



Terbit *online* pada laman web jurnal :
<http://ejournal.amikompurwokerto.ac.id/index.php/telematika/>

Telematika

Terakreditasi Sinta “3” KEMENRISTEKDIKTI, No. 21/E/KPT/2018



Optimasi Algoritme *Naïve Bayes* untuk Klasifikasi Data Gempa Bumi di Indonesia Berdasarkan Hiposentrum

Rastri Prathivi

Teknik Informatika
 Fakultas Teknologi Informasi dan Komunikasi
 Universitas Semarang
 Email : vivi@usm.ac.id

INFO ARTIKEL

Sejarah Artikel:

Menerima 21 November 2019
 Revisi 13 Januari 2020
 Received 27 Februari 2020
 Online 29 Februari 2020

Keyword:

Earthquake Classification
 Hypocenter
 Adaboost
 Naïve Bayes

Kata Kunci:

Klasifikasi Gempa
 Hiposentrum
 Adaboost
 Naïve Bayes

Korespondensi:

Telepon: +62 085950251065
 E-mail: vivi@usm.ac.id

ABSTRACT

The Hiposentrum or epicentre is the source of an earthquake which is at a certain depth on earth. The classification of earthquake powers based on the depth of Hiposentrum needed to examine the potential earthquake powers spread in Indonesian territory. The results of the classification process often experience problems, namely inaccuracy in classification. To solve that problem, then algorithms optimising classification must be increased. This research uses the Naïve Bayes algorithm, which is optimized using the Adaboost algorithm. Evaluation of the results of the optimized classification algorithm is needed to determine the level of accuracy using prescriptions and recall. In this study, the object of research is earthquake data in Indonesia which will be used as training data and testing data. The average accuracy of the Naïve Bayes algorithm is 72.3%, and the Naïve Bayes and Adaboost algorithm is 85.3%.

ABSTRAK

Hiposentrum atau pusat gempa merupakan sumber gempa yang terdapat pada kedalaman tertentu di bumi. Klasifikasi kekuatan gempa berdasarkan kedalaman hiposentrum diperlukan untuk mengetahui potensi kekuatan gempa yang tersebar di wilayah Indonesia. Hasil dari proses klasifikasi seringkali mengalami masalah yaitu ketidaktepatan dalam klasifikasi. Untuk mengatasi masalah tersebut maka algoritme klasifikasi perlu ditingkatkan optimasinya. Penelitian ini menggunakan algoritme *Naïve Bayes* yang dioptimasi menggunakan algoritme *Adaboost*. Evaluasi terhadap hasil dari algoritme klasifikasi yang telah dioptimasi diperlukan untuk mengetahui tingkat akurasi menggunakan *presicion* dan *recall*. Dalam penelitian ini objek penelitian berupa data gempa bumi di Indonesia yang akan digunakan sebagai data *training* dan data *testing*. Hasil rata - rata akurasi algoritme *Naïve Bayes* sebesar 72,3% dan algoritme *Naïve Bayes* dan *Adaboost* sebesar 85,3%.

PENDAHULUAN

Hiposentrum atau pusat gempa merupakan sumber gempa yang terdapat pada kedalaman tertentu di bumi. Kekuatan suatu gempa sangat tergantung pada lokasi hiposentrumnya. Menurut Hartuti (2009) klasifikasi gempa berdasarkan hiposentrumnya diklasifikasikan menjadikan tiga bagian yaitu gempa bumi dalam adalah gempa bumi yang terjadi dengan kedalaman hiposentrum >300 km di bawah permukaan bumi. Gempa bumi menengah yaitu gempa bumi yang terjadi dengan kedalaman hiposentrum berkisar antara 60 km sampai 300 km di bawah permukaan bumi. Gempa bumi dangkal yaitu gempa bumi yang terjadi dengan kedalaman hiposentrum <60 km di bawah permukaan bumi. Berdasarkan letak wilayah geografisnya Indonesia adalah negara kepulauan yang memiliki potensi gempa bumi yang besar.

Data pada Badan Meteorologi, Klimatologi dan Geofisika (BMKG) mencatat terjadinya gempa bumi di Indonesia secara berkala dalam selang waktu beberapa bulan saja. Bahkan potensi gempa bumi yang lokasi hiposentrumnya cukup dalam hampir terjadi setiap hari.

Klasifikasi kekuatan gempa berdasarkan kedalaman hiposentrum diperlukan untuk mengetahui potensi kekuatan gempa yang tersebar di wilayah Indonesia. Hasil dari proses klasifikasi seringkali mengalami masalah yaitu ketidaktepatan dalam klasifikasi. Untuk mengatasi masalah tersebut maka algoritme klasifikasi perlu ditingkatkan optimasinya. Menurut Saritas (saritas, 2019) algoritme *Naïve Bayes* merupakan algoritme probabilitas sederhana untuk mengklasifikasikan data berdasarkan probabilitas data dengan menghitung frekuensi dan kombinasi nilai pada dataset yang digunakan. Di dalam algoritme *Naïve Bayes* setiap data diasumsikan sebagai variabel bebas yang dapat mempertimbangkan nilai dari variabel klasifikasi. Pada penelitian yang dilakukan oleh Saritas (saritas 2019) performa akurasi algoritme *Naïve Bayes* sebesar 83.54 terhadap algoritme ANN 86.95 yang digunakan untuk mengklasifikasikan data kanker payudara. Dalam penelitian yang ditemukan oleh Nakra (2019) algoritme *Naïve Bayes* memiliki tingkat akurasi yang cukup baik yaitu 79.6935 terhadap algoritme Bayes Net sebesar 78.5441 yang dihitung dengan tool WEKA. Dalam penelitian tersebut algoritme *Naïve Bayes* masih dimungkinkan untuk ditingkatkan akurasinya dengan mengoptimasi algoritme *Naïve Bayes* tersebut. Dalam penelitian ini penulis yang akan meningkatkan optimasi algoritme *Naïve Bayes* yang diterapkan untuk mengklasifikasikan dataset gempa bumi berdasarkan hiposentrumnya. Untuk meningkatkan optimasi algoritme *Naïve Bayes*, penulis menggunakan algoritme *Adaboost*. Evaluasi terhadap hasil dari algoritme klasifikasi yang telah dioptimasi diperlukan untuk mengetahui tingkat akurasi menggunakan *presicion* dan *recall*. Dalam penelitian ini objek penelitian berupa data gempa bumi di Indonesia yang akan digunakan sebagai data *training* dan data *testing*. Berbagai penelitian yang terkait dengan klasifikasi kekuatan gempa dan lokasi gempa telah banyak dilakukan oleh para peneliti sebelumnya. Sari (2018) menggunakan algoritme K-AP Clustering untuk mengklasifikasikan gempa bumi di Indonesia. Fatichah (2017) mengklasifikasi data gempa berdasarkan data dari twitter menggunakan algoritme Decision Tree, Random Forest dan SVM. Halim (2017) mengelompokkan dampak gempa di Indonesia menggunakan algoritme Kohonen Self Organizing Maps (SOM). Saraswathi (2014) membuat komparasi beberapa algoritme *clustering* untuk mendapatkan akurasi dari algoritme K-Means, DB Scan, Hirarki dan Optic.

METODE PENELITIAN

1. Jenis dan Sumber Data

Jenis data primer yang digunakan dalam penelitian ini adalah data gempa bumi tahun 2017 yang terdiri dari data terjadinya gempa. Sumber data dalam penelitian ini berasal dari Badan Meteorologi, Klimatologi dan Geofisik (BMKG) <http://repogempa.bmkg.go.id/>.

2. Variabel Penelitian

Variabel dalam penelitian ini dibagi menjadi tiga bagian atribut yaitu kedalaman hiposentrum, magnitudo dan lokasi gempa. Yang menjadi data primer dalam proses klasifikasi adalah data kedalaman hiposentrum. Data tersebut akan dibagi menjadi tiga klasifikasi yaitu gempa dalam, gempa menengah dan gempa dangkal. Data primer merupakan data utama yang digunakan sebagai acuan dalam suatu penelitian dan merupakan data yang diolah dalam suatu penelitian. Data

3. Analisa Data Penelitian

Data primer yang digunakan dalam penelitian masih berupa data mentah yang belum diolah. Ada sekitar 6090 data yang digunakan dalam penelitian ini. Data mentah tersebut memiliki atribut yang datanya hilang (*missing values*) terutama pada atribut magnitudo. Jumlah atribut yang datanya hilang ada sekitar 3678 data. Data yang demikian perlu diolah atau *dipreprocessing*. Untuk melakukan *preprocessing* data menggunakan algoritme normalisasi data. Dataset dibagi menjadi dua bagian yaitu dataset sebagai data pelatihan (*training*) sebanyak 80% dari 6090 data yaitu 4872 data. Sedangkan data uji (*testing*) sebanyak 20% yaitu sebanyak 1218 data. Contoh data gempa terlihat pada Tabel 1.

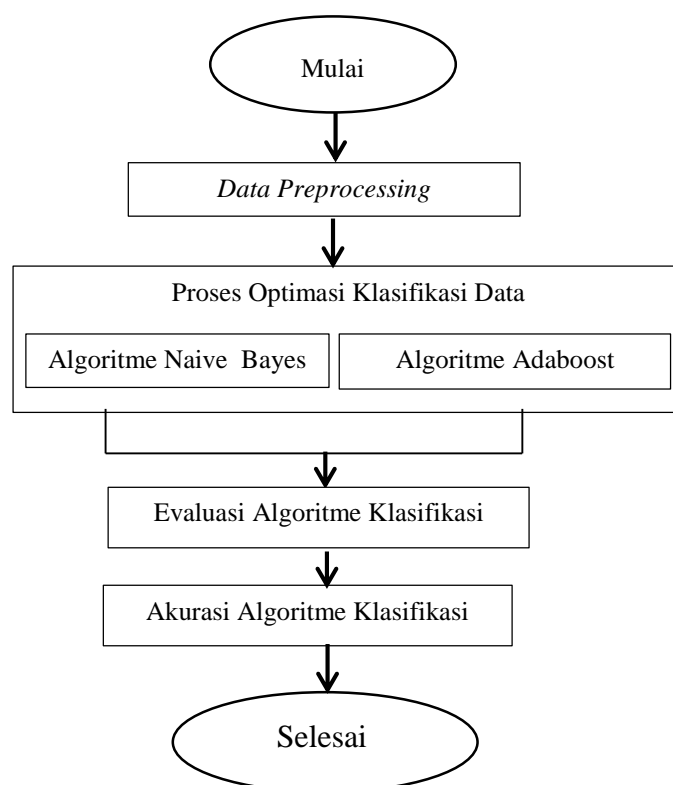
Tabel 1. Contoh data gempa di Indonesia

<i>Data ke-</i>	<i>Tanggal</i>	<i>Waktu (WIB = UTC + 7 Jam)</i>	<i>Hipotesis</i>	<i>Mag</i>	<i>Type Mag</i>	<i>smaj</i>	<i>smin</i>	<i>az</i>	<i>rms</i>	<i>Region</i>
1	1/1/2017	13:25.9	76	5	MLv	62.78	0.96	160	1.415	Northern Molucca Sea
2	1/1/2017	45:54.6	74	4	MLv	1.79	0.55	158	0.601	Near North Coast of Irian Jaya
3	1/1/2017	58:13.7	10	4	MLv	7.68	1.16	210	0.894	Java, Indonesia
4	1/1/2017	05:01.1	533	4.2	mb	24.47	1.13	51	0.595	Flores Sea
5	1/1/2017	39:18.2	10	4.1	MLv	4.95	0.87	215	0.816	Southwest of Sumatra, Indonesia
6	1/1/2017	12:09.8	66	5.1	mb	24.41	2.07	144	0.958	North of Halmahera, Indonesia
7	1/1/2017	13:39.1	32	5.5	Mw(mB)	84.39	2.03	126	1.342	North of Halmahera, Indonesia
8	1/1/2017	17:51.7	10	4.8	mb	17.66	2.33	262	1.238	North of Halmahera, Indonesia
9	1/1/2017	13:39.4	10	4.4	mb	12.43	2.1	242	0.545	North of Halmahera, Indonesia
10	1/1/2017	27:23.9	63	1.8	MLv	1.71	0.44	101	0.225	Minahassa Peninsula, Sulawesi
11	1/1/2017	36:49.8	10	3.5	MLv	4.01	1.01	213	0.416	South of Bali, Indonesia
12	1/1/2017	25:16.8	15	3	MLv	2	0.95	270	0.318	Southwest of Sumatra, Indonesia
13	1/1/2017	46:40.7	21	2.9	MLv	4.25	0.68	268	0.107	Sumbawa Region, Indonesia
14	1/1/2017	01:15.3	14	2.7	MLv	3.27	1.43	157	0.092	Sulawesi, Indonesia
15	1/1/2017	18:52.1	293	3.7	MLv	6.76	0.47	72	0.685	Bali Sea
16	1/1/2017	36:03.7	155	4.4	mb	9.38	1.41	210	0.851	North of Halmahera, Indonesia
17	1/1/2017	45:22.1	10	3.5	MLv	4.09	0.1	59	1.128	Northern Sumatra, Indonesia
18	1/1/2017	58:00.8	35	4.7	mb	53.59	1.89	144	0.982	Off West Coast of

<i>Data ke-</i>	<i>Tanggal</i>	<i>Waktu (WIB = UTC + 7 Jam)</i>	<i>Hipotesis</i>	<i>Mag</i>	<i>Type Mag</i>	<i>smaj</i>	<i>smin</i>	<i>az</i>	<i>rms</i>	<i>Region</i>
										Northern Sumatra
19	1/2/2017	56:50.3	10	2.9	MLv	1.35	0.76	247	0.466	Southern Sumatra, Indonesia
20	1/2/2017	52:57.9	87	4.9	mb	16.06	1.41	79	0.819	Banda Sea

4. Tahapan Penelitian

Berikut tahapan penelitian yang dilakukan oleh penulis, terlihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. *Data Preprocessing*

Pada proses ini untuk data yang hilang atau *missing value* perlu dilakukan pembersihan data atau *data cleaning*. Pada *data preprocessing* ini data yang mengalami pengulangan *record* atau redundansi dinormalisasi.

b. Proses Klasifikasi Data

Proses klasifikasi data menggunakan algoritme *Naive Bayes* dengan rumus

c. Optimasi

Data yang sudah dinormalisasi kemudian diklasifikasi dengan algoritme *Naive Bayes* dan dioptimasi dengan algoritme *Adaboost*.

Algoritme *Naive Bayes* terdiri dari beberapa tahap yaitu:

1. Menghitung jumlah kelas yang akan diklasifikasikan
2. Menghitung jumlah kasus dari setiap kelas
3. Kalikan dengan varian variabel yang telah ditentukan

4. Bandingkan hasilnya pada setiap kelas

Konsep dasar dari algoritme *Adaboost* adalah memberikan bobot yang lebih pada klasifikasi yang tidak tepat (*weak classification*). *Boosting* bisa dikombinasi dengan *classifier* algoritme lainnya untuk menambah performa klasifikasi (Listiana, 2017). Tahapan dalam algoritme *Adaboost* (Listiana, 2017) yaitu:

1. Input: Suatu kumpulan *dataset* penelitian dengan label $\{(x_i, y_i) (x_N, y_N)\}$, suatu *component learn algoritme*, jumlah perputaran T.
2. Inisialisasi: Bobot $\{w_i\}$ data *training* $w_i^1 = 1/N$ untuk semua $i = 1, \dots, N$
3. Iterasi for $m = 1, \dots, M$

- a. Gunakan *component learn algoritme* untuk melatih suatu komponen klasifikasi, y_m , pada bobot data pelatihan. Cocokkan *clasifier* $y_m(x_i)$ dengan data pelatihannya untuk meminimalisir bobot fungsi *error* pada:

$$J_m = \sum_{i=1}^N w_i^m I, y_m(x_i) \neq t_n \quad (1)$$

dimana $I(y_m(x_i) \neq t_n)$ sebagai fungsi indikator akan bernilai 1 ketika $y_m(x_i) \neq t_n$ dan 0 untuk $y_m(x_i) = t_n$

- b. Evaluasi kuantitas

$$\epsilon_m = \frac{\sum_{i=1}^N w_i^m I, y_m(x_i) \neq t_n}{\sum_{i=1}^N w_i^m} \quad (2)$$

Gunakan persamaan 2. untuk mengevaluasi persamaan 3.

$$\alpha_m = \ln \left\{ \frac{1 - \epsilon_m}{\epsilon_m} \right\} \quad (3)$$

- c. *Update* koefisien bobot

$$w_i^{(m+1)} = \frac{w_i^{(m)} \exp\{\alpha_m I(y_m(x_i) \neq t_n)\}}{C_m} \dots \dots \dots (4)$$

dengan C_m adalah konstanta untuk normalisasi

- d. *Output*:

$$Y_M(x) = \text{sign} \left(\sum_{m=1}^M \alpha_m y_m(x) \right) \quad (5)$$

- d. Evaluasi Algoritme Klasifikasi

Metode klasifikasi *Naïve Bayes* yang sudah dioptimasi pada data hiposentrum gempa dievaluasi menggunakan algoritme *confusion matrix*.

Pada penelitian ini klasifikasi dibagi menjadi tiga kelas maka ada cara menghitung akurasi, presisi dan recall dapat dilakukan dengan menghitung rata-rata dari nilai akurasi, presisi dan recall pada setiap kelas.

$$\text{Akurasi} = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \times 100\% \quad (6)$$

dimana:

- TP_i adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i.
- TN_i adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i.

- FN_i adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem untuk kelas ke-i.
 - FP_i adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem untuk kelas ke-i
- e. Akurasi Algoritme Klasifikasi

Akurasi dari algoritme optimasi *Naïve Bayes* menggunakan *precision* dan *recall*.

$$Precision = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (FP_i + TP_i)} * 100\% \quad (7)$$

$$Recall = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (FP_i + FN_i)} * 100\% \quad (8)$$

HASIL DAN PEMBAHASAN

1. Hasil dari *Preprocessing Data* menggunakan proses normalisasi

(*Z-Transformation*) dan *replace missing values*. Normalisasi yang dilakukan adalah normalisasi pada atribut magnitudo dengan hasil normalisasi mean: 4.047947569271616, variance: 113.84787442937836. Untuk hasil *replace missing values* dengan menambahkan data dengan perhitungan rata-rata (*average*) pada atribut magnitudo.

2. Hasil klasifikasi gempa bumi dan evaluasi algoritme klasifikasi

Hasil klasifikasi gempa bumi terlihat pada tabel 1 berdasarkan data testing yang diujikan yaitu sebanyak total 1218 data. Gempa dangkal hiposentrumnya kurang dari 60 km, gempa sedang hiposentrumnya antara 60 km sampai 300 km dan gempa dalam memiliki hiposentrum lebih dari 300 km. Akurasi yang dihitung adalah akurasi dari hasil klasifikasi yaitu akurasi gempa dangkal, sedang dan dalam. Algoritme yang digunakan untuk menghitung akurasi menggunakan *algoritme confusion matrix*. Perhitungan akurasi dari setiap metode *Naïve Bayes* dan *Naïve Bayes + Adaboost* dapat dilihat pada Tabel 2 sampai dengan Tabel 7.

Tabel 2. Akurasi Gempa Dangkal dengan *Naïve Bayes*
Banyak Gempa =214

	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN=38	FP=37
Actual True	FN=29	TP = 110
Akurasi =	69%	

Tabel 3. Akurasi Gempa Sedang dengan *Naïve Bayes*
Banyak Gempa =998

	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN=96	FP=99
Actual True	FN=89	TP = 714
Akurasi =	81%	

Tabel 4. Akurasi Gempa Dalam dengan *Naïve Bayes*
Banyak Gempa 6

	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN=1	FP=2
Actual True	FN=0	TP = 3
Akurasi =	67%	

Tabel 5. Akurasi Gempa Dangkal dengan Optimasi *Naïve Bayes* dan *Adaboost*
Banyak Gempa = 119

	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN=23	FP=13
Actual True	FN=7	TP = 76
Akurasi =	83%	

Tabel 6. Akurasi Gempa Sedang dengan Optimasi *Naïve Bayes* dan *Adaboost*
Banyak Gempa = 1093

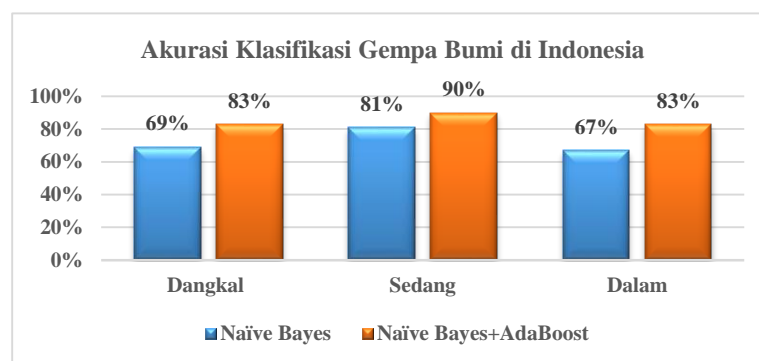
	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN =110	FP = 11
Actual True	FN = 96	TP = 876
Akurasi =	90%	

Tabel 7. Akurasi Gempa Dalam dengan Optimasi *Naïve Bayes* dan *Adaboost*
Banyak Gempa = 6

	<i>Predicted False</i>	<i>Predicted True</i>
Actual False	TN=2	FP=1
Actual True	FN=0	TP = 3
Akurasi =	83%	

Tabel 8. Jumlah Lokasi Berdasarkan Klasifikasi Hiposentrum Gempa Bumi

<i>Algoritme Klasifikasi</i>	<i>Gempa Dangkal</i>	<i>Akurasi Gempa Dangkal</i>	<i>Gempa Sedang</i>	<i>Akurasi Gempa Sedang</i>	<i>Gempa Dalam</i>	<i>Akurasi Gempa Dalam</i>	<i>Rata- Rata</i>
<i>Naïve Bayes</i>	214	69%	998	81%	6	67%	72,3%
<i>Naïve Bayes + Adaboost</i>	119	83%	1093	90%	6	83%	85,3%



Gambar 2. Hasil Akurasi Setiap Klasifikasi Gempa

Pada Gambar 2. Disimpulkan bahwa hasil rata - rata akurasi algoritme *Naïve Bayes* sebesar 72,3% dan optimasi algoritme *Naïve Bayes* dan *Adaboost* sebesar 85,3%

KESIMPULAN DAN SARAN

Kesimpulan dan saran dari penelitian ini adalah Hasil rata - rata akurasi algoritme *Naïve Bayes* sebesar 72,3% dan algoritme *Naïve Bayes* dan *Adaboost* sebesar 85,3%. Dari hasil klasifikasi terlihat bahwa jumlah lokasi yang paling banyak terjadi gempa bumi adalah lokasi gempa bumi sedang sebanyak 1093 lokasi. Penelitian ini juga bisa dikembangkan untuk klasifikasi data gempa yang berpotensi tsunami dan dengan menggunakan algoritme *clustering* untuk mengelompokkan gempa bumi berdasarkan bagian wilayah Indonesia.

DAFTAR PUSTAKA

- Chengsheng, T., Huacheng, L., & Bing, X. (2017). *Adaboost* typical Algorithm and its application research. In *MATEC Web of Conferences* (Vol. 139, p. 00222). EDP Sciences.
- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using *Naïve Bayes* and k-nn classifier. *arXiv preprint arXiv:1610.09982*.
- E. Prasetyo, (2012) *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1 ed. Yogyakarta: Andi Offset.
- Fatihah, C., & Purwitasari, D. (2017). Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM. *Jurnal Teknik ITS*, 6(1), 153-158.
- Haixiang, G., Yijing, L., Yanan, L., Xiao, L., & Jinling, L. (2016). BPSO-*Adaboost*-KNN ensemble learning algorithm for multi-class imbalanced data classification. *Engineering Applications of Artificial Intelligence*, 49, 176-193.
- Halim, N. N., & Widodo, E. (2017). Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM). In *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai-Nilai Islami)* (Vol. 1, No. 1, pp. 188-194).
- Hartuti, E. R. 2009. *Buku Pintar Gempa*. Yogyakarta: Diva Perss
- Listiana, E., & Muslim, M. A. (2017). Penerapan *Adaboost* untuk Klasifikasi Support Vector Machine Guna Meningkatkan Akurasi pada Diagnosa Chronic Kidney Disease. *Prosiding SNATIF*, 875-881.
- Nakra, A., & Duhan, M. (2019). Comparative Analysis of Bayes Net Classifier, *Naïve Bayes* Classifier and Combination of both Classifiers using WEKA. *IJ Inf. Technol. Comput. Sci*, 11, 38-45.
- Saraswathi, S., & Sheela, M. I. (2014). A comparative study of various clustering algorithms in data mining. *International Journal of Computer Science and Mobile Computing*, 11(11), 422-428.
- Sari, N. N. (2018). K-Affinity Propagation (K-AP) Clustering Untuk Klasifikasi Gempa Bumi (Studi Kasus: Gempa Bumi di Indonesia Tahun 2017).
- Saritas, M. M., & Yasar, A. (2019). Performance Analysis of ANN and *Naïve Bayes* Classification Algorithm for Data Classification. *International Journal of Intelligent Systems and Applications in Engineering*, 7(2), 88-91.