

Application of the Decision Tree Method to Forest Fire Detection (Case Study: in Palembang, South Sumatra)

Penerapan Metode *Decision Tree* terhadap Deteksi Kebakaran Hutan (Studi Kasus: Palembang, Sumatera Selatan)

Awanda Putra Mahendra¹, Dody Pradipta², Moh. Rizal Bayu Saputro³, Kusri⁴

^{1,2,3}Program Magister Komputer, Universitas AMIKOM Yogyakarta

⁴Universitas AMIKOM Yogyakarta, Indonesia

E-mail: ¹awanda.m@students.amikom.ac.id, ²pradiptadody@students.amikom.ac.id,

³moh.saputro@students.amikom.ac.id, ⁴kusri@amikom.ac.id

Abstract – Forest and land fires are very detrimental to the community since these occasions have a direct affect on the environment, particularly for the individuals around the area of the incident. This consider points at forest and land fires that happen based on natural conditions around the city of Palembang, South Sumatra This study uses data from the BPS Palembang city as much as 60 data in four years (2015-2019). The model used in this study is the Decision Tree C4.5 algorithm with the variables used in a month, wind speed, rainfall, and humidity. This study applies five scenarios with a combination of variables with the best level of accuracy. The research results are within the form of a classification of whether a forest or land fire occurs or not. The accuracy value of the dataset is within four years (2015-2019), with the scenario of sharing 80 training datasets and 25% of testing data getting an accuracy value of 92%. At that point the Decision Tree C4.5 algorithm model can be utilized in this study, since it has the potential to classify the potential for forest and land fires in the city of Palembang, South Sumatra.

Keywords — decision tree c4.5, data mining, forest fires

Abstrak – Kebakaran hutan dan lahan sangat merugikan bagi masyarakat dikarenakan peristiwa ini berdampak langsung pada lingkungan terutama bagi masyarakat disekitar lokasi kejadian. Penelitian ini digunakan untuk klasifikasi kebakaran hutan dan lahan yang terjadi berdasarkan kondisi lingkungan sekitar pada kota Palembang, Sumatera Selatan. penelitian ini menggunakan data dari BPS kota Palembang sebanyak 60 data dalam empat tahun (2015-2019). Model yang digunakan pada penelitian ini adalah algoritma *Decision Tree C4.5* dengan menggunakan variabel berupa bulan, kecepatan angin, curah hujan dan kelembaban udara. Penelitian ini menerapkan 5 skenario dengan kombinasi variabel dengan tingkat akurasi terbaik. Hasil penelitian yang dikeluarkan berupa klasifikasi apakah suatu hutan dan lahan terjadi kebakaran atau tidak. Nilai akurasi dari *dataset* tersebut dalam waktu empat tahun (2015-2019) dengan skenario pembagian *dataset training* 80 dan *data testing* 25% mendapatkan nilai akurasi sebesar 92%. Maka model algoritma *Decision Tree C4.5* dapat digunakan pada penelitian ini, karena memiliki potensi untuk melakukan klasifikasi terhadap potensi kebakaran hutan dan lahan pada kota Palembang, Sumatera Selatan.

Kata Kunci — pohon keputusan c4.5, data mining, kebakaran hutan

1. PENDAHULUAN

Kebakaran hutan di Indonesia menimbulkan banyak kerugian baik dalam hal ekonomi maupun ekologi, bukan hanya manusia tetapi ekosistem juga mendapatkan dampak yang luar biasa. Sumatera merupakan salah satu daerah yang memiliki luas kebakaran dengan kerugian yang cukup tinggi.

Sumatera selatan menjadi salah satu dari penyumbang kebakaran tertinggi di Indonesia. Data karhutla KLHK menyebutkan 35.231 Hektare terbakar pada bulan januari sampai bulan Mei 2020, selain itu, juga terjadi penurunan 9,13 persen teridentifikasi sebanyak 54 persen kebakaran terjadi pada lahan gambut, dikarenakan masih banyaknya lahan gambut yang rawan akan terjadinya kebakaran [1].

Kebakaran hutan dan lahan dapat terjadi pada kawasan hutan bahkan dapat juga terjadi pada di luar hutan, dan juga dapat terjadi pada tanah mineral maupun lahan gambut. Kebakaran pada lahan gambut tidak mudah untuk dipadamkan karena biomassa yang terdapat pada tanah dapat menyebarkan api dengan cepat dan juga dapat menyebar pada lapisan gambut yang berada di bawah permukaan tanah. Kondisi pada lahan gambut yang kering akibat perusakan lahan dapat menyebabkan mudahnya materi tersebut untuk terbakar, terutama pada musim kemarau. Sehingga diperlukan penanganan khusus terhadap pengendalian potensi kebakaran hutan yang lebih massif [2].

Data Mining merupakan langkah untuk ekstraksi data tidak berguna dan berkarakter konotatif menjadi informasi dari data yang memiliki jumlah yang banyak [1]. *Data mining* memiliki beberapa metode salah satunya adalah metode klasifikasi. Klasifikasi dapat diterapkan pada kasus kebakaran hutan yaitu dengan menggunakan *Decision Tree* untuk deteksi dini terhadap ancaman kebakaran hutan. Primajaya A. dkk (2020) meneliti tentang prediksi kebakaran hutan menggunakan algoritma klasifikasi C4.5 menyimpulkan bahwa algoritma C4.5 menggunakan skenario sebesar 80% untuk *data training* dan 20% untuk *data testing* dapat memberikan nilai akurasi sebesar 89,8%. Pacheco dkk (2021) yang meneliti *Assessment of k-Nearest Neighbor and Random Forest Classifiers for Mapping* menghasilkan bahwa Akurasi terbaik didapatkan pada KNN nilai K=5 dan pada RFC 400 pohon dipilih sebagai nilai optimal [1].

Beberapa penelitian mengatakan ada banyak faktor yang dapat mempengaruhi kebakaran hutan dan lahan seperti iklim, kondisi lahan, jenis tanah dan faktor lingkungan bio-fisik lainnya Yusuf (2019) [7]. Penelitian Primajaya (2020) algoritma C4.5 interpretasi serta evaluasi pengetahuan. *Data training* dan *testing* menggunakan *Use training Set*, *Percentage split*, *Cross validation* untuk skenario persentase yang selanjutnya akan dipilih model terbaik. Penelitian ini menggunakan indikator evaluasi yaitu nilai akurasi. Penelitian ini memberikan kesimpulan bahwa algoritma C4.5 yang memiliki *percentage split* 80% untuk *data training* serta 20% untuk *data testing* dapat memberikan nilai akurasi tinggi yaitu sebesar 89,7859% [2].

Berdasarkan penelitian Pratiwi T. A (2021) yang menggunakan algoritma *Naive Bayes* yang menggunakan *Split validation* untuk membagi *data testing* dan juga *data training* menjabarkan 25% digunakan sebagai *data testing* serta 75% digunakan sebagai *data training*. Lokasi yang terjadi kebakaran hutan yang dideteksi menggunakan titik panas sebagai indikator karena suhu sekitar lokasi kebakaran hutan lebih rendah sehingga suhu pada lokasi titik kebakaran hutan relatif tinggi. Perhitungan akurasi menggunakan *confusion matriks* memberikan hasil bahwa pada tahun 2015 memiliki hasil terbaik dengan nilai akurasi sebesar 99.45% dan Kappa sebesar 98.82% untuk wilayah pada kabupaten Pelandayan [1].

Penelitian Primajaya A. (2020) meneliti tentang “Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat”. Menggunakan metodologi *knowledge discovery in database* (KDD). Metodologi KDD berisikan beberapa tahapan seperti pengolahan data, transformasi data, pemrosesan data pengumpulan dan pemilihan data dengan algoritma C4.5 serta interpretasi dan evaluasi pengetahuan. *Use training Set*, *Percentage split*, *Cross validation* digunakan untuk membagi skenario data training dan testing dengan persentase dan model terbaik yang akan dipilih. Indikator evaluasi model yang dibangun menggunakan akurasi. Penelitian ini memberikan kesimpulan bahwa C4.5 dengan *percentage split* 80% sebagai *data training* dan 20% sebagai *data testing* memberikan nilai akurasi tertinggi sebesar 89,7859% [2].

Penelitian Noviansyah M. R (2018) yang meneliti tentang “Penerapan Data Mining Menggunakan Metode *K-Nearest Neighbor* untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data AWS (*Automatic Weather Station*)”. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* (K-NN). Penelitian ini menerapkan algoritma klasifikasi *K-Nearest Neighbor* (K-NN) dalam klasifikasi tingkat bahaya kebakaran hutan berdasarkan *Fire Watcher Index* (FWI). atribut variabel yang digunakan dalam penelitian ini berupa suhu, kecepatan angin, curah hujan dan kelembaban udara. Pengujian yang dilakukan dengan menghitung jarak terhadap data latih dengan data uji menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk mendapatkan nilai K terbaik. Berdasarkan hasil pengujian menghasilkan persentase data uji sebesar 80,16% dengan nilai K=5 menggunakan 262 data uji [3].

Penelitian Fitriyani (2018) tentang “Komparasi Algoritma LR, K-NN Dan SVM Untuk Estimasi Area Kebakaran Hutan”. Metode yang menggunakan algoritma *Linear Regression* (LR), *K-*

Nearest Neighbor (K-NN) dan *Support Vector Machine* (SVM). Metode tersebut digunakan untuk proses regresi dan klasifikasi. Penelitian ini dilakukan dengan cara membandingkan tiap algoritma untuk menghasilkan algoritma terbaik yang dapat digunakan sebagai estimasi area kebakaran hutan. *Dataset* yang digunakan *dataset* kebakaran hutan dari *UCI Repository* yang merupakan *dataset* publik. Hasil pengujian terhadap *dataset* yang digunakan untuk setiap algoritma, menunjukkan bahwa nilai RMSE terbaik dihasilkan pada pengujian *Linear Regression* (LR) yang memberikan hasil nilai RMSE sebesar 1.496, Selanjutnya pada algoritma *K-Nearest Neighbor* (K-NN) memiliki nilai RMSE sebesar 1.981, dan terakhir pada algoritma *Support Vector Machine* (SVM) memiliki nilai RMSE sebesar 3.088 [4].

Penelitian Karo I. M. K (2020) mengenai “Implementasi Metode XGBoost dan *Feature Importance* untuk Klasifikasi pada Kebakaran Hutan dan Lahan”. Metode yang digunakan pada penelitian ini menggunakan XGBoost dan *feature importance* yang terdapat pada pulau Sumatera. Data yang digunakan berupa data titik panas yang diperoleh dari *Global Forest Watch*. terdapat enam dan tujuh variabel yang sangat berpengaruh sebagai penentu titik panas. XGBoost dan *feature importance* menghasilkan nilai akurasi sebesar 89,52%. *Sensitivity* (SE) sebesar 91,32%, *Specificity* (SP) sebesar 93,16%, dan *Matthews Correlation Coefficient* (MCC) sebesar 92,75% [5].

Penelitian Sugianto C. A (2015) tentang “Analisis Komparasi Algoritma Klasifikasi untuk Menangani data Tidak Seimbang pada Data kebakaran Hutan”. Metode yang digunakan untuk klasifikasi menggunakan algoritma *K-Means + Neural Network*, *K-Means + Random Forest*, *K-Means + Naive Bayes*, dan *K-Means + C4.5*. *Dataset* yang digunakan terdiri dari dua kombinasi variabel meteorologi dan *fire weather index* (FWI) untuk memprediksi ukuran kebakaran hutan. Hasil akurasi dari penelitian ini didapatkan algoritma *K-Means + Naive Bayes* sebesar 80,49%, *K-Means + C4.5* sebesar 94,01%, *K-Means + Random Forest* sebesar 79,90%, dan *K-Means + Neural Network* sebesar 90,12%. Hasil pengujian menghasilkan algoritma *K-Means + C4.5* dapat menangani *dataset* yang tidak seimbang [6].

Berdasarkan fenomena diatas, peneliti tertarik dan berinisiatif mengangkat judul “Penerapan Metode Algoritma *Decision Tree* terhadap Deteksi Kebakaran Hutan (Studi Kasus: Palembang, Sumatera Selatan)” dalam penelitiannya.

2. LANDASAN TEORI

2.1. Kebakaran Hutan

Kebakaran hutan dan lahan merupakan bencana yang disebabkan oleh perbuatan manusia ataupun alami. Tanda terjadinya kebakaran hutan ditandai dengan penyebaran api secara cepat yang melahap hutan dan lahan yang dilaluinya [8].

2.2. Kecepatan Angin

Angin adalah pergerakan udara yang sejajar dengan permukaan bumi. Udara bergerak ke daerah bertekanan rendah dari daerah bertekanan tinggi. Sirkulasi umum dari atmosfer merupakan gerak rata-rata angin di permukaan bumi. Tekanan angin merupakan penyebab terjadinya angin. Sedangkan kecuraman beda tekanan merupakan penyebab kecepatan angin. Kecepatan angin sendiri memiliki satuan *knot* [9].

2.3. Curah Hujan

Curah hujan adalah volume air hujan yg memiliki satuan per luas yang merupakan kumpulan air hujan yang jatuh sampai ke tanah. Curah hujan yang tinggi banyak mendukung kegiatan ekonomi dan sosial di Indonesia yang dapat dijadikan sebagai informasi berguna bagi kehidupan sehari-hari, seperti pemadaman kebakaran hutan, produksi pertanian, keselamatan masyarakat, perikanan dan perkebunan [10].

2.4. Kelembaban Udara

Kelembaban udara merupakan komponen dari cuaca dan iklim. Kandungan uap air di udara digambarkan sebagai kelembaban udara yang dinyatakan sebagai kelembaban mutlak, kelembaban

relatif dan *deficit* tekanan uap air. Suhu dan curah hujan juga dapat mempengaruhi tingkat kelembaban pada udara. Kelembaban udara juga salah satu faktor lingkungan yang sangat mempengaruhi terjadinya penyebaran penyakit [11].

2.5. Data Mining

Data mining merupakan istilah yang digunakan untuk mendalami pengetahuan dalam bidang *big data*. *Data Mining* merupakan langkah untuk ekstraksi data tidak berguna dan berkarakter konotatif menjadi informasi dari data yang memiliki jumlah yang banyak [12]. *Data mining* digunakan untuk identifikasi informasi menggunakan matematika, *machine learning*, kecerdasan buatan dan statistic [13]. *Data mining* menggunakan tiga tipe data antara lain tipe data kategorial, tipe data rentang waktu, dan tipe data numerik [14].

3. METODE PENELITIAN

3.1. Perumusan Masalah

Kunci dalam metode penelitian adalah perumusan masalah. Pada penelitian yang rumusan masalah berupa bagaimana melakukan klasifikasi terhadap kebakaran hutan dan lahan pada kota Palembang menggunakan model Algoritma C4.5.

3.2. Pengumpulan Data

Tahap pengumpulan data akan dibagi menjadi dua tahap yaitu observasi dan studi literatur.

a. Observasi

- (1) Data Primer yang digunakan merupakan data BPS Kota Palembang pada tahun 2015 hingga tahun 2019 dengan format .xlsx. Data tersebut digunakan untuk melakukan tahap *Training* dan *Testing*. Variabel yang digunakan berupa bulan, kecepatan angin, curah hujan dan kelembaban udara. Jumlah data yang digunakan sebanyak 61 data dengan jangka waktu 4 tahun (2015-2019).
- (2) Data Sekunder digunakan untuk melabeli data pada dataset yang digunakan. Data label diambil dari *Global Forest Watch* wilayah Indonesia dengan fokus di Kota Palembang. Variabel yang digunakan berupa hasil yang digunakan untuk melabeli data primer.

b. Studi Literatur

Studi literatur merupakan tahap pertama pada pengumpulan data. Metode ini dilakukan dengan mencari sumber terkait penelitian yang dilakukan.

c. Preprocessing

Pada tahap *preprocessing* dilakukan untuk pembersihan data yang akan digunakan terhadap data kosong, inkonsisten, dan ketidaksesuaian data. Tahap ini juga dilakukan *preprocessing* dengan memeriksa *missing value*. Jika ditemukan *missing value* maka dilakukan penghapusan data.

d. Transformation

Transformation dilakukan untuk mengubah data yang memiliki tipe data numerik menjadi tipe kategorikal. Tahap ini pemecahan data dilakukan dengan tujuan untuk melakukan testing data dan training data yang dilakukan dengan beberapa skenario. Pemisahan data juga dilakukan dengan teknik *percentage split*.

e. Data Mining

Tahap awal dari proses perhitungan algoritma C4.5 adalah dengan melakukan pengambilan data training dan data kebakaran hutan dan lahan. Variabel yang digunakan yaitu :

- (1) Bulan merupakan variabel yang digunakan untuk menentukan waktu kebakaran hutan dan lahan pada masa lampau.

- (2) Kecepatan angin merupakan variabel dengan satuan *knot*.
- (3) Curah hujan merupakan variabel dengan satuan mm.
- (4) Kelembaban udara merupakan variabel dengan satuan persen.

3.3. Dataset

Dataset yang digunakan untuk klasifikasi kebakaran hutan adalah: kecepatan angin, curah hujan, kelembaban udara dan klasifikasi kebakaran dengan jumlah data yang digunakan berjumlah 60 data.



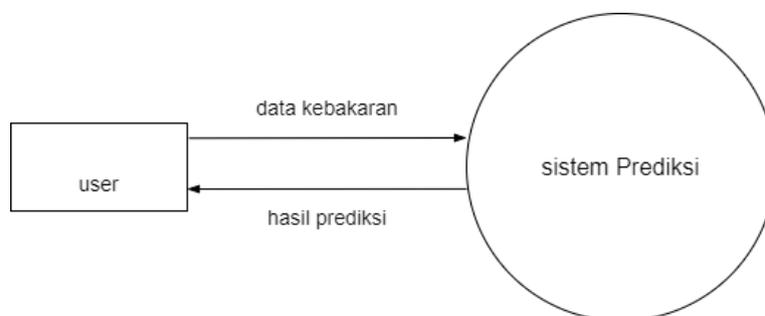
	bulan	kecepatan_angin	curah_hujan	kelembaban_rerata	hasil
0	01/01/2015	390	22160	8400	0
1	01/02/2015	340	13220	8400	0
2	01/03/2015	270	39050	8500	0
3	01/04/2015	220	37560	8300	0
4	01/05/2015	270	17790	8100	0
5	01/06/2015	310	17020	8100	0
6	01/07/2015	460	2140	7600	1
7	01/08/2015	490	2120	7500	1
8	01/09/2015	480	530	7100	1

Gambar 1. Dataset

Pada penelitian ini menggunakan beberapa model skenario yaitu *data training* sebesar 80% dan *data testing* sebesar 20%, skenario ke-2 yaitu *data training* sebesar 75% dan *data testing* sebesar 25%, dengan membuat kombinasi antara variabel yang digunakan. Untuk menguji model yang sudah dibangun pada penelitian ini akan digunakan perhitungan *Confussion Matrix*.

3.4. Rancangan Data Flow Diagram (DFD)

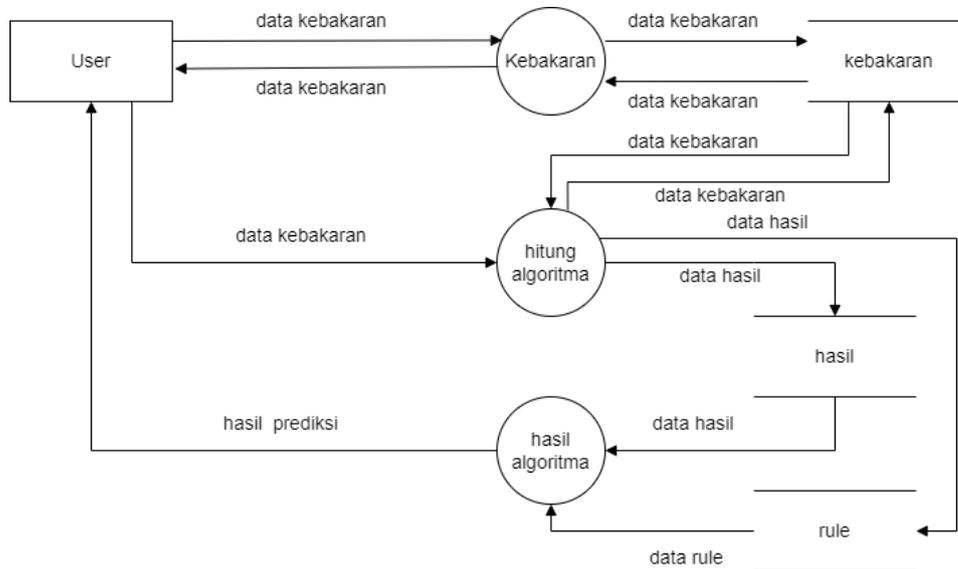
3.4.1. Data Flow Diagram Lv. 0



Gambar 2. DFD Lv. 0

User menginput data kebakaran yang terjadi kemudian sistem akan memberikan hasil prediksi sesuai dengan algoritma C4.5.

3.4.2. Data Flow Diagram Lv. 1



Gambar 3. DFD Lv. 1

User akan meng-input data kebakaran yang sudah pernah terjadi dan akan disimpan kedalam table kebakaran, yang selanjutnya akan digunakan untuk membentuk model bagi algoritma C4.5. setelah mendapatkan model sistem akan menampilkan hasil dari prediksi kebakaran untuk setiap 3 jam.

4. HASIL DAN PEMBAHASAN

Tabel 1. Data hasil Pengujian

Skenario	Data Training	Data Testing	Kombinasi Variabel
Skenario 1	80%	20%	kecepatan angin, curah hujan, kelembaban rerata dan klasifikasi kebakaran
Skenario 2	70%	30%	kecepatan angin, curah hujan, kelembaban rerata dan klasifikasi kebakaran
Skenario 3	80%	20%	curah hujan, kelembaban rerata dan klasifikasi kebakaran
Skenario 4	80%	20%	kecepatan angin, kelembaban rerata dan klasifikasi kebakaran
Skenario 5	80%	20%	kecepatan angin, curah hujan dan klasifikasi kebakaran

Pada penelitian ini mendapatkan akurasi tertinggi pada skenario 1 mendapatkan tingkat akurasi sebesar 92%. Pada skenario 2 - skenario 5 menghasilkan akurasi sebesar 72%, 67%, 75% dan 67%.

5. KESIMPULAN

Kesimpulan dari penelitian ini dari fakta hasil pengujian adalah klasifikasi kebakaran hutan menggunakan data kebakaran BPS dapat menggunakan algoritma *Decision Tree*. Algoritma ini merupakan salah satu dari metode klasifikasi, digunakan untuk membantu dalam proses pengambilan keputusan menggunakan struktur pohon atau struktur hirarki. *Dataset* yang digunakan adalah perpaduan antara dataset BPS dan *dataset Global Forest Watch* pada wilayah Sumatera Selatan pada tahun 2015 - 2019 menghasilkan model terbaik dengan skenario 1 dengan tingkat akurasi yang didapatkan sebesar 92%.

DAFTAR PUSTAKA

- [1] Pratiwi, T. A., Irsyad, M., Kurniawan, R., Agustian, S., & Negara, B. S. (2021). Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan. *CESS (Journal of Computer Engineering, System and Science)*, 6(1), 139-148.
- [2] Primajaya, A., Sari, B. N., & Khusaeri, A. Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4. 5 Studi Kasus Provinsi Kalimantan Barat. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 6(2), 188-192.
- [3] Noviansyah, M. R., Rismawan, T., & Midyanti, D. M. (2018). Penerapan data mining menggunakan metode k-nearest neighbor untuk klasifikasi indeks cuaca kebakaran berdasarkan data AWS (automatic weather station) (studi kasus: kabupaten Kubu Raya). *Coding Jurnal Komputer dan Aplikasi*, 6(2).
- [4] Fitriyani, F., & Sanjaya, R. (2018). KOMPARASI ALGORITMA LR, K-NN DAN SVM UNTUK ESTIMASI AREA KEBAKARAN HUTAN. *Infotronik: Jurnal Teknologi Informasi dan Elektronika*, 3(2), 103-110.
- [5] Karo, I. M. K. Implementasi Metode XGBoost dan Feature Important untuk Klasifikasi pada Kebakaran Hutan dan Lahan. *Journal of Software Engineering, Information and Communication Technology (SEICT)*, 1(1), 10-16.
- [6] Sugianto, C. A. (2015). Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan. *Techno. com*, 14(4), 336-342.
- [7] Yusuf A., Hapsih, Siregar S.H., Nurrochmat D.R., (2019). Analisis Kebakaran Hutan dan Lahan di Propinsi Riau. (*Dinamika Lingkungan Indonesia*), 6(2).
- [8] Maulida, L. (2018). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167-174.
- [9] Tayyib, M., & Winahju, W. S. (2014). Pemodelan Kecepatan Angin di Perairan Pulau Bawean dengan Menggunakan Fungsi Transfer. *Jurnal Sains dan Seni ITS*, 3(2), D248-D253.
- [10] Fauziah, N., Wahyuningsih, S., & Nasution, Y. N. (2016). Peramalan Menggunakan Fuzzy Time Series Chen (Studi Kasus: Curah Hujan Kota Samarinda). *Jurnal Statistika Universitas Muhammadiyah Semarang*, 4(2).
- [11] Alizkan, U. (2017). Analisis Korelasi Kelembaban Udara Terhadap Epidemi Demam Berdarah Yang Terjadi Di Kabupaten Dan Kota Serang. *Gravity: Jurnal Ilmiah Penelitian dan Pembelajaran Fisika*, 3(1).
- [12] Widiastuti, D. (2012). Analisa Perbandingan Algoritma Svm, Naive Bayes, Dan Decision Tree Dalam Mengklasifikasikan Serangan (Attacks) Pada Sistem Pendeteksi Intrusi.
- [13] Mohamed, H. (2015). *Implementasi algoritma naïve bayes untuk proses seleksi penerima beasiswa Libyan Embassy berbasis web* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

- [14] Setiawan, A. (2018). Kebijakan Penanganan Kebakaran Hutan dan Lahan di Indonesia. *Political Ecology*, VI (May), 1-13.

