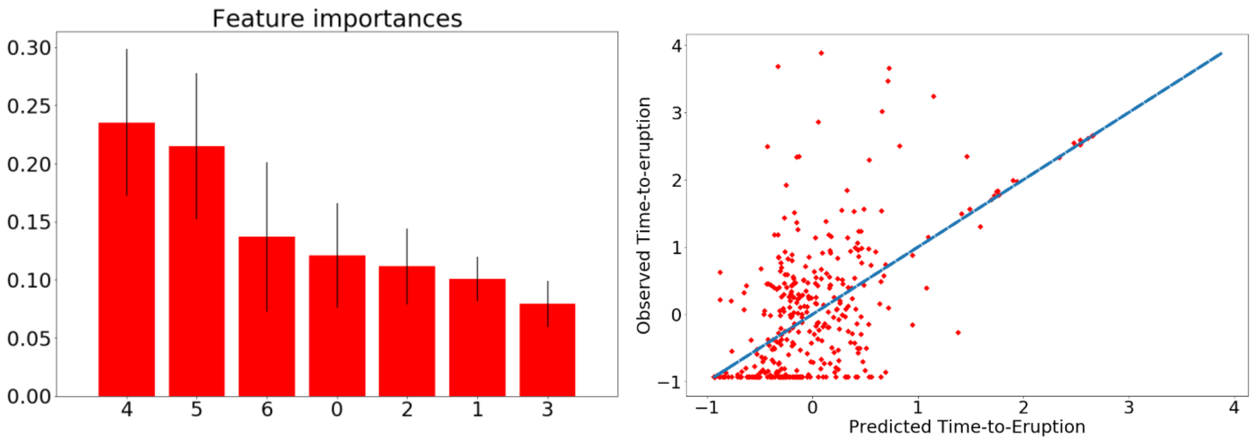

Figure 6: Random forest results with relative feature importance on the left and comparison of the predicted and observed time to eruption values on the right

7 Conclusion Future Work

Based on our analysis, it appears that certain machine learning methods, such as Random Forest, have the potential to accurately predict volcanic eruptions. However, the weak correlation between observed and predicted times to eruption suggests that additional features beyond interpreted catalogue earthquakes will need to be incorporated. We propose including earthquake focal mechanisms, low frequency ground deformation data, continuous background geo-phone records, and gas emission data to improve the accuracy of predictions. Additionally, we plan to expand the dataset to include a larger time period and consider incorporating data from other tectonically analogous volcanic systems. These efforts will undoubtedly provide more comprehensive insights into the nature of volcanic eruptions and their predictability.

9 Contributions

All team members contributed to the project effort. Mazin downloaded data and did random forest. Hilal made neural network. Nabeel did K-means clustering.

References

[1] Katharine V Cashman and R Stephen J Sparks. How volcanoes work: A 25 year perspective. GSA Bulletin, 125(5-6):664–690, 2013.

[2] R Steve J Sparks. Forecasting volcanic eruptions. Earth and Planetary Science Letters, 210(1- 2):1–15, 2003.

[3] RW Decker. State-of-the-art in volcano forecasting. Bulletin Volcanologique, 37(3):372–393, 1973.

[4] BBC. White island volcano: New zealand troops airlift six bodies off island. BBC, Dec 2019.

[5] Ragnar Stefánsson. Advances in earthquake prediction: research and risk mitigation. Springer Science & Business Media, 2011.

[6] Andrea Bevilacqua, Raffaele Azzaro, Stefano Branca, Salvatore D'Amico, Emanuela De Beni, Franco Flandoli, and Augusto Neri. Estimates and models of the statistical dependence between local earthquakes and flank eruptions at mt. etna volcano (italy): an old topic revised through new historical data. In AGU Fall Meeting 2019. AGU, 2019.

[7] Christopher X Ren, Aline Peltier, Valerie Ferrazzini, Bertrand Rouet-Leduc, PA Johnson, and Florent Brenguier. Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano. arXiv preprint arXiv:1909.12395, 2019.

[8] Daphne Koller, Nir Friedman, Sašo Džeroski, Charles Sutton, Andrew McCallum, Avi Pfeffer, Pieter Abbeel, Ming-Fai Wong, David Heckerman, Chris Meek, et al. Introduction to statistical relational learning. MIT press, 2007.

[9] Nicole J-M Blackman and John J Koval. Interval estimation for cohen's kappa as a measure of agreement. Statistics in medicine, 19(5):723–741, 2000.