e-ISSN :
2776-2521

JOCPES

Journal of Computation Physics and Earth Science
Volume 1 No. 1
April, 2021

Email:
adm.jocpes@gmail.com
admin@physan.org

https://journal.physan.org/index.php/jocpes

Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 1, Number 1, April 2021, Page 1 - 5
https://journal.physan.org/index.php/jocpes/index

Spatial Distribution and Implementation of the K-Means Clustering Method at Hotspots in North Sumatra

Sri Wahyuni [1]
[1]Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Article Info

Article history:

Received February 1, 2021
Revised February 16, 2021
Accepted March 11, 2021

Keywords:

hotspot
clustering
spatial
k-menas
sattelites

A B S T R A C T

Hotspots are indicators of forest and land fires. Hotspot monitoring can be carried out with the help of remote sensing tools and geographic information systems. Hotspot data is obtained from the MODIS sensors from the TERRA and AQUA satellites which contain information on latitude and longitude coordinates and the level of confidence divided by three levels, namely low, medium and high confidence levels. Based on the spatial results, the number of hotspots in North Sumatra Regency is in February, March, June, July, and August. Districts that are dominant with hotspots are Karo Regency, Labuhan Batu Regency, Mandailing Natal Regency, Padang Lawas Regency and South Tapanuli Regency. Based on the results, the process of applying the k-means clustering method to the weka application, the data obtained is in the form of a clustered group and the results can be made into indicators in determining hotspots in districts in North Sumatra province per month.

Corresponden Author:

Kartika Dewi,
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: sinambela.m@gmail.com

1. INTRODUCTION

Forest and land fires in Indonesia are an important note for the government because they produce haze and other impacts due to forest fires on the world community. Forest fires usually occur during the dry season. For the island of Sumatra, the provinces that experience high potential for forest fires are Riau, Jambi and South Sumatra every year. For the province of North Sumatra, there is the potential for forest fires in several districts. Various efforts have been made by the government to tackle these forest fires [1].

Hotspots can be interpreted as indicators of forest fires. These land fires can be identified using remote sensing technology [2]. One of the remote sensing technologies is the MODIS (Moderate resolution imaging spectroradiometer) sensor from the TERRA and AQUA satellites owned by NASA (National Aeronautics and Space Administration). These satellites record and monitor hotspots around the world in real time 24 hours. The TERRA satellite which carries the MODIS sensor is an environmental observation satellite that can be used for regional surface temperature data extraction [3].

Based on the results of the research that has been done, it can be seen that the analysis of the area of forest and land fires using remote sensing technology can be carried out using two approaches, namely digital and visual interpretation [4]. Hotspots can be validated by direct and indirect observation [5]. Direct observations are carried out by checking directly into the area where there are many hotspots.

Hotspot data obtained from the MODIS satellites contains information on the coordinates of the location of the area recorded by the satellite. The data of the coordinate location points were analyzed spatially and used the K-Means algorithm. The k-means method is one of the clustering methods which aims to divide the data set into clusters [6]. This method is used because the K-Means algorithm is easy to adapt and can be used for processing large datasets.

2. RESEARCH METHOD

Hotspots are hotspots that are indicated as points that have a higher temperature than the surroundings. In the hotspot data generated from the MODIS sensor, the hotspot is the result of heat detection on a certain pixel scale, for example 1 km x 1 km. Generally, the more hotspots that are detected in an area, the greater the possibility of forest fires in that area [7].

LAPAN (National Institute of Aeronautics and Space) has issued technical guidance on the level of confidence or confidence level of a hot spot to determine actions in dealing with it. Table 1 shows the meaning of the level of confidence in the data taken from the technical guidelines [7].

Table 1. Meaning of Trust Interval in Hotspot Information

Confidence Level (C)	**Class**
$0\% \leq C < 30\%$	Low
$30\% \leq C < 80\%$	Medium
$80\% \leq C < 100\%$	High

The K-Means algorithm is one of the most frequently used data grouping techniques. K-Means groups similar objects in the same cluster. An overview of the algorithm is as follows [9] determine the value of "K" or the number of clusters desired and determine the centroid or cluster center. The centroid is determined randomly from the existing data objects according to the value "K", if the value "K" is two then the centroid that is determined is also two. Divide each object into the most similar clusters. The most similar clusters are those with the closest centroid. Calculation of the proximity between the object and the centroid can be done using a distance function, such as Euclidean Distance. Recalculate cluster "K" with the average of all objects assigned to each cluster. Repeat the second and third steps until there are no more objects moving clusters [6].

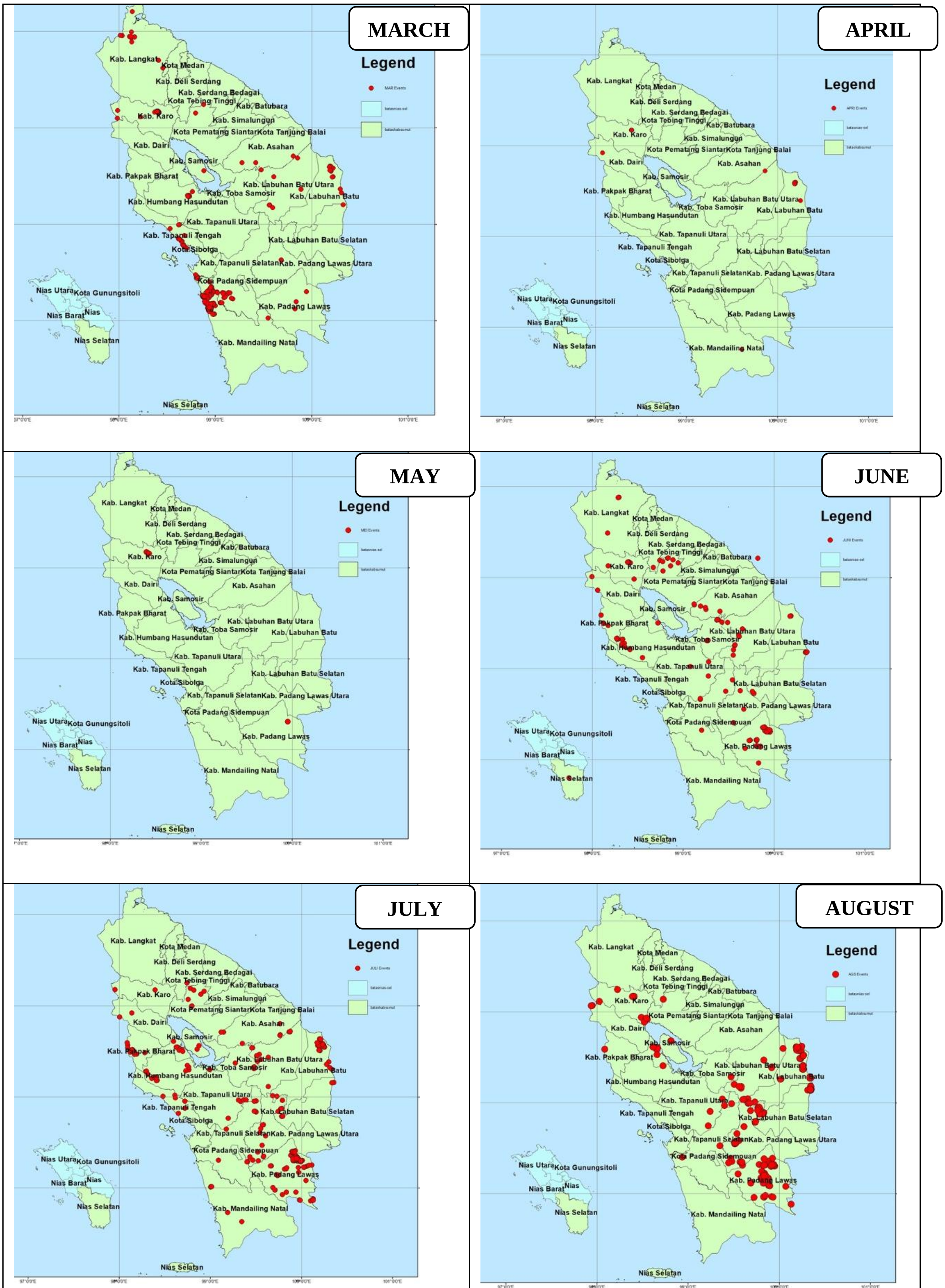
The formulas used for k-means are:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \{x_{ik} - x_{jk}\}^2} \tag{1}$$

Where: d: distance, i: i1, i2, i3, ……, ip, j: j1, ji2, ji3, ……, jeep, k: attribute value, n: dimension data, x: data object.

3. RESULTS AND DISCUSSION

In this study, using hotspots or hotspots to provide solutions to existing problems using spatial distribution and clustering using the k-means algorithm.


MARCH
APRIL
MAY
JUNE
JULY
AUGUST
Legend
Kab. Langkat
Kota Medan
Kab. Deli Serdang
Kab. Serdang Bedagai
Kota Tebing Tinggi
Kab. Batubara
Kab. Karo
Kab. Simalungun
Kota Pematang Siantar
Kota Tanjung Balai
Kab. Dairi
Kab. Asahan
Kab. Samosir
Kab. Pakpak Bharat
Kab. Labuhan Batu Utara
Kab. Toba Samosir
Kab. Labuhan Batu
Kab. Humbang Hasundutan
Kab. Tapanuli Utara
Kab. Tapanuli Tengah
Kab. Labuhan Batu Selatan
Kota Sibolga
Kab. Tapanuli Selatan
Kab. Padang Lawas Utara
Kota Padang Sidempuan
Nias Utara
Kota Gunungsitoli
Nias Barat
Nias
Kab. Padang Lawas
Nias Selatan
Kab. Mandailing Natal

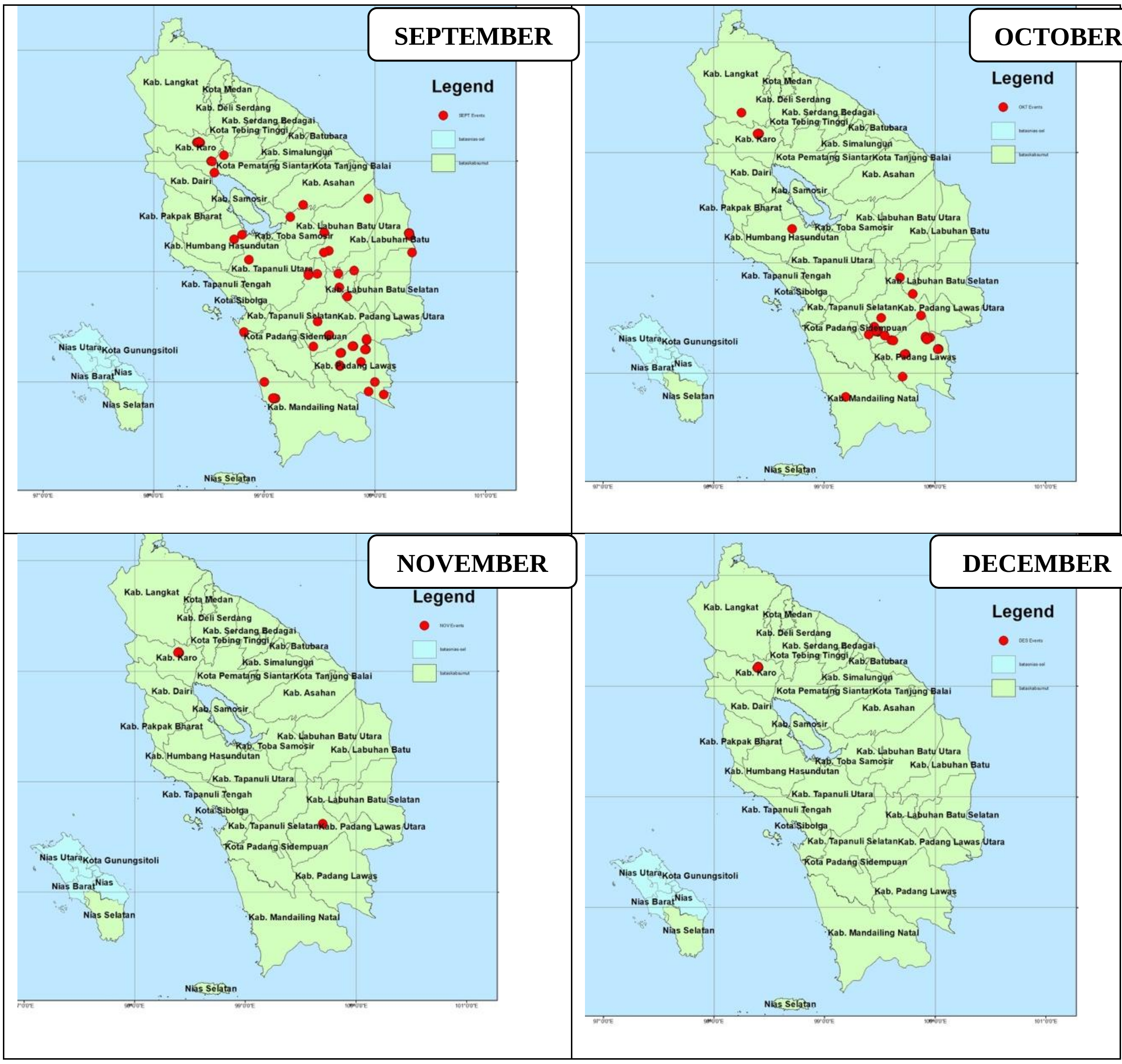


Figure 1. Distribution of cumulative monthly spatial hotspots in 2014 - 2020

The spatial results of hotspots using the Arcgis Map application can be seen that many hotspots in the districts of North Sumatra province occur in February, March, June, July, and August. In terms of climatological factors and environmental factors, it affects the number of hotspots each month. From climatological factors, the number of hotspots in North Sumatra is influenced by the reduced amount of rainfall that month.

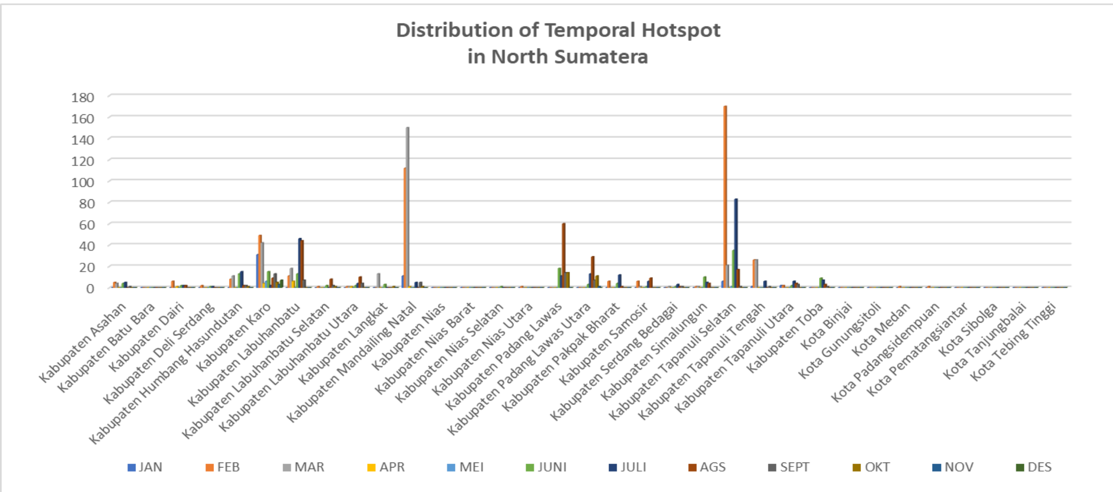


Figure 2. Graphics of monthly hotspot time series for 2014 – 2020

The distribution of temporal hotspots in the districts of North Sumatra province, there are several regencies where there are dominant hotspots per month. From Figure 2 it can be seen that there are 5 districts that have many hotspots per month, such as Karo Regency, Labuhan Batu Regency, Mandailing Natal Regency, Padang Lawas Regency and South Tapanuli Regency. The number of hotspots in the province of North Sumatra in the period 2014 to 2020, the highest peak occurred in 2014 in February.

After doing the mapping using the arcgis map, then using the k-means algorithm clustering application using the weka application. Using the Weka application by using sample data then creating the tested data and it will produce predictions of hotspots that have the possibility of occurring in the district.

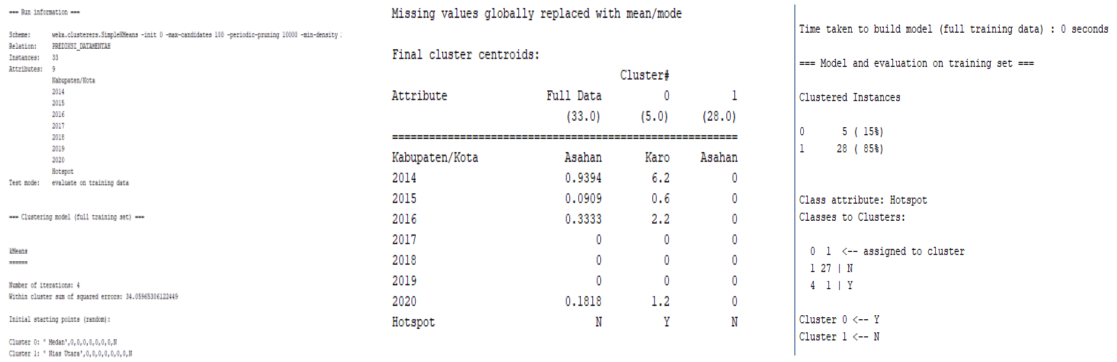


Figure 3. Implementation of the K-Means Clustering Method

In the implementation results using the k-mean clustering algorithm in weka, in January there are several districts that are predicted to have hotspots, namely Karo, Mandailing Natal and South Tapanuli districts.

4. CONCLUSION

Based on the analysis of that can be drawn from this research, namely, among others, based on the results of the spatial distribution of hotspots that are processed using the arcgis map, it can produce data on the number of distribution of hotspots that are dominant in districts in North Sumatra province. The dominant districts to watch out for in the dry season are Karo, Labuhan Batu, South Labuhan Batu, Mandailing Natal, South Tapanuli and Padang Lawas. From the results (output) obtained from the process of applying the k-means clustering method to the weka application, the data obtained is in the form of a clustered group and the results can be made into indicators in determining hotspots in districts in North Sumatra province per month.

REFERENCE

[1] N. A. Indrawan and H. A. Adrianto, “Spatio - Temporal Clustering Hotspot di Sumatera Selatan Tahun 2002 - 2003 Menggunakan Algoritme ST - DBSCAN dan Bahasa Pemrograman R Spatio - Temporal Clustering Hotspot in South Sumatera From 2002 - 2003 Using ST - DBSCAN Algorithm and R Programming Lan,” vol. 3, pp. 112–121, 2003.

[2] T. Handayani, A. J. Santoso, and Y. Dwiandiyanta, “Pemanfaatan Data Terra Modis untuk Identifikasi Titik Api Pada Kebakaran Hutan Gambut (Studi Kasus Kota Dumai Provinsi Riau),” Semin. Nas. Teknol. Inf. dan Komun., 2014.

[3] I. Prasasti, K. A. Sambodo, and I. Carolita, “Pengkajian Pemanfaatan Data TERRA-MODIS untuk Ekstraksi Data Suhu Permukaan Lahan (SPL) Berdasarkan Beberapa Algoritma,” Penginderaan Jauh, 2007.

[4] Suwarsono, Rokhmatuloh, and T. Waryono, “Pengembangan Model Identifikasi Daerah Bekas Kebakaran Hutan Dan Lahan (Burned Area) Menggunakan Citra Modis Di Kalimantan (Model Development of Burned Area Identification Using Modis Imagery in Kalimantan),” J. Penginderaan Jauh, vol. 10, no. 2, pp. 93–112, 2013.

[5] I. A. Csiszar, J. T. Morisette, and L. Giglio, “Validation of active fire detection from moderate-resolution satellite sensors: The MODIS example in Northern Eurasia,” IEEE Trans. Geosci. Remote Sens., 2006, doi: 10.1109/TGRS.2006.875941.

[6] R. A. Soelaiman Isye; Gosno, Eric Budiman, “Implementasi KD-Tree K-Means Clustering untuk Klasterisasi Dokumen,” J. Tek. ITS, vol. 2, no. Vol 2, No 2 (2013), pp. A432–A437, 2013, [Online]. Available: http://ejurnal.its.ac.id/index.php/teknik/article/view/3872.

[7] LAPAN, PANDUAN TEKNIS (V.01) INFORMASI TITIK PANAS (HOTSPOT) KEBAKARAN HUTAN/LAHAN. 2016.

Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 1, Number 1, April 2021, Page 6 - 9
https://journal.physan.org/index.php/jocpes/index

Average Temperature Forecasting Based on Deli Serdang Station Using Long Short-Term Memory Model

Ilham Junaedi[1], Endah Paramita[1], Nora Valencia Sinaga[1], Sri Wahyuni[1], Syahrul Humaidi[2]
[1]Badan Meteorologi Klimatologi dan Geofisika, Indonesia
[2]Department of Physics, FMIPA, Universitas Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received February 5, 2021
Revised February 20, 2021
Accepted March 14, 2021

Keywords:

Average temperature
Deliserdang Station
lstm

ABSTRACT

An understanding of designs and gage of typical temperature joined of parameter climate and climate data for better water resource organization and orchestrating amid a bowl is uncommonly imperative. Examine climate designs utilizing ordinary and neighborhood every year typical temperatures, compare and make discernments. amid this consider, we'll analyze adjacent and conventional typical temperature data in 96031 Station backed recognition station input. the preeminent objective of this considers to appear the execution of the conventional temperature in an exceedingly single station and to predict the ordinary temperature data utilizing the Long memory Illustrate approach. bolstered the comes about of standard informatics of exploring temperature with adjacent temperature relationship, we got the appear of preparing bend, remaining plot, and thus the diffuse plot is showed up utilizing these codes. the decent execution of 96031 Station had a Mean Squared Error esteem of 0.01 and R squared esteem 0.98, concerning zero will speak to superior quality of the indicator.

Corresponden Author:

Ilham Junaedi,
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: ilham_bmg@yahoo.co.id

1. INTRODUCTION

The 96031 station is one of the stations within the Medan zone which conveyed to watch the climate information. As a climate station, the 96031 stations deliver climate parameter information to educate the climate in-flight route zone, environment, climate issue, and horticulture [1], [2]. One of the vital climate information in climate issues is the temperature normal. The temperature is an imperative calculate in all stages of the climate issue, and the foremost imperative climate figure influencing fire behavior. Higher temperatures cruel that heatwaves are likely to happen more frequently and final longer. Hotter temperatures can moreover lead to a chain response of other varieties around the world [3], [4]. That's since expanding discussion temperature too influences the seas, climate designs, snow and ice, and plants and creatures. In this think about, we centered to assess the time-series of normal temperature collected from 96031 stations in Meteorology, Climatology and Geophysics Territorial 1, North Sumatera, Indonesia.

Long Short-Term Memory (LSTM) networks are a substantial branch of Recurrent Neural Networks (RNN), capable of learning long-term territories [5], [6]. LSTM is a model approach to solved some forecast issues. As a type of recurrent neural network, LSTM can learn the order dependence between items in a sequence. It also had the potential of being able to learn the context required to make predictions in time series forecasting problems, rather than having this context pre-specified and fixed [7], [8]

The most objective of this consider to analyze middle factors, assess and anticipate the time-series of normal temperature from 2008 until 2020 by utilizing the Long Short-Term Memory (LSTM) Demonstrate. The models were tried utilizing climate time arrangement information at each station which collected and compares the Root Cruel Square Mistake (RMSE) come about from the LTSM show. The comes about of this test appear that the LSTM demonstrate with/or without middle of the road variable has way better execution in case the time arrangement were accessible.

2. DATA AND METHOD

For this study, we utilized temperature data from the climate catalog of the Meteorology, Climatology, and Geophysics Regional I. The recorded temperature data was observed from 2008 to 2020, with the category of the average temperature on the Medan City as observed of 96031 stations. The selection of temperature series data provides a solid basis for evaluating the performance of observed that have been installed and reviewed in previous reports. The 96031station located with latitude and longitude coordinates are 3.61, 98.78. In this case, we focused on analyst the temperature data as time series recorded from each station in the Medan area and using LSTM to train, test, and predict the time series as a previous study [5]–[7].

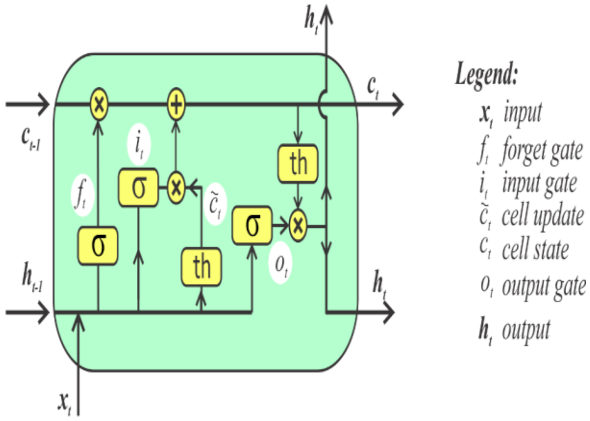


Figure 1. The LSTM Model [8]

In figure 1, as a brief depiction, the input of the LSTM cell could be a time arrangement set of information x that experiences a few sigmoid actuation doors σ. Each entryway calculates a certain work to calculate the cell states. What we are given is as it were an awfully brief clarification of how LSTM works. For more data with respect to LSTM, it is still partway better to require a lesson on profound learning to advance get it the concepts.

The LSTM cell comprises of an input door – which controls the stream of the input enactments into the memory cell, yield door which controls the yield stream of the cell actuation, and overlooks door, which channels the data from the input and past yield and chooses which one ought to be recollected or overlooked and dropped out. Other than three doors the LSTM cell contains cell upgrade which is more often than not tanh layer to be portion of the cell state. In each LSTM cell, the three factors are coming into the cell, the current input xt, past yield ht-1, and past cell state ct-1. On the other hand, from each LSTM cell, two factors are getting out the current yield ht and the current cell state ct.

In this section, you will learn about the concepts of mean squared error and R-squared. Mean squared error (MSE) is the average sum of the squared difference between the actual value and the predicted or estimated value. It is also termed as mean squared deviation (MSD) [9].

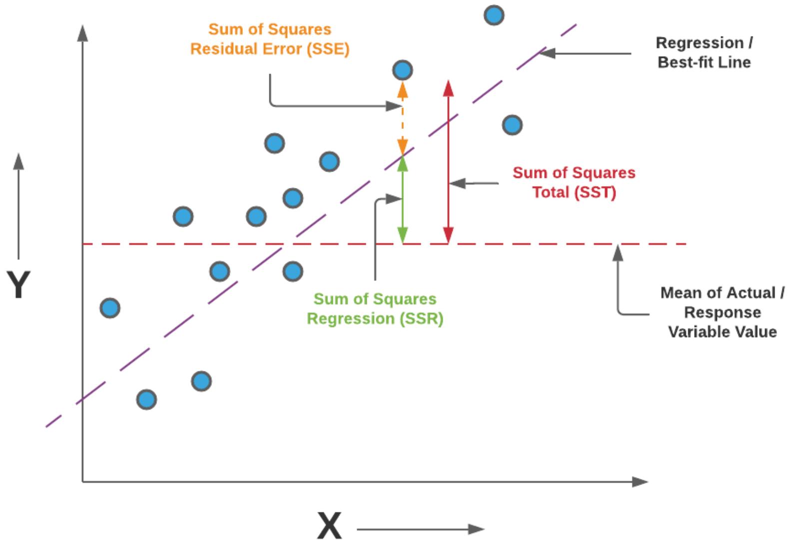


Figure 2. Diagrammatic representation for understanding R-Squared [9]

R-Squared can also be represented using the following formula in eq.1. The diagram and note that greater the value of SSR, more is the variance covered by the regression / best fit line out of total variance (SST).

$$R^2 = \frac{SST}{SST} = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(\hat{y}_i - \bar{y})^2} \tag{1}$$

Based on the diagram and note that smaller the value of SSE, smaller is the value of (SSE/SST) and hence greater will be value of R-Squared as show in eq.2.

$$R^2 = 1 - \left(\frac{SSE}{SST}\right) \tag{2}$$

3. RESULTS AND DISCUSSION

The initial information of normal temperature had been computed to see that we demonstrate forecast was effective. In any case, it can be watched from the anticipated (n days) that the blunders are ordinarily from the unexpected rise or decrease within the data such as in days 350-360. But, based on the primary 75 days, the model can appropriately take after the design of the information. The visual of all information of succinct Balai Besar (96041) can be appeared in fig.3.

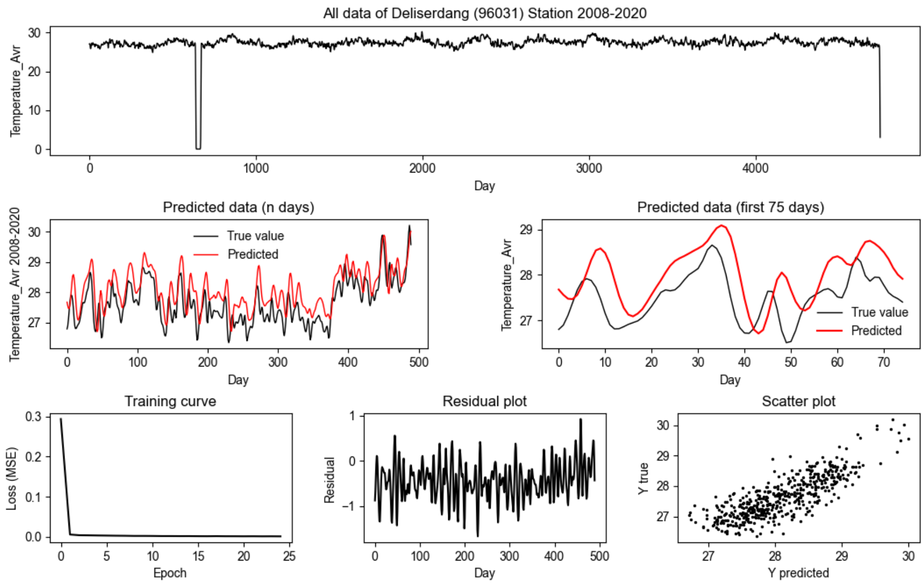


Figure 3. Stacked LSTM prediction results in 96031 Station

The preparing bend, remaining plot, and diffuse plot are appeared utilizing these codes. The representation of 96031 stations appears the esteem near to zero will speak to way better quality of the indicator within the relapse demonstrate. It can be seen from figure 3 that the Stacked LSTM performed best compared to the types of the station. We will also see that the single-cell worked incredibly utilizing 100 input days but we found that this kind of setup was as well computationally costly. The bidirectional LSTM moreover performed more awful with more input days. Based on the series data of temperature of 96031 stations, the MSE value is 0.36, and R2 value is 0.25.

4. CONCLUSION

In this study, we found that LSTM may be a great device for foreseeing normal temperature information. Based on the representation of 96031 stations that appeared, had an esteem near to zero will speak to superior quality of the estimator/predictor (relapse show). Be that as it may, ready to see from here that there are a few things to require domestic as lessons in utilizing LSTM. To begin with, more input days does not cruel that the demonstrate will be more precise. Other than that, temperature information conditioning may, offer-assistance in making the demonstrate more accurate. Finally, indeed in spite of the fact that we haven't appeared, LSTM needs a certain sum of information to be connected. From there, able to envision that LSTM can be utilized for anticipating climate and patterns of normal temperature.

REFERENCE

[1] B. Choubin, G. Zehtabian, A. Azareh, E. Rafiei-Sardooi, F. Sajedi-Hosseini, and Ö. Kişi, "Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches," Environ. Earth Sci., vol. 77, no. 8, pp. 1–13, Apr. 2018, doi: 10.1007/s12665-018-7498-z.

[2] F. S.-H. & Ö. K. Bahram Choubin, Gholamreza Zehtabian, Ali Azareh, Elham Rafiei-Sardooi, "Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches | SpringerLink," https://link.springer.com/article/10.1007/s12665-018-7498-z?shared-article-renderer. https://link.springer.com/article/10.1007/s12665-018-7498-z?shared-article-renderer (accessed Mar. 10, 2021).

[3] Wikepedia, "Effects of climate change - Wikipedia," https://en.wikipedia.org/wiki/Effects_of_climate_change, 2021. https://en.wikipedia.org/wiki/Effects_of_climate_change (accessed Mar. 10, 2021).

[4] NASA, "Global Warming," https://earthobservatory.nasa.gov/features/GlobalWarming, 2010. https://earthobservatory.nasa.gov/features/GlobalWarming (accessed Mar. 10, 2021).

[5] S. Afshin, H. Fahmi, A. Alizadeh, H. Sedghi, and F. Kaveh, "Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in karoon basin," Sci. Res. Essays, vol. 6, no. 6, pp. 1200–1208, 2011, doi: 10.5897/SRE10.448.

[6] M. I. Hutapea, Y. Y. Pratiwi, I. M. Sarkis, I. K. Jaya, and M. Sinambela, "Prediction of relative humidity based on long short-term memory network," AIP Conf. Proc., vol. 2221, no. March, 2020, doi: 10.1063/5.0003171.

[7] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Weather forecasting using merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model," J. Comput. Sci., vol. 14, no. 7, pp. 930–938, 2018, doi: 10.3844/jcssp.2018.930.938.

[8] Stanford, "Understanding LSTM Networks," https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf, pp. 1–13, 2015, [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[9] A. Kumar, "Mean Squared Error or R-Squared, Data Analytics," https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/, 2020. https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/ (accessed Mar. 10, 2021).

Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 1, Number 1, April 2021, Page 10 - 15
https://journal.physan.org/index.php/jocpes/index

Shear Wave Splitting Analysis Beneath Sumatra For-Arc Inferred from Broadband Seismic Network Station

Lewi Ristiyono[1], Chichi Nurhafizah[1], Suhenri Purba[1], Marhaposan Situmorang[2]
[1]Badan Meteorologi Klimatologi dan Geofisika, Indonesia
[2]Department of Physics, FMIPA, Universitas Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received February 5, 2021
Revised February 23, 2021
Accepted March 16, 2021

Keywords:

Shear Wave Spliting
Seismic observation
anisotropy

A B S T R A C T

The observation of broadband network seismic had been deployed in Sumatra For-Arc. The waveform data for this study were recorded from January 2014 – December 2016. The earthquake event data were selected with the epicenter of around 950 – 1800 in distance and Magnitude with more than 7 Mw. In this case, we use shear wave splitting to determine an anisotropic structure in Sumatra For-arc. Seismic Anisotropy can perform as a tool to classify and observe anisotropic structures of subsurface deformation processes beneath Sumatra For-Arc. The valid outcomes, in this case, have been gained that they only correspond to the upper layer, which has the delay time duration of 0.5 – 0.8 s is the anisotropic layer located in the Mentawai Island. The fast an anisotropic polarization direction found in Sumatra For-arc are parted into NE-SW direction found on the upper layer.

Corresponden Author:

Lewi Ristiyono,
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: lewi.r@yahoo.co.id

1. INTRODUCTION

Sumatra is section of the Indonesia archipelago which has special geological conditions. It is positioned between two massive plates, Indo-Australia Plate in the south and Eurasia Plate in the north. The subduction of these two plates results in one of the most energetic seismic things to do in the world (Figure 1). And if we seem to be the proportion of the earthquake in Indonesia (Figure 2), indicates Indonesia are is a very energetic area. The Sumatra Subduction Zone is one of the fantastic examples to find out about ocean-continental-subduction zone on a large scale.

Thus, Sumatra is considered as an perfect surroundings to study the reason of earthquake mechanism in a subduction zone is a vicinity where the Indo-Australia Plate is converging underneath the Eurasia Plate.

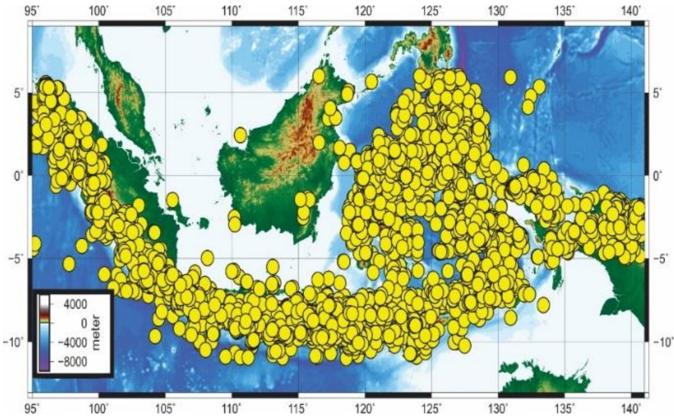


Figure 1. Seimicity of Indonesia

The Indo-Australian Plate is moving northward at a rate of 6-7 cm/year [1], [2]. This Subduction movement influences tectonic activity that occurs in Sumatra and some smaller islands around it.

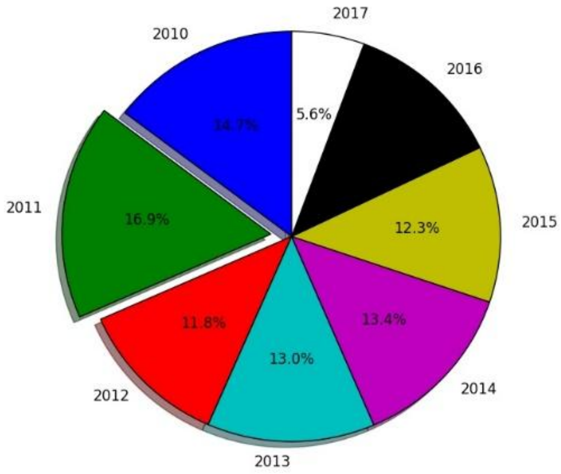


Figure 2. Percentage of Earthquake in Indonesia 2010-2017, August

A study by [3], [4] is one of the most detailed studies focusing on the subduction zone and fault in the north of Sumatra. In addition, a study by [5], [6] estimated anisotropic layers in Sumatra and Java by using a few stations. Consequently, it only involves some of the investigated regions. In this case, the Sumatra For-arc is a region that has a significant tectonic type. Up to this time, the study is still limited to an anisotropic observation which is close to the Sumatra For-Arc.

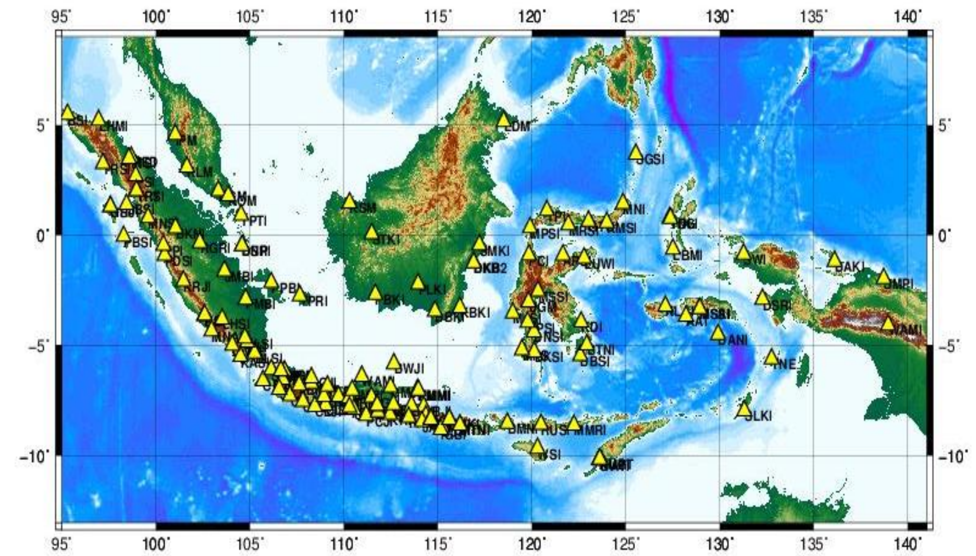


Figure 3. BMKG (IA) Seismic Network

These two factors are the heritage of the study of anisotropic shape in Sumatra. This is file is furnished with shear wave splitting commentary using facts taken from some seismic broadband station of the BMKG community in Indonesia (Figure 3). The statement is constant with the preceding studies [7]–[9]. Yet, it gives a denser density of earthquake monitoring station which deployed in Sumatra For-arc. Thus, it is feasible to supply a better constraint to decide an anisotropic shape in Sumatra For-arc. In this research we choose to center of attention on Sumatra Island, even though the preceding research have investigated a tectonic records of Sumatra Island in Figure 4, starting from the plate reconstruction to the tomography, the research have not explored but the subduction precipitated by the mantle motion.

2. DATA AND METHOD

This analyze about employed data that have been taken from six eternal stations of BMKG (IA) at the identical time as GFZ (Geo For-Schungs-Zetrum) neighborhood and BMKG Libra Networks used to be used to estimate the anisotropic characteristic below Sumatra For-Arc. Table 1 shows the neighborhood of earthquake monitoring stations. The statistics for this find out about was once earthquake records recorded from January 2014 – December 2016. The earthquake match data had been chosen with the epicenter of spherical 850 – 1800 in distance and Magnitude with greater than 7 Mw. The motive for deciding on the epicenter due to the fact the event gorgeous a shear wave splitting analysis on SKS/SKKS wave phase.

Table 1. Station Seismic Monitoring

No	Station	Lat	Long	Sensor
1.	SNSI	2.408	96.32	STS 2
2.	PBSI	-0.054	98.28	STS 2
3.	SISI	-1.326	99.08	STS 2
4.	PPSI	-2.7630	100.09	STS 2
5.	EGSI	-5.3524	102.27	STS 2
6.	GSI	1.3039	97.575	STS2, Triaxial

The distance additionally may want to warranty that the seismic wave nonetheless had ample energy [10]–[13]. Shear wave splitting used the teleseismic data, specially SKS/SKSS phase, ordinarily to reflect the anisotropic characteristics of the higher mantle [14], [15]. Figure 5 suggests most of the occasions that the BMKG station discovered in Sumatra For-arc and surrounding regions. Before conducting the shear-wave splitting analysis, first of all, the data of the SKS/SKKS wave section had been filtered, aiming to expand the signal-to-noise ratio (SNR) and steadiness of the shear-wave splitting measurement by way of the use of a bandpass filter with the frequency of 0.05- 4 Hz. This filtering resulted in, extra or less, the same splitting parameter. Consequently, it showed that the high-frequency scattering does now not influence the considerable result.

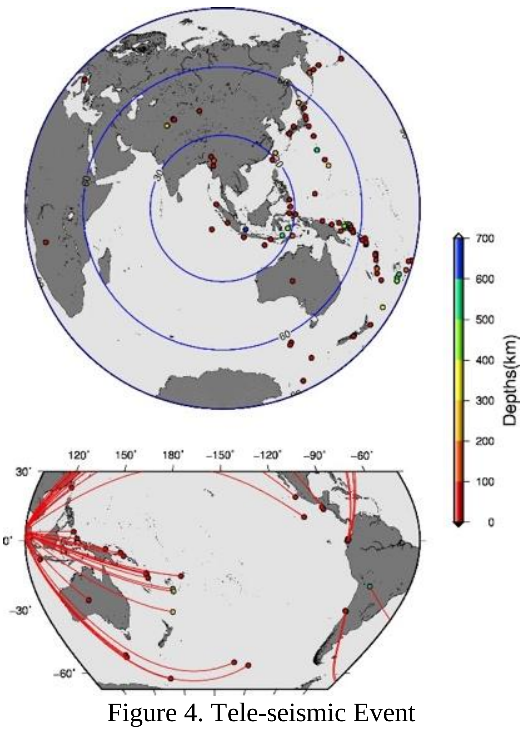


Figure 4. Tele-seismic Event

3. RESULTS AND DISCUSSION

Before plotting the parameter of the result of the shear-wave splitting size the use of a technique developed through [14], some seismograms recorded from the earthquake data that did not meet the standards had been deleted. A seismogram recorded from the earthquake statistics that did not meet the criteria had been deleted. A waveform of the earthquake that produced a extend time of extra than three seconds would be deleted. The prolong time usually takes place for round 1 or 2 seconds. Every delay time of more than three seconds would be regarded unrealistic to the appreciation of anisotropic media underneath the earth mantle [16], [17]. The seismic section was then checked visually and analyzed by the usage of a seismogram displayer as Figure 6 shows.

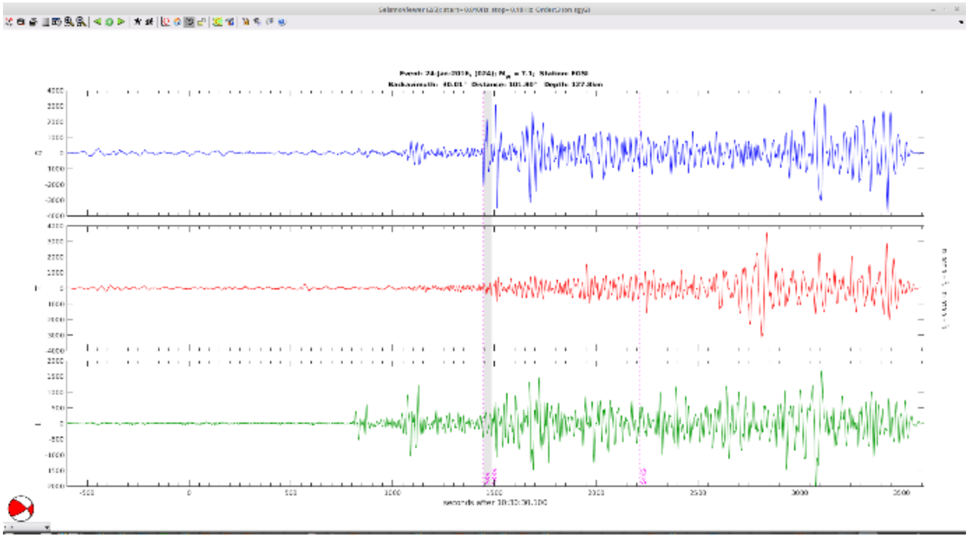


Figure 5. Seismogram viewer after changing into LTQ (Radial, transverse, and Vertical) and applied Band Pass Filter to reduce noise

The preparing bend, remaining plot, and diffuse plot are appeared utilizing these codes. The representation of 96041 stations appears the esteem near to zero will speak to way better quality of the indicator within the relapse demonstrate. It can be seen from figure 3 that the Stacked LSTM performed best compared to the types of the station. We will also see that the single-cell worked incredibly utilizing 100 input days but we found that this kind of setup was as well computationally costly. The bidirectional LSTM moreover performed more awful with more input days. Based on the series data of temperature of 96041 stations, the MSE value is 0.01, and R2 value is 0.98.

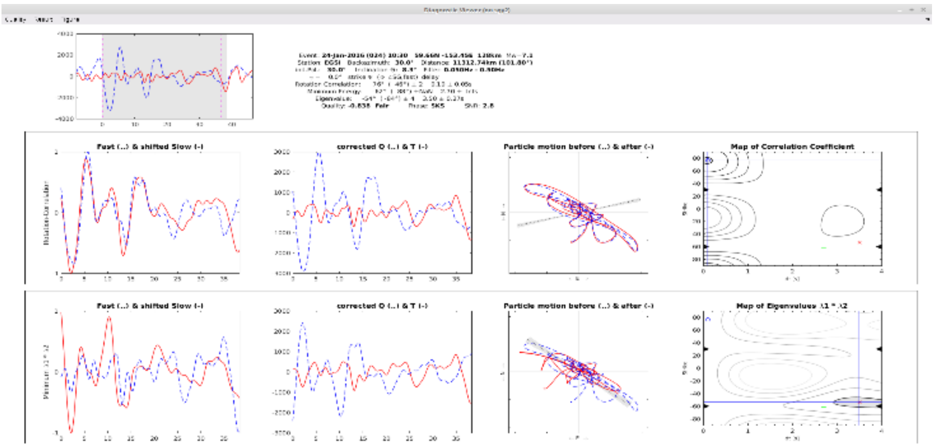


Figure 6. Example of splitting measurements of EGSI station obtained using Split-Lab

The observation of the seismic anisotropy in Sumatra For-Arc shows the end result of the shear-wave splitting calculation. For-Arc location is a place engulfing trench in the front of Sumatra For-Arc consists of six recording stations; SNSI, PBSI, SISI, PPSI, EGSI, and GSI. On the different hand, in For-Arc area of Mentawai Island, the end result of shear wave splitting measurements exhibit that there are two anisotropic layers; upper and lower layers. The decrease anisotropic layer produces the lengthen time at the price of about 1.4-1.8 s that possesses the identical polarization course of NE – SW (northeast-southwest), and it is perpendicular to the trench located in front of Sumatra For-Arc. The result of the shear-wave splitting measurement on the Mentawai Island indicates that the path is usually parallel with Indo-Australian Absolute Plate Movements.

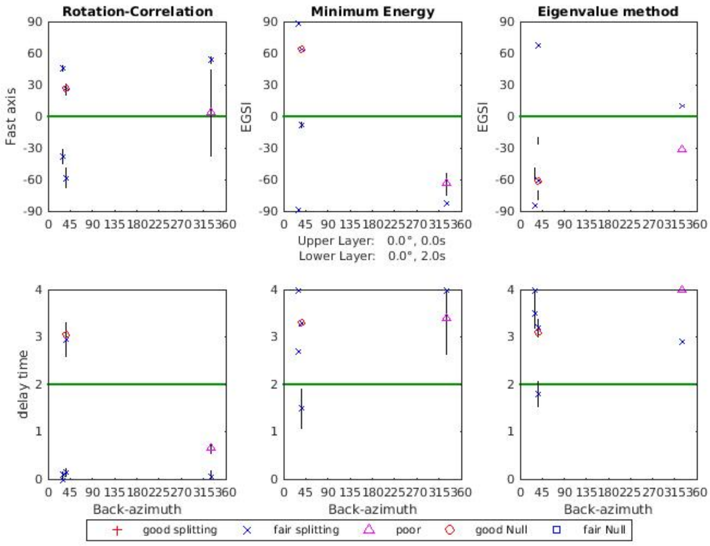


Figure 7. Example of double layer model fit at EGSI stations

The top left panel indicates the uncorrected radial (dashed) and transverse (solid) components. The middle and bottom rows of panels show the diagnostic plots for the rotation-correlation technique and the transverse aspect minimization method, respectively: from left to right, the corrected quickly (dashed) and sluggish (solid) components, the corrected radial (dashed) and transverse (solid) components, the uncorrected (dashed) and corrected (solid) particle motion diagrams, and the maps of correlation.

Three extraordinary splitting methods are compared (Rotation-correlation, Minimum Energy, and Eigenvalue). For every technique, the returned azimuthal version of speedy orientation estimate (top row) and lengthen time estimates (center row) are shown.

This result does now not swimsuit the international commentary of trench that had been conducted by way of [14]. However, this result is relevant to the result of studies performed through [5], [12], [16] which show that the result of the shear-wave splitting observation carried out on the Mentawai Islands possesses a polarization route that is parallel with the Absolute Plate Movements (APM). This discovering indicates that there are anisotropic plates underneath the Mentawai Islands.

These layers are anticipated to be subduction plates that exist below the Mentawai Islands. On the higher anisotropic layer, it confirmed the extend time at the rate of around 0.5-0.8 s that possesses a polarization direction which W-E (west-east) and parallel with the trench. This discovering potentially shows that there is a shallower anisotropic quarter than a subduction zone. This anisotropic layer is presumed upon a Mentawai Fault that exists below the Mentawai Islands. An appropriate rationalization describes the polarization course which is perpendicular to the trench is the anisotropic layers under the earth crust.

4. CONCLUSION

In this study, we current a new shear wave splitting end result in Sumatra For-Arc. The shear-wave splitting measurement from SKS/SKKS segment used to be conducted by using teleseismic facts taken from a everlasting station in Sumatra For-Arc. The valid effects have been won that they only correspond to six stations. The findings show that there are two foremost anisotropic layers fashioned in the For-Arc place (the Mentawai Island). The top layer, which has a extend time period of 0.5 – 0.8 s is the anisotropic layer placed in the Mentawai Island. The quick and anisotropic polarization path determined in Sumatra For-Arc is divided into the North East-South West direction located on the top layer.

REFERENCE

[1] R. Hall, “Australia-SE Asia collision: Plate tectonics and crustal flow,” Geol. Soc. Spec. Publ., vol. 355, pp. 75–109, 2011, doi: 10.1144/SP355.5.

[2] K. H. Liu, “NA-SWS-1.1: A uniform database of teleseismic shear wave splitting measurements for North America,” Geochemistry, Geophys. Geosystems, vol. 10, no. 5, 2009, doi: 10.1029/2009GC002440.

[3] E. Van Rijsingen, “Subduction interface roughness and megathrust earthquakes : Insights from natural data and analogue To cite this version : Subduction Interface Roughness and Megathrust Earthquakes,” 2018.

[4] R. McCaffrey, “The Tectonic Framework of the Sumatran Subduction Zone,” Annu. Rev. Earth Planet. Sci., vol. 37, no. 1, pp. 345–366, 2009, doi: 10.1146/annurev.earth.031208.100212.

[5] J. O. S. Hammond, J. Wookey, S. Kaneshima, H. Inoue, T. Yamashina, and P. Harjadi, “Systematic variation in anisotropy beneath the mantle wedge in the Java-Sumatra subduction system from shear-wave splitting,” Phys. Earth Planet. Inter., vol. 178, no. 3–4, pp. 189–201, 2010, doi: 10.1016/j.pepi.2009.10.003.

[6] J. F. Di Leo et al., “Mantle flow in regions of complex tectonics: Insights from Indonesia,” Geochemistry, Geophys. Geosystems, vol. 13, no. 12, pp. 1–20, 2012, doi: 10.1029/2012GC004417.

[7] K. Tarigan, M. Sinambela, A. T. Simanullang, H. Sunandar, and S. B. Sinaga, “The Characteristics Influence of the Seismic Signal Noise Using Spectral Analysis,” J. Phys. Conf. Ser., vol. 1116, no. 3, 2018, doi: 10.1088/1742-6596/1116/3/032041.

[8] K. Tarigan, M. Sinambela, M. Panjaitan, P. Simangunsong, and H. K. Siburian, “Machine Learning for Waveform Spectral Analysis on Nuclear Explosion Signal and Performance of Broadband Vertical Component,” J. Phys. Conf. Ser., vol. 1120, no. 1, pp. 1–6, 2018, doi: 10.1088/1742-6596/1120/1/012083.

[9] E. Darnila, M. Ula, K. Tarigan, T. Limbong, and M. Sinambela, “Waveform analysis of broadband seismic station using machine learning Python based on Morlet wavelet,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012048.

[10] A. Wüestefeld, “Methods and applications of shear wave splitting: The East European Craton,” Origins, pp. 1–57, 2007.

[11] G. Polat, N. M. Özel, and S. Crampin, “Preliminary investigating results from azimuthal seismic anisotropy beneath Western Anatolia and the hellenic subduction zone,” Yerbilim. Earth Sci., vol. 38, no. 3, pp. 229–240, 2017.

[12] E. Walsh, R. Arnold, and M. K. Savage, “Silver and Chan revisited,” J. Geophys. Res. Solid Earth, vol. 118, no. 10, pp. 5500–5515, 2013, doi: 10.1002/jgrb.50386.

[13] A. Sieminski, H. Paulssen, J. Trampert, and J. Tromp, “Finite-frequency SKS Splitting: Measurement and sensitivity kernels,” Bull. Seismol. Soc. Am., vol. 98, no. 4, pp. 1797–1810, 2008, doi: 10.1785/0120070297.

[14] M. D. Long and P. G. Silver, “Shear wave splitting and mantle anisotropy: Measurements, interpretations, and new directions,” Surv. Geophys., vol. 30, no. 4–5, pp. 407–461, 2009, doi: 10.1007/s10712-009-9075-1.

[15] Z. Huang, L. Wang, D. Zhao, N. Mi, and M. Xu, “Seismic anisotropy and mantle dynamics beneath China,” Earth Planet. Sci. Lett., vol. 306, no. 1–2, pp. 105–117, 2011, doi: 10.1016/j.epsl.2011.03.038.

[16] A. Wüstefeld, G. Bokelmann, G. Barruol, and J. P. Montagner, “Identifying global seismic anisotropy patterns by correlating shear-wave splitting and surface-wave data,” Phys. Earth Planet. Inter., vol. 176, no. 3–4, pp. 198–212, 2009, doi: 10.1016/j.pepi.2009.05.006.

[17] M. Assumpção, M. Guarido, S. van der Lee, and J. C. Dourado, “Upper-mantle seismic anisotropy from SKS splitting in the South American stable platform: A test of asthenospheric flow models beneath the lithosphere,” Lithosphere, vol. 3, no. 2, pp. 173–180, 2011, doi: 10.1130/L99.1.

Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 1, Number 1, April 2021, Page 16 - 20
https://journal.physan.org/index.php/jocpes/index

Spectral Probability Density for Broadband Seismic Ambient Noise Level

Endah Puspita[1], Eridawati[1]
[1]Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Article Info

Article history:

Received March 1, 2021
Revised March 10, 2021
Accepted March 23, 2021

Keywords:

Spectral
Ambient Noise
PDFPSD
Seismic
broadband

A B S T R A C T

The time-series approach is commonly utilized to get to the estimation of the likelihood thickness work of control ghostly densities (PDF PSD) of waveform information. This paper is concerned with the introduction of the evaluation of waveform commotion to degree the likelihood thickness work (PDF) be done inside, we utilized the metadata from a stock, a parser occurrence of DNP (Denpasar, Bali, Indonesia), IGBI (Ingas, Bali, Indonesia), and PLAI (Plampang, NTB, Indonesia) from BMKG IA-Networks and computations are based on the schedule utilized by McNamara Demonstrate. The point of this paper to characterize the current and past execution of the stations and recognizing the data on clamor levels at BMKG IA-Networks Station. The result of this paper shows the consistency of the unearthly is displayed the DNP, IGBI, and PLAI organize to confirm the quality of information conjointly acts as a test execution broadband arrange to the time taken by the broadband organize within the field and examination the Lombok earthquake in 2018.

Corresponden Author:

Nancy Damanik,
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: dewi90kartika@gmail.com

1. INTRODUCTION

Ambient noise level or background noise level is the contextual sound level sound pressure level at a given location in situ. The furthermost common source of ambient noise is from the exploit of human beings at or near the surface of the Earth. In the previous noise studies of the earthquake, there is no attempt to screen the continuous waveforms to eliminate body and surface waves from naturally occurring processing earthquake signals. In this case, they are low-probability occurrences even at low power levels (small magnitude events) related to the ambient surroundings at the seismic station.

The observed probability of the power spectral density as a tool to assess the field performance of earthquake monitoring system and the statistical distribution of noise levels across the frequency spectrum [1]. The more statistically approach requires large sample sizes, which become the norm as advances in probability power spectral density [2], [3]. In this study, we use the datasets of the broadband network from DNP, IGBI, and PLAI which deployed in BMKG network, Indonesia. Datasets on seismic background noise are collected, amongst others, for assessing the suitability of sites for permanent seismic recordings. One of the most important steps to be undertaken before the selection of suitable seismological recording sites is the investigation of seismic background noise.

The power spectral density spectrum estimation, periodogram is not a consistent estimate of the power spectrum [4], [5]. So, non-parametric methods are used for power spectrum estimation by smoothing and averaging operations which are directly implied on the periodogram or the autocorrelation [6], [7]. The aim of this paper to characterize the current and past performance and detecting the information on noise levels at BMKG IA-Networks Station.

2. DATA AND METHOD

The time series of waveform data were recorded in three locations in Indonesia: Denpasar-Bali, Ingas - Bali, and Plampang - Nusa Tenggara Barat (NTB) for one month in August 2018. The waveform data were employed from Indonesia Tsunami Early Warning System (Ina-TEWS) – BMKG-IA network in Table 1.

Table 1. Seismic Station of BMKG-IA Network and Location

No	Station	Latitude	Longitude	Networks
1.	PLAI-IA	-8.83	117.78	Indonesia
2.	IGBI –IA	-8.62	115.15	Indonesia
3.	DNP -IA	-8.68	115.21	Indonesia

In Table 1, shown the seismic station which recorded the SEED channel, the network code is IA (Indonesia), with channel code is BHZ (component Z).

The time series data in mini-SEED format obtained from the three stations were used for this study. An average of an hour-long vertical, Z component data for one month on August 2018 from each station and their respective noise spectra were plotted using spectral analysis methods. We use the power spectra to amplitude measurements [8]–[10]. The spectral analysis is commonly used to make the correction of attenuation and instrument displacement spectrum and determine the flat spectral level and corner frequency from which the seismic [6], [11]–[13].

The flow diagram of Probability Density Function and Power Spectral Density can show in Fig.1. The waveform data indicated the pre-processing by read of waveform data in hour.

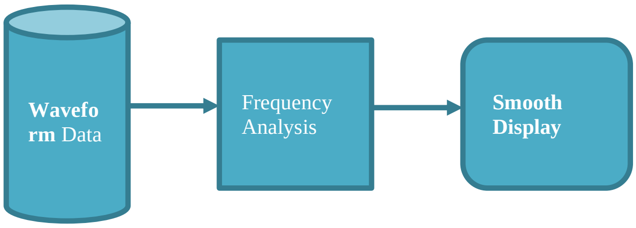


Figure 4. Tele-seismic Event

In this case, we remove instrument response and split into segment. The frequency analysis using FFT each segment and not only compute power for each segment but also compute average power for the hour interval. The smooth display shown the store of PSD and smoothing of power in octave.

3. RESULTS AND DISCUSSION

The figure [2-4] shows the result of probability power spectral density, which related plotted in power amplitude (dB) versus periods (s) [7], [12], [14]–[17]. The upper black solid line means that the New High Noise Model (NHNM) and the lower means that the New Low Noise Model. The DNP in figure 1 and PLAI in figure 3 networks are represented the spectral probability density closer to the upper line or the network means the new high noise model. In this case, the DNP and PLAI have a characterization of outliers and their influence on the noise level. The PSD distribution in Fig.2 shows the noise at the DNP, the minimum and mode power closely follow the especially in the frequency range above 1s.

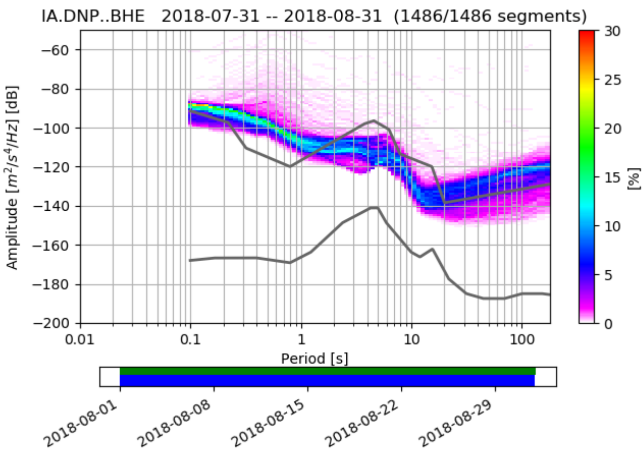


Figure 2. The PSD and PDF for DNP Broadband Seismic Network. The function is calculated using data from 1 to 30 August 2018

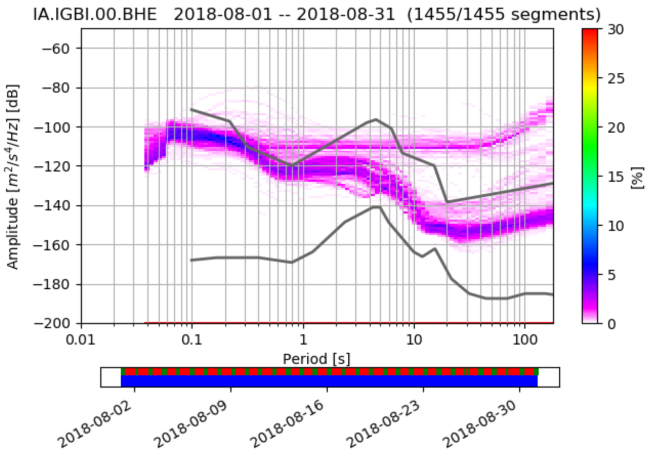


Figure 3. The PSD and PDF for IGBI Broadband Seismic Network. The function is calculated using data from 1 to 30 August 2018

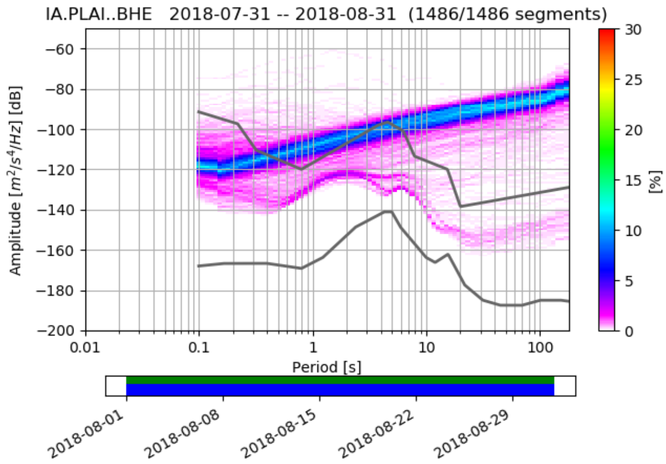


Figure 4. The PSD and PDF for PLAI Broadband Seismic Network. The function is calculated using data from 1 to 30 August 2018

In this case, the spectrum probability density of IGBI network is too flat, it means that there are several of the waveforms in August is missed, it's related to the gap of the waveform. The gaps can be represented in the period in Fig. 4. The other case, the spectral probability density is represented between the NHNM and NLNM. The green patches in periods represent available data, and red patches represent gaps in the stream that were added to the probability power spectral density [9][18][10].

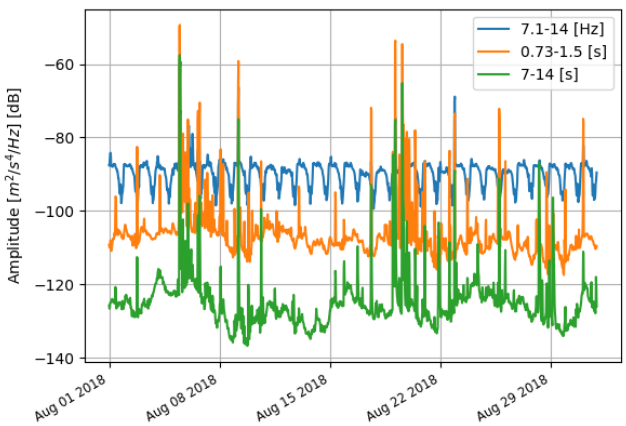


Figure 5. The Spectral Probability Density value of DNP

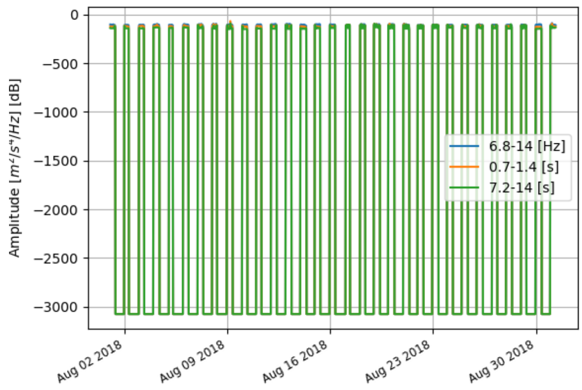


Figure 6. The spectral probability density value of IGBI

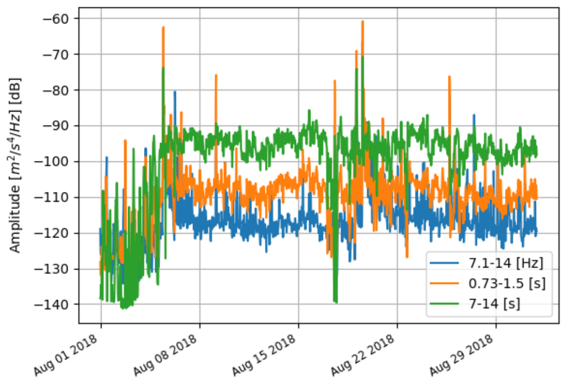


Figure 7. The Spectral probability density value of PLAI

In figure [5-7] show the information on metadata will be correctly picked for the respective start time of the data-trace. This means that DPN, IGBI and PLAI instrument changes are correctly taken into account during response removal. The time series of power spectral density of DPN, IGBI and PLAI network values can also be extracted from the probability power spectral density by accessing the property PSD values and plotted using the plot temporal method [19], [20].

4. CONCLUSION

The data used in the calculations are not separated out of earthquake, system and instrumental failures. The PSDPDF is determined power spectral of background noise level. We have presented the consistency of the spectral probability density is presented the DNP, IGBI, and PLAI network to verify the quality of data and also acts as a test performance broadband network to the time taken by the broadband network in the field and analysis the Lombok earthquake in 2018.

REFERENCE

[1] N. D. Merchant, T. R. Barton, P. M. Thompson, E. Pirotta, D. T. Dakin, and J. Dorocicz, “Spectral probability density as a tool for ambient noise analysis,” J. Acoust. Soc. Am., vol. 133, no. 4, pp. EL262–EL267, 2013, doi: 10.1121/1.4794934.

[2] K. Tarigan, M. Sinambela, A. T. Simanullang, H. Sunandar, and S. B. Sinaga, “The Characteristics Influence of the Seismic Signal Noise Using Spectral Analysis,” J. Phys. Conf. Ser., vol. 1116, no. 3, 2018, doi: 10.1088/1742-6596/1116/3/032041.

[3] E. Darnila, M. Ula, K. Tarigan, T. Limbong, and M. Sinambela, “Waveform analysis of broadband seismic station using machine learning Python based on Morlet wavelet,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, p. 012048, 2018, doi: 10.1088/1757-899X/420/1/012048.

[4] Y. Vaezi et al., “Seismic signals and noise,” Geophys. J. Int., vol. 2, no. 1, pp. 33–47, 2004, doi: 10.2312/GFZ.NMSOP-2.

[5] M. P. Mounika, K. Himaja, K. S. Ramesh, S. K. Rao, and T. V. Chandra, “Power spectrum analysis of seismic data for an earthquake using bartlett algorithm,” Int. J. Pure Appl. Math., vol. 114, no. 10 Special Issue, pp. 221–229, 2017.

[6] Á. Katalin, “Studying noise measurement and analysis,” Procedia Manuf., vol. 22, pp. 533–538, 2018, doi: 10.1016/j.promfg.2018.03.078.

[7] H. R. Gupta, R. Mehra, and S. Batan, “Power Spectrum Estimation using Welch Method for various Window Techniques,” Int. J. Sci. Res. Eng. Technol., vol. 2, no. 6, pp. 389–392, 2013, [Online]. Available: www.ijsret.org.

[8] P. Bormann, “Seismic signals and noise,” Bormann, P.(ur.), vol. 1, no. August, pp. 1–34, 2002, doi: 10.2312/GFZ.NMSOP-2.

[9] D. E. Mcnamara et al., “A Real-time Seismic Noise Analysis System for Monitoring Data Quality and Station Performance,” p. 2006, 2006.

[10] D. E. McNamara and R. P. Buland, “Ambient noise levels in the continental United States,” Bull. Seismol. Soc. Am., vol. 94, no. 4, pp. 1517–1527, 2004, doi: Doi 10.1785/012003001.

[11] M. M. Haney, J. Power, M. West, and P. Michaels, “Causal Instrument Corrections for Short-Period and Broadband Seismometers,” Seismol. Res. Lett., vol. 83, no. 5, pp. 834–845, 2012, doi: 10.1785/0220120031.

[12] A. el-aziz K. Abd el-aal and M. S. Soliman, “New Seismic Noise Models Obtained Using Very Broadband Stations,” Pure Appl. Geophys., vol. 170, no. 11, pp. 1849–1857, 2013, doi: 10.1007/s00024-013-0640-7.

[13] R. Andrés and P. Guridy, “Seismic Background Noise of Puerto Rico By,” 2008.

[14] M. Ula, E. Darnila, P. Siagian, L. Siagian, Peristiwanto, and M. Sinambela, “Machine learning on waveform spectral analysis of nuclear explosion from broadband seismic station in Indonesia,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, p. 012047, 2018, doi: 10.1088/1757-899X/420/1/012047.

[15] A. M. Reddy, B. Jayasree, S. K. Rao, and V. L. Bharathi, “Analysis of Power Spectrum Density on Earthquake Data Using Modified Covariance Algorithm,” vol. 114, no. 10, pp. 183–190, 2017.

[16] J. Berger, P. Davis, and G. Ekström, “Ambient Earth noise: A survey of the Global Seismographic Network,” J. Geophys. Res. Solid Earth, vol. 109, no. 11, pp. 1–10, 2004, doi: 10.1029/2004JB003408.

[17] K. U. Afegbua and F. O. Ezomo, “Evaluation of performance of z-component of Nigerian seismographic stations from spectral analysis,” vol. 8, no. 11, pp. 428–442, 2013, doi: 10.5897/IJPS2013.3833.

[18] D. E. McNamara, C. R. Hutt, L. S. Gee, H. M. Benz, and R. P. Buland, “A Method to Establish Seismic Noise Baselines for Automated Station Assessment,” Seismol. Res. Lett., vol. 80, no. 4, pp. 628–637, 2009, doi: 10.1785/gssrl.80.4.628.

[19] D. Rodríguez-Navarro, J. L. Lázaro-Galilea, I. Bravo-Muñoz, A. Gardel-Vicente, F. Domingo-Perez, and G. Tsirigotis, “Mathematical model and calibration procedure of a PSD sensor used in local positioning systems,” Sensors (Switzerland), vol. 16, no. 9, pp. 1–26, 2016, doi: 10.3390/s16091484.

[20] P. Hall and U. S. River, Spectral Analysis of Signals. 2009.

Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 1, Number 1, April 2021, Page 21 - 27
https://journal.physan.org/index.php/jocpes/index

Visualization of Fire Weather Index in Aek Godang Based on the Machine Learning Approach

Kerista Tarigan[1], Edison Kurniawan[2], Sri Wahyuni[2]
[1]Department of Physics, FMIPA, Univesitas Sumatera Utara, Medan, Indonesia
[2]Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Article Info

Article history:

Received March 9, 2021
Revised March 20, 2021
Accepted March 26, 2021

Keywords:

Fire Weather Index
Machine Learning
SVM

A B S T R A C T

Forest fires are a major natural issue, making temperate and environmental harm whereas dangering human lives. The examined and study for timberland fire had been worn out Aek Godang, Northern Sumatera, Indonesia. There are 26 hotspots in 2017 near Aek Godang, North Sumatera, Indonesia. In this consider, we utilize an information mining approach to prepare and test the information of woodland fire and Fire Weather Index (FWI) from meteorological information. The point of this ponders to anticipate the burned range and distinguish the woodland fire in Aek Godang ranges, North Sumatera. The result of this considers shown the Fire battling and avoidance movement may be one reason for the watched need of relationship. The reality that this dataset exists demonstrates that there's as of now a few exertions going into fire avoidance.

Corresponden Author:

Kerista Tarigan,
Department of Physics, FMIPA, Universitas Sumatera Utara, Medan, Indonesia
Email: kerista@usu.ac.id

1. INTRODUCTION

Every year Northern Sumatra of Indonesia spends hundreds of thousands in order to deal with the wildfire breakout. This situation now not solely motives monetary damage however can additionally disrupt the ecological stability by way of destroying vegetation and plants and fauna [1]. Wildfire is additionally accountable for air pollution and changes in climatic circumstance over the period of time [2]. Over the decade forest fire has turn out to be a major problem as it has endangered the lives of species. regardless of the massive charges concerned in controlling these dead fires, they are additionally a essential problem in forest fires [3]. The forests on the border of Aek Godang areas, North Sumatra had been badly affected and would be impacted by means of different areas in North Sumatra. The primary trouble of this study, how to computation the hotspot in this location to predict the woodland fire. Firefighters are conscious of how forest fires can be unpredictable [4]. However, if this data is obtained through them as a warning about the breakout in a timely manner then this form of phenomenon can be anticipated, controlled mainly can be prevented. There are many typical applied sciences that deal with wildfire hazard analysis. In this study, based on the description above, we are aiming to remedy this trouble via a historical analysis of woodland and land furnace facts and the usage of weather data to predict the extent of fires that have occurred. Then we also explored information mining strategies to locate out and predict the depth of wooded area and land fires [5]. Fast detection is a key component for controlling such a phenomenon. In achieving this, alternative options are needed. one of them is the use of nearby sensor-based automatic equipment furnished by using a number of meteorological stations [6]. causing meteorological conditions (such as temperature and wind) to affect wooded area and land fires, as well as knowing what a furnace index, such as the Fire Weather Index (FWI), makes use of this data. FWI is primarily based on the Index Spread Index (ISI) about the spread of furnace and wind speed, then the Buildup Index (BUI) to calculate the quantity of gasoline that reasons a fire. All of this is used as a measure for the well-known index of hearth hazard in woodland areas. In this work, we conducted statistics exploration with a information mining (DM) strategy so that we may want to predict the place of forest fires and burned land [7].

In this study we use Support Vector Machines (SVM)[9][10] and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), by carrying out tests on the latest real-world data, data collected from the northern Sumatera. The satisfactory configuration end result is the use of the SVM method with 4 meteorological enter parameters (namely relative humidity, rain, temperature, and wind) and is capable to predict burnt areas from several widely wide-spread small fires. So, this know-how is very supportive and useful in enhancing preventive motion and administration of firefighting sources (equipment and people).

2. DATA AND METHOD

The dataset of this study had been collected in BMKG of Aek Godang Station, North Sumatera from 2017 years [1], form the LAPAN based on the Satellite of NOAA [2] and from PKHL Direktorat Pengendalian Kebakaran Hutan [3]. There are more than 26 hotspots in the 2017 close to Aekgodang, Northern Sumatera was recorded [4].

The FWI file is a pointer measuring chimney escalated and combining the two going before components. In spite of the fact that the scale utilized is unmistakable for each issue of the FWI, the culminate fetched may to show additional extreme combustion conditions. At that point the distinctive imperative component is that the gasoline mugginess code requires memory (time slack) of the going before climate conditions: is 12 days for DMC, sixteen hours for FFMC, and 52 days for DC. Usually, a basic pointer in figuring out the profundity of lush region and arrive fires that take place.

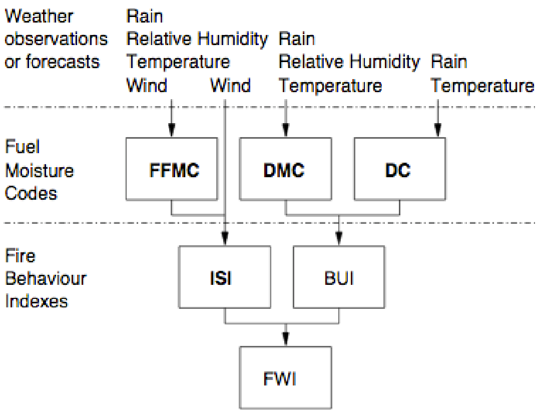


Figure 1. The Fire Weather Index structure [5]

A regression dataset D is made up of $k \in \{1,\ldots,N\}$ examples, each mapping an input vector $(x_1^k, \ldots\, x_A^k)$ to a given target y_k. The error is given by: $e = y_k - \bar{y}_k$, where $\bar{y}_k$ represents the predicted value for the k input pattern. The overall performance is computed by a global metric, namely the Mean Absolute Deviation (MAD) and Root Mean Squared (RMSE) [6], which can be computed as eq.1.

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|y_k - \hat{y}_k|$$
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (1)$$

In both metrics, lower values result in better predictive models. However, the RM SE is more sensitive to high errors. Another possibility to compare regression models is the Regression Error Characteristic (REC) curve, which plots the error tolerance (x-axis), given in terms of the absolute deviation, versus the percentage of points predicted SVM present theoretical advantages over NN, such as the absence of local minimum the model optimization phase. In SVM regression, the input $x \in \Re A$ is transformed into a high-dimensional feature space, by using a nonlinear mapping. Then, the SVM finds the best linear separating hyperplane in the feature space:

$$\hat{y} = w_0 + \sum_{i=1}^{m} w_i\phi_i(x) \qquad (2)$$

Where $\phi_i(x)$ represents a nonlinear transformation, according to the kernel function $K(x,x') = \sum_{i=1}^{m}\phi_i(x)\phi_i(x')$. To estimate the best SVM, the ϵ-insensitive loss function (Figure 4) is often used [6]. In presenting hyper parameters and less numerical difficulty than other kernels such as polynomials and sigmoid by using the popular Kernel Radial Basis Function.

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \gamma > 0 \tag{3}$$

The SVM performance is affected by three parameters: $C-$ a trade-off between the model complexity and the amount up to which deviations larger than ϵ are to related; ϵ – the width of the ϵ-insensitive zone; and γ – the parameter of the kernel. Since the search space for the three parameters is high, the C and ϵ values will be set using the heuristics proposed in: $C = 3$ (for standardized inputs) and $\epsilon = 3\hat{\sigma}\sqrt{\frac{\ln(N)}{N}}$ where $\hat{\sigma}$ is the standard deviation as predicted by a 3-nearest neighbor algorithm.

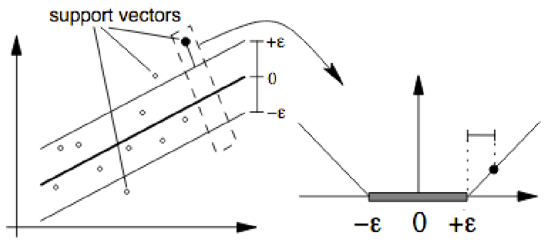


Figure 2. Example of a linear SVM regression and the ϵ-insensitive loss function

3. RESULTS AND DISCUSSION

Predicting fireplace burn of Aek Godang, Northern Sumatera region must assist in directing resources over large areas. An exceptionally interpretable model might provide records on hearth prevention. One may consequently be inclined to seem at multi-linear regression or generalized additive models.

The result of the split the statistics into coaching and trying out sets as shown in Figure 3.

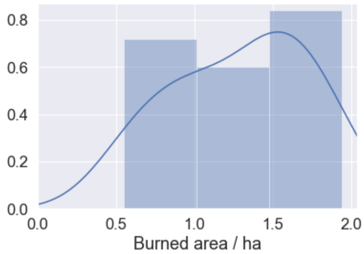


Figure 3. The distribution of the response variable

The response variable burned of Aek Godang, Northern Sumatera area, is extraordinarily skewed towards small fires. It might be beneficial to transform this with e.g. a Log10() scaling. The visual-spatial statistic result of this study proven in Figure four Most fires manifest at central and low X-Y coordinates, with the exception of one very high hearth count grid reference at (8, 6). Comparing complete fires with the total burned region there is some proof that fires at low X are small and numerous, where fires at high X are much less accepted however larger.

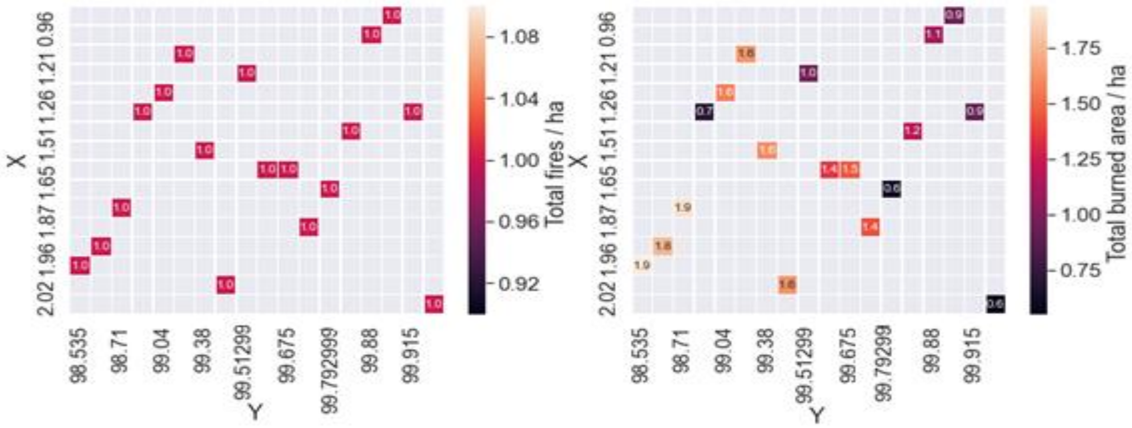


Figure 4. Visualize spatial statistics

The median burned region in Figure 5, reinforces the remaining bullet, that is to say, smaller fires dominate low-X regions the place large fires dominate at high-X regions.

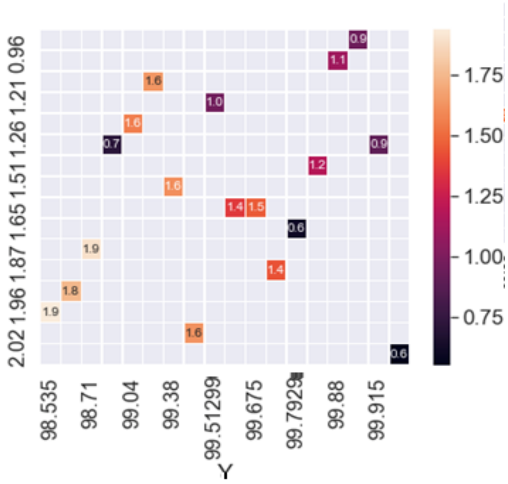


Figure 5. Median Burned

Based on the express the average burned area is biggest in March (Figure 6). However, this may also be the end result of a single or a few fires in view that the width of the distribution is small. The greatest fires have a tendency to occur in the summer time months, Aug via Sep. There is no obvious fashion in location burned on a given day of the week.

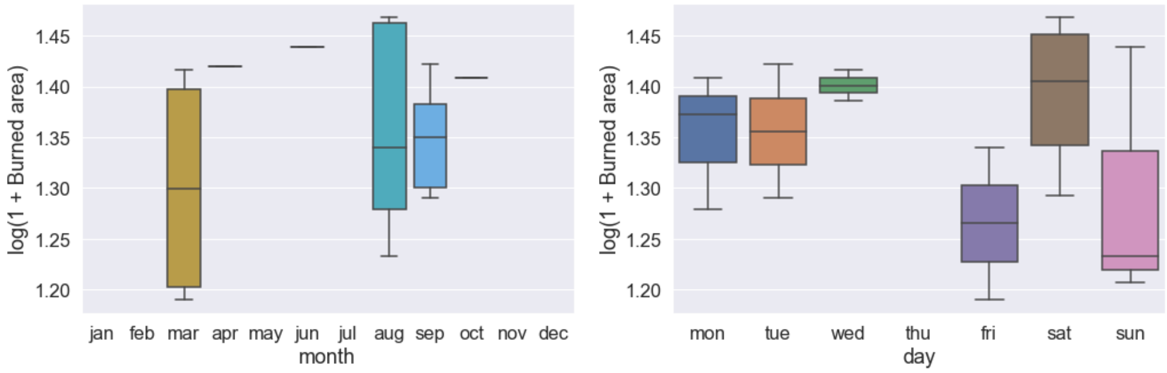


Figure 6. Categorical Variable

In Figure 7, the fire depend as a characteristic of month and day appears like most fires take place in the summer time months of August through September. Most fires manifest on the weekend, possibly pointing to human recreation.

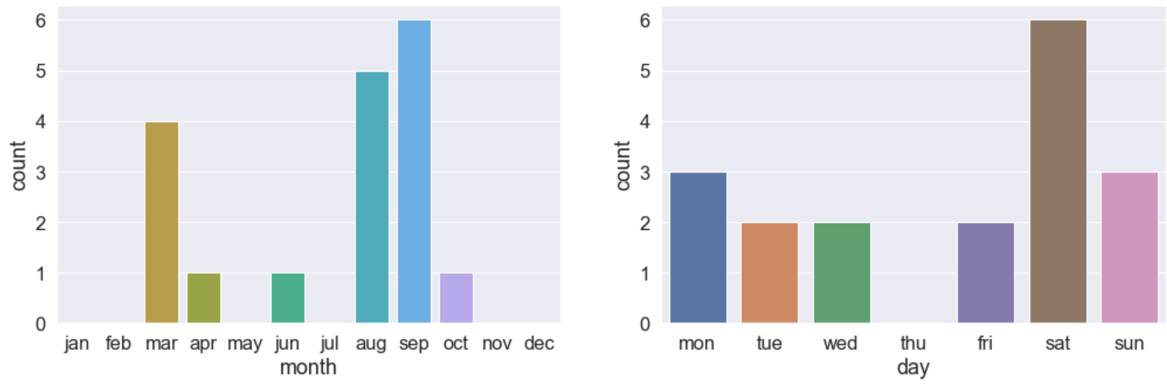


Figure 7. The fire count as a function of month and day looks like

The FWI symptoms are all correlated with one another and with temperature. There may also be some (multi) collinearity, which will amplify the variance of a geared up model. It may be beneficial to mix these into a single predictor. We'll stick with the full set of predictors for now

	X	Y	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
X	1.000000	-0.104349	-0.510968	0.163899	0.173017	-0.445531	-0.379804	0.525578	-0.320112	0.182948	0.152108
Y	-0.104349	1.000000	0.651102	-0.698352	-0.731524	0.603644	0.590436	-0.580939	0.239158	-0.241677	-0.697455
FFMC	-0.510968	0.651102	1.000000	-0.383860	-0.417636	0.966017	0.906448	-0.975714	0.234824	-0.463451	-0.367122
DMC	0.163899	-0.698352	-0.383860	1.000000	0.995787	-0.313409	-0.271251	0.274650	-0.212321	0.241341	0.991549
DC	0.173017	-0.731524	-0.417636	0.995787	1.000000	-0.339549	-0.307984	0.304645	-0.224787	0.251641	0.992462
ISI	-0.445531	0.603644	0.966017	-0.313409	-0.339549	1.000000	0.834487	-0.978134	0.215194	-0.337063	-0.293979
temp	-0.379804	0.590436	0.906448	-0.271251	-0.307984	0.834487	1.000000	-0.862024	0.186511	-0.519856	-0.236786
RH	0.525578	-0.580939	-0.975714	0.274650	0.304645	-0.978134	-0.862024	1.000000	-0.230477	0.425679	0.262224
wind	-0.320112	0.239158	0.234824	-0.212321	-0.224787	0.215194	0.186511	-0.230477	1.000000	-0.195281	-0.215478
rain	0.182948	-0.241677	-0.463451	0.241341	0.251641	-0.337063	-0.519856	0.425679	-0.195281	1.000000	0.216059
area	0.152108	-0.697455	-0.367122	0.991549	0.992462	-0.293979	-0.236786	0.262224	-0.215478	0.216059	1.000000

Figure 8. Correlation Matrix

Based on figure 8, the cross-validated imply absolute error from bagging is 0.09. Using default hyperparameters is not the strongest way to examine fashions in this way but we'll assume that the default hyperparameters are set to give practical starting points for most problems. This is very disappointing, the mannequin predicts a nearly regular response. There also appears to be a lower limit on the expected burned area. Does this mirror a lower restriction in the coaching data? It would be prudent to inspect this further.

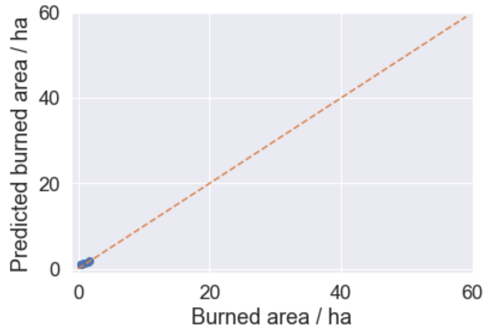


Figure 9. Test predictions against true burned area

The test set deviance increases beyond ~100 iterations, a clear signal that the model is overfitting. It would have been useful to do this checking out on a separate validation dataset as an alternative of the check set. This would have allowed us to go back and address the overfitting. Unfortunately, this is a very difficult dataset to work within that it is small with few if any predictors nicely correlated with the response. We would likely now not get any reward for similarly reducing the measurement of the coaching set.

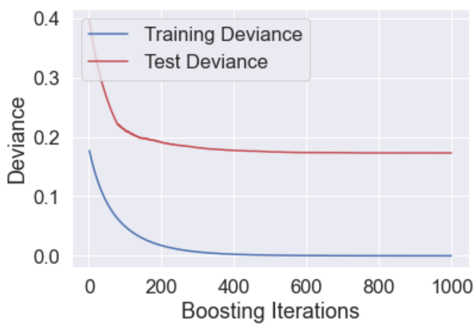


Figure 10. Training and test set deviance

All the fashions give comparable consequences and are tremendously poor, gradient boosting gives the lowest cross-validation error so we will take this forward and attempt to tune the parameters. There is no huge correlation between any one of the predictors and the response. A multilinear regression or generalized additive model is probably no longer going to eke out a signal. Highly nonlinear techniques might be better suited at the rate of interpretation.

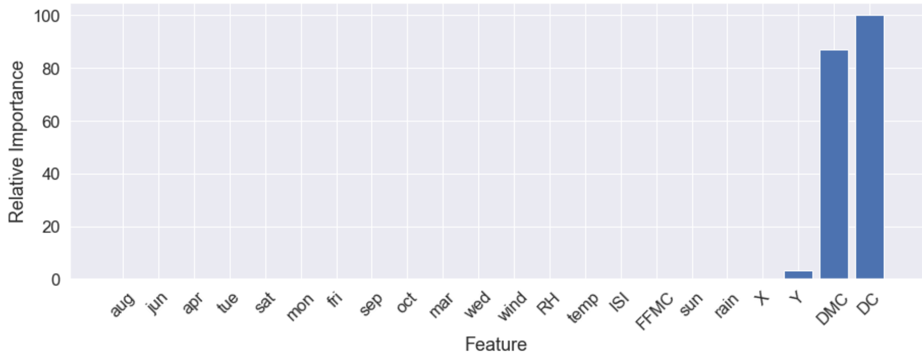


Figure 11. Feature importance

Many of the points have little or no significance in the closing model, they are probable adding noise, indicating that some feature selection might be prudent. The temperature has the easiest significance of all the features, which makes a lot of sense. However, each wind and rain have no importance. One might have expected fires to burn much less vicinity at instances of high precipitation and for high winds to fan the flames.

The wooded area fires dataset used to be presented in Cortez and Morais 2007 [6], the place the authors current an answer to this trouble the use of a trained support vector machine. In assessing the accuracy of their mannequin they produce a REC curve, which plots the error tolerance (x-axis), given in terms of the absolute deviation, versus the percentage of points envisioned in the tolerance (y-axis). The ideal regressor should existing a REC region close to 0.5.

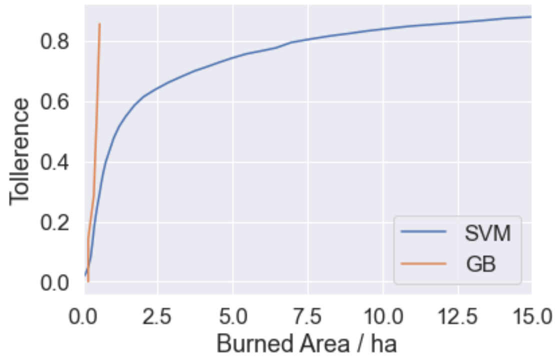


Figure 12. REC Curve

4. CONCLUSION

Based on the result, there is little correlation between the predictor variables and the response variable. It would be prudent to attempt and understand this lack of correlation better. Fire hostilities and prevention exercise may be one motive for the found lack of correlation. The truth that this dataset exists suggests that there is already some effort going into hearth prevention. One ought to imagine a situation in which many fires have the ability to emerge as very massive but are extinguished before they have the risk to do so. If statistics pertaining to furnace prevention is reachable it would likely be an extraordinarily precious addition to this dataset. It was once shown that the gradient boosting model used to be likely overfitting. Controlling the depth of timber and studying charge are two methods which were used to stop overfitting. Scikit-learn gives numerous more, which includes the capacity to enforce a decrease bound on the number of samples in a leaf. This limits the ability of the boosting algorithm to structure leaves that seize single outlying data points, hence decreasing variance and overfitting. As with random forests, introducing randomization into the boosting algorithm can additionally minimize variance. Scikit-learn affords two methods. First by using developing each tree with a random subsample of the education set and 2nd via randomly subsampling the points viewed for each node. In summary, a whole lot greater tuning of the mannequin is possible.

Gradient boosting based totally on the cross-validated mean absolute error from tuned gradient boosting is 0.07, it performs characteristic selection naturally. However, with the use of a validation set, it would have been feasible to use the feature importance plot above to do some guide characteristic selection. In particular, most of the days and months have no relevance to the problem and are probably simply including noise. Unfortunately, the use of a validation set for this motive would always reduce the coaching data, in addition contributing to the situation of making an attempt to eke out a susceptible sign from a small dataset.

REFERENCE

[1] BMKG, “Badan Meteorologi, Klimatologi dan Geofisika,” 2020. .
[2] LAPAN, “Lembaga Penerbangan dan Antariksa Nasional,” 2020. .
[3] D. P. K. H. PKHL, “SiPongi Karhutla Monitoring Sistem,” Jakarta, 2019.
[4] D. Bidang and P. Jauh, INFORMASI TITIK PANAS (HOTSPOT) KEBAKARAN HUTAN / LAHAN. 2016.
[5] Nrcan, “Canadian Wildland Fire Information System | Canadian Forest Fire Weather Index (FWI) System,” https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi, 2020. .
[6] P. Cortez and A. Morais, “A Data Mining Approach to Predict Forest Fires using Meteorological Data,” Proc. 13th Port. Conf. Artif. Intell., no. January 2007, pp. 512–523, 2007.