



Implementasi Data Mining Menggunakan Metode *K-Nearest Neighbor* Untuk Memonitoring Data Pelanggan Oleh PT. Telkom Akses Pematangsiantar

¹Putri Natasya Lubis ²Nur Adha Pasaribu ³Nela Meilani ⁴Muhammad Siddik Hasibuan
¹Universitas Islam Negeri Sumatera Utara

Alamat Surat

Email: pnatasyalubis@icloud.com, nuradhapasaribu@gmail.com,
nelameilani2305@gmail.com, muhammadsiddik@uinsu.ac.id

Article History:

Diajukan: 13 April 2023; Direvisi: 25 April 2023; Accepted: 28 April 2023

ABSTRAK

Penelitian ini dilatarbelakangi oleh meningkatnya penggunaan WiFi di Indonesia yang menyebabkan peningkatan jumlah pelanggan WiFi PT. Telkom Akses. Efek ini sering menyebabkan keterlambatan proses layanan. Penelitian ini bertujuan untuk mengklasifikasikan keadaan pelayanan sebagai normal atau tidak normal berdasarkan data pelanggan Telkom. Pada penelitian ini, kami mengimplementasikan data mining dengan menggunakan metode *k-nearest neighbor*. Data yang digunakan adalah 100 dataset dari 10.121 data pelanggan telekomunikasi dari bulan September sampai Oktober 2022. Data dimasukkan ke dalam weka, dengan pemilihan atribut yang terdiri dari 60 data training dan 40 data testing, weka menangani perhitungan dan menyajikan hasil klasifikasi dalam bentuk normal atau tidak normal. Hasil perhitungan klasifikasi pada *dataset* 100 data menggunakan algoritma *k-nearest-neighbor* menghasilkan hasil terbaik menggunakan weka dengan nilai kebenaran 81% dan tingkat eror 19%.

Kata kunci: Data Mining, KNN, Weka, Pematangsiantar.

ABSTRACT

This research is motivated by the increasing use of WiFi in Indonesia which has led to an increase in the number of WiFi subscribers of PT. Telkom Access. This effect often causes delays in the service process. This study aims to classify service conditions as normal or abnormal based on Telkom customer data. In this study, we implement data mining using the k-nearest neighbor method. The data used is 100 datasets from 10,121 telecommunications customer data from September to October 2022. The data is entered into Weka, with attribute selection consisting of 60 training data and 40 testing data, Weka handles calculations and presents classification results in normal form or not normal. The results of classification calculations on a dataset of 100 data using the k-nearest-neighbor algorithm produce the best results using Weka with a truth value of 81% and an error rate of 19%.

Keywords: Data Mining, KNN, Weka, Pematangsiantar.

1. PENDAHULUAN

Ketersediaan sarana dan prasarana milik PT. Telekomunikasi tersebar di seluruh Indonesia. Contoh akses internet yang dimiliki oleh PT. Telkom yaitu speedy. Speedy adalah layanan koneksi internet untuk kualitas hidup Telkom juga menawarkan layanan jaringan WiFi. Layanan yang ditawarkan antara lain memasang WiFi baru, menambah atau mempercepat akses WiFi,

dan mengganti kabel. PT. Telkom sangat mementingkan kepuasan pelanggan dalam pelayanannya.

Perkembangan dunia yang pesat di Teknologi data telah menyebar ke berbagai bidang. Kemajuan pesat ini mendukung pergeseran pemikiran tentang bagaimana melakukan sesuatu dengan cara yang lebih sederhana dan lebih efektif. Pola pikir orang berubah sebagai akibat dari teknologi informasi, dan mereka mencari yang lebih mudah dan lebih efektif. cara untuk melakukan pekerjaan mereka. Teknologi mengubah konsep tradisional yang belum berkembang menjadi konsep modern dan digital. Sebagian besar aktivitas sekarang digital. Banyak hal yang telah didigitalkan masih diubah, dan yang baru menjadi lebih sederhana dan lebih efisien. (Lubis & Zufria, 2022)

Dengan bertambahnya jumlah pengguna Internet di Indonesia, demikian pula jumlah pelanggan PT. Telkom Akses. Namun karena banyaknya pertanyaan tentang pesanan dari pelanggan telekomunikasi, beberapa faktor mempengaruhi waktu layanan untuk pelanggan telekomunikasi, sehingga status layanannya normal dan ada yang tidak normal.

Pada permasalahan di atas, peneliti mengimplementasikan Metode K-nearest neighbor digunakan dalam data mining informasi pelanggan telekomunikasi untuk menentukan status layanan dan melaporkan kepada administrator untuk memastikan bahwa administrator dapat menemukan waktu untuk menggunakan dan memungkinkan teknisi untuk mengoptimalkan lebih baik dalam proses pelayanan. (Agusetiana et al., 2022)

1.1. PT. Telkom Akses

PT.Telkom Access (PTTA) adalah anak perusahaan industri Telkom yang mengerjakan pembuatan dan administrasi layanan infrastruktur jaringan. PT. Telkom Access, didirikan pada 12 Desember 2012.PT.Telekomunikasi Indonesia, Tbk (Telkom) memiliki 100% saham Telkom Access (PTTA) yang merupakan anak perusahaan PT Telekomunikasi Indonesia, Tbk (Telkom). PTTA bergerak ke arah pengelolaan infrastruktur jaringan dan penyediaan layanan arsitektur.

1.2. Kualitas Pelayanan

Melayani orang lain adalah cara langsung untuk memenuhi kebutuhan mereka melalui kegiatan. Kualitas layanan terkait dengan baik atau buruknya layanan. Ukuran seberapa baik tingkat layanan memenuhi harapan pelanggan adalah kualitas layanan. Kemampuan fasilitator layanan memenuhi kebutuhan pelanggan secara konsisten berhubungan langsung dengan kualitas layanan. Kualitas layanan adalah upaya untuk memuaskan kebutuhan pelanggan dan memberikan layanan yang melebihi harapan mereka.

1.3. Pelayanan Telkom

Algoritma untuk menemukan grup K dalam data pelatihan yang sangat mirip dengan data uji coba dan melakukan pencarian merek dalam kategori tertentu dalam cakupan informasi dikenal sebagai K-Naerest Neighbor, yang merupakan interpretasi lain dari KNN. (Haryadi, 2012)

Kelebihan dari algoritma KNN adalah kemampuan untuk melatih informasi dengan banyak noise, dan bekerja paling baik ketika terdapat banyak informasi pelatihan. Algoritma KNN, di sisi lain, memiliki sejumlah kekurangan, termasuk persyaratan untuk menentukan nilai K benchmark (jumlah tetangga terdekat), pelatihan didasarkan pada ketidakjelasan jarak mengenai jenis jarak yang harus digunakan dan fitur yang harus digunakan untuk mendapatkan hasil terbaik, serta biaya komputasi yang tinggi akibat perhitungan jarak antara setiap instance kueri dan total sampel pelatihan. Berikut rumus pencarian jarak menggunakan rumus Euclidean : (Reza Noviansyah et al., 2018)

$$d_{ij} = \sqrt{\sum_{i=1}^p (x_{ik} - x_{jk})^2} \dots\dots\dots 1$$

Penjelasan dari formula umum algoritma *K-Nearest Neighbor* adalah sebagai berikut:

XJK = sampel data/data training

Xik = data uji atau testing

Ij = variabel data

d = jarak

p = dimensi

Jarak Euclidean Ini cenderung dianggap sebagai jarak yang sangat singkat dari 2 tempat, dan pada dasarnya situasi konversi Pythagoras ketika digunakan dalam 2 struktur. Karena jarak selalu memiliki nilai positif, tidak perlu menghitung basis 2 saat menggunakan Euclidean fungsi jarak untuk menyamakan jarak. (Syukriyawati et al., 2013)

1.4. WEKA

WEKA termasuk citra, pengelompokan, penyembuhan, aturan federasi, dan alat untuk pra-pemrosesan data. Pelanggan tidak perlu menulis kode program apa pun untuk memproses data, memasukkannya ke dalam desain kursus, atau menganalisis pengklasifikasi yang dihasilkan.

Di WEKA, Anda dapat mencoba berbagai metode untuk penggalian informasi, salah satunya yang terkenal adalah pohon keputusan. Salah satu cara paling menarik untuk mengelompokkan tumbuhan tetap adalah melalui tumbuhan permanen, yang menghubungkan arsitektur tumbuhan tetap, yang memiliki simpul tetap yang dihubungkan oleh celah dari simpul dasar ke simpul daun (akhir). Setiap hasil akan menghasilkan pembuatan agen dalam bentuk penentuan karakteristik. Agar setiap agen membuat keputusan, mereka akan ditampilkan ke node lain atau node akhir. Plant, menurut Maimon dan Rokach, adalah strategi pengelompokan yang dikenal sebagai partisi rekursif. Node yang menghasilkan tanaman bersumber membentuk pabrik penyediaan, dan setiap node hanya memiliki satu input. Node uji adalah titik keluar node. Node yang tersisa disebut sebagai node keputusan atau node daun. Ketika memilih dua atau lebih sub-ruang, setiap node akan dibagi sesuai dengan jumlah jenis masalah dan sesuai dengan jenis fitur. Kasus menghasilkan tanaman penentu ion, yang menghasilkan masalah yang dapat dipecahkan. Determinasi tanaman adalah diagram alir seperti pohon di mana setiap simpul menggambarkan pengujian pada suatu fitur, setiap agen menggambarkan hasil pengujian, dan simpul daun menggambarkan kategori. Pabrik keputusan biasanya digunakan untuk mengumpulkan data untuk suatu keputusan. Pengguna dapat mengutip tindakan dari simpul akar, atau titik awal, di pabrik penyediaan. Kecocokan untuk algoritme pohon keputusan dimuat oleh pengguna dari simpul akar ini. Kesimpulan adalah tanaman tekad, dengan masing-masing cabang menunjukkan naskah potensial untuk keputusan yang dibuat dan hasilnya. (Pujiono et al., 2013)

2. METODE

2.1.1 Data Mining

Sederhananya, Ekstraksi dan "penambangan" pengetahuan dari sejumlah besar informasi dikenal sebagai penambangan data. Penambangan data, menurut Gartner Group Larose, memeriksa file data yang disimpan dalam memori besar untuk mengidentifikasi asosiasi, pola, dan metode yang bermakna menggunakan identifikasi pola statistik dan matematis teknik. Penambangan data dapat ditafsirkan dengan cara yang berbeda. Singkatnya, penambangan data adalah proses penggalian informasi yang berguna dan pola menarik dari data dalam jumlah besar. Basis data, situs web, repositori, atau data semuanya dapat berfungsi sebagai basis data. Tujuan dari data mining adalah

menggunakan dan mengolah data untuk memperoleh informasi baru dan berguna. (Dunhan, 2013)

2.2.1 Klasifikasi

Klasifikasi adalah layanan untuk membuat formulir (penggunaan) yang menggambarkan dan membedakan kategori informasi atau desain yang dimaksudkan untuk memperhitungkan kategori mata pelajaran yang penunjukan kelasnya ambigu. Ada dua langkah yang terlibat dalam pengelompokan informasi. Yang pertama adalah pelatihan, atau langkah pelatihan pembibitan, di mana algoritme pengelompokan dikembangkan dan disajikan dalam bentuk aturan pengelompokan untuk tujuan menganalisis data pelatihan pembibitan. Pengelompokan, metode kedua, memanfaatkan data eksperimen untuk berspekulasi tentang keakuratan aturan klasifikasi. (Muslim Rejeki & Tarmuji, 2018)

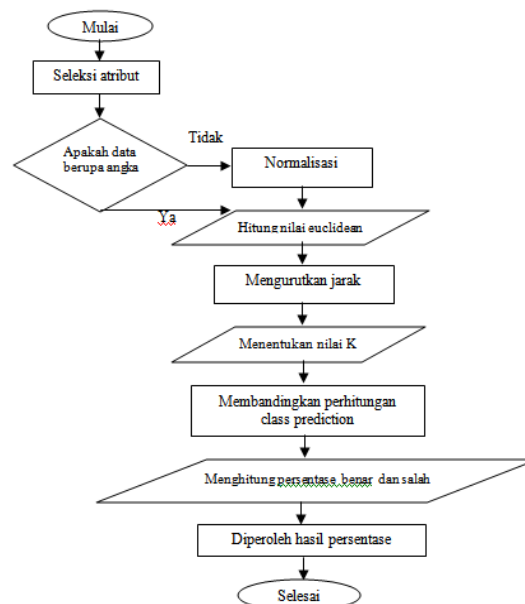
2.3.1 Metode K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah algoritme yang dapat digunakan tidak hanya untuk mengelompokkan berbagai hal tetapi juga untuk memperkirakan dan memperkirakan. K-Nearest Neighbor adalah contoh dasar pemutakhiran yang menyertakan informasi tentang pemutakhiran pembibitan. Ini memungkinkan kita menemukan klasifikasi data yang tidak diklasifikasikan dibandingkan dengan kumpulan data pelatihan. (Ong, 2013)

Penambangan dunia nyata, khususnya pengelompokan. Hampir semua program dapat menjalankan perangkat lunak ini, yang berorientasi pada subjek dan ditulis dalam bahasa Jawa. tingkat kategori. Pada sejumlah tingkatan, WEKA mudah digunakan dan diterapkan. Ini memiliki implementasi algoritma pembelajaran canggih yang dapat diterapkan ke kumpulan data dari baris perintah. (Marks Hall, 2007)

2.4.1 Flowchart

Berikut ini merupakan gambaran flowchart metode klasifikasi algoritma K-NN pada data training dan data testing untuk memprediksi status pelayanan di PT. Telkom Akses Pematangsiantar:



Gambar 1. Flowchart Proses KNN

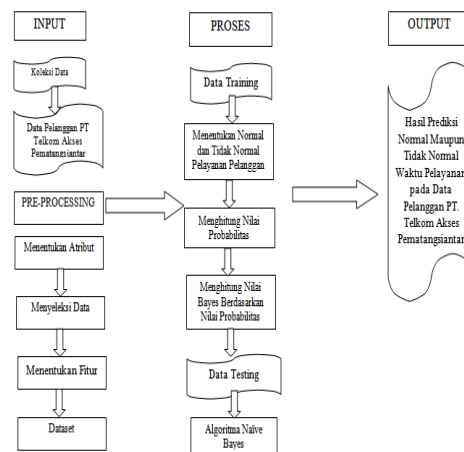
Dengan adanya gambaran flowchart diatas, maka tahapan secara umum sesuai klasifikasi algoritma knn pada data training sebagai berikut:

1. Mulai
2. Menentukan nilai K yang digunakan
3. Menghitung jarak Euclidean
4. Mengurutkan hasil pengurutan jarak
5. Memilih objek yang termasuk dalam nilai K
6. Diperoleh hasil prediksi
7. Selesai

3. HASIL DAN PEMBAHASAN

3.1. Diagram Klasifikasi

Perancangan diagram klasifikasi dengan menggunakan algoritma knn yaitu sebagai berikut:



Gambar 2. Diagram Klasifikasi KNN

3.2. Data Pelanggan Telkom

Data yang digunakan sebagai penelitian adalah data request order dari pelanggan PT. Telkom Akses Pematangsiantar. Data yang akan digunakan untuk memprediksi normal maupun tidaknya waktu pelayanan kepada pelanggan Telkom adalah sebanyak 10.121 data yang diperoleh dari bulan September-Oktober tahun 2022.

Table 1. Data Pelanggan Telkom

NO	ONU SN	STO	STATUS
1	ZTEGCF55G3910	SDK	TIDAK NORMAL
2	48575443BD8A73A3	PMS	TIDAK NORMAL
3	487543291ACX9E	PMS	TIDAK NORMAL
4	4648545499BB59B8	PMS	NORMAL
5	48575443318A109E	PMS	TIDAK NORMAL
6	4857544388E5DD9A	PMS	TIDAK NORMAL
7	48575443ADE751A0	PMS	TIDAK NORMAL
8	485754436C3D62A4	PMS	TIDAK NORMAL
9	48575443D3DA7CA3	PMS	TIDAK NORMAL
10	4857544313480CA5	SIH	TIDAK NORMAL

3.3. Pre-processing

Pada tahap ini dilakukan pre-processing data berupa penentuan atribut dan fitur, seleksi atribut, penentuan dataset, penentuan data training dan penentuan data testing pada data pelanggan Telkom Akses.

a) Variabel dan Fitur

Variabel adalah kumpulan bagian dari suatu entitas. Data klien Telekom Access terdiri dari tiga karakteristik dan satu kategori prediktif, status. Fitur, di sisi lain, adalah konten variabel. Informasi dari PT. Telkom Access Pematangsiantar dapat mengetahui variabel dan karakteristik.

b) Pemilihan Atribut

Dari data yang diterima, telah dilakukan tahap pemilihan atribut, yang ditangani oleh tiga atribut termasuk ONU SN, STO, dan Status. Berikut adalah hasil preprocessing data menjadi data request order.

Table 2. Dataset 100

NO	ONU SN	STO	STATUS
1	5A544547CEFD55A1	SIH	NORMAL
2	5A544547CE996A83	BLG	NORMAL
3	48575443AB540095	BLG	TIDAK NORMAL
4	ZTEGC852AE4A	TJB	TIDAK NORMAL
5	ZTEGC4574F51	TJB	TIDAK NORMAL
6	ZTEGCC4826F8	TJB	TIDAK NORMAL
7	ZTEGC88178BF	GNT	TIDAK NORMAL
8	48575443B3D45A0	TBT	TIDAK NORMAL
9	48575443B3C5A39D	TBT	TIDAK NORMAL
10	48575443CD6597A0	TUK	TIDAK NORMAL

Table 3. Data Training 60

NO	ONU SN	STO	STATUS
1	4648545499BC3990	PYB	TIDAK NORMAL
2	5A544547CF4EF3E7	PYB	NORMAL
3	48575443661C1A66	PYB	TIDAK NORMAL
4	5A544547CE94F768	DLS	NORMAL
5	48575443F94E3DA2	DLS	TIDAK NORMAL
6	4A454547CE94F4D4	DLS	NORMAL
7	475754436476879E	RAP	TIDAK NORMAL
8	48575443DDEB649E	RAP	TIDAK NORMAL
9	48575443825467A4	RAP	TIDAK NORMAL
10	4648545498A24F70	RAP	TIDAK NORMAL

Table 4. Data Testing 40

NO	ONU SN	STO	STATUS
1	485754431385AEA5	PMS	?
2	5448545499AA0998	PMS	?
3	48575443A59406A2	PMS	?
4	545445479006107C	TJB	?
5	48575443D3CA5AA3	PMS	?
6	48575443BE45399D	KPI	?
7	48575443D3DFA3A3	TGD	?
8	5A544547CCB8F8DB	TGD	?
9	485754432E0DAC9F	TGD	?
10	48575443BD8062A3	BGO	?

3.4. Pembahasan

Bagian pembahasan menjelaskan langkah-langkah untuk melakukan perhitungan menggunakan metode Naive Bayes untuk memprediksi status layanan normal atau tidak normal untuk data pelanggan Telkom Access. Data pelanggan diuji dengan alat weka.

3.4.1 Beranda Weka

Halaman ini merupakan halaman saat weka dibuka. Weka memiliki lima menu: *Explore*, *Experimenter*, *KnowledgeFlow*, *Workbench* dan *Simple CLI*. Data diuji di menu *Explore*

3.5.1 Halaman Menu Explorer

Halaman menu explorer memiliki lima pilihan menu: preprocessing, klasifikasi, cluster, asosiasi, pemilihan atribut, dan visualisasi. Tahap pengujian ini dijalankan pada menu *Preprocessing* dengan membuka *file* dengan *Open File* dan memasukkan data yang akan diuji yaitu data training.

3.6.1 Halaman Atribut Data Training

Setelah memasukkan *file*, akan terlihat atribut yang terdapat pada data pelanggan. Halaman Atribut menggunakan tiga atribut, antara lain ONU SN, STO, dan Status. Kelas yang digunakan untuk pengujian memiliki status dengan nilai Normal 11, dan 49 Tidak Normal dari 60 data pelanggan.

Untuk melihat atribut data training yang telah diubah pada format arff adalah masuk ke halaman awal Weka, masuk ke menu *Tools* lalu pilih *ArffViewer*. Pilih Masukkan data dengan memilih *File* dan klik *Open* untuk memasukkan data yang ingin digunakan

No.	1	ONU SN	2	STO	3	STATUS
31	48485454...	PVB	TIKAK NO...			
32	54544547...	PVB	NORMAL			
33	48575443...	PVB	TIKAK NO...			
34	54544547...	DLS	NORMAL			
35	48575443...	DLS	TIKAK NO...			
36	48485454...	DLS	NORMAL			
37	47975443...	RAP	TIKAK NO...			
38	48575443...	RAP	TIKAK NO...			
39	48575443...	RAP	TIKAK NO...			
40	48485454...	RAP	TIKAK NO...			
41	48575443...	RAP	NORMAL			
42	48575443...	RAP	TIKAK NO...			
43	48485454...	RAP	TIKAK NO...			
44	48475443...	RAP	TIKAK NO...			
45	48575443...	RAP	TIKAK NO...			
46	48575443...	RAP	NORMAL			
47	48475443...	PMS	NORMAL			
48	5875443...	PMS	TIKAK NO...			
49	48575443...	PMS	TIKAK NO...			
50	48575443...	PMS	TIKAK NO...			
51	5754431A...	PMS	TIKAK NO...			
52	48575443...	PMS	TIKAK NO...			
53	48575443...	PMS	TIKAK NO...			
54	48575443...	KS	TIKAK NO...			
55	54544547...	PVB	NORMAL			
56	48485454...	PMS	NORMAL			
57	48575443...	PMS	TIKAK NO...			
58	48575443...	PMS	TIKAK NO...			
59	48485454...	PMS	TIKAK NO...			
60	48485454...	PMS	NORMAL			

Gambar 3. Halaman Viewer Atribut Data Training

Menunjukkan halaman tampilan yang digunakan untuk menampilkan modifikasi atribut data training dalam format arff. Halaman ini menggunakan 40 atribut data Training.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, maka dapat ditarik kesimpulan sebagai berikut.

- 1) Prediksi klasifikasi status pelayanan pada data pelanggan Telkom Akses dilakukan dengan menggunakan metode klasifikasi KNN serta menggunakan 3 variabel yang telah ditentukan.
- 2) Hasil prediksi data pelanggan Telkom Akses dibagi menjadi 100 dataset. 60% dari setiap dataset digunakan sebagai data pengujian, dan 40% dari setiap dataset digunakan sebagai data pelatihan.
- 3) Hasil yang diperoleh dengan menggunakan weka untuk mendapatkan nilai kebenaran 81% dan 19% nilai kesalahan.

4.1. SARAN

Penelitian berikut dapat menggunakan data dari PT. Telkom Akses yang diperoleh dari kota lain. Bandingkan hasil klasifikasi metode KNN dengan metode yang berbeda menggunakan jumlah dataset yang sama.

5. DAFTAR PUSTAKA

- Agusetiana, E., Fitriani, A. S., & Kunci, K. (2022). *Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode K-Nearest Neighbor untuk Memprediksi Status Pelayanan*. 115–119.
- Dunhan, M. H. (2013). *Data Mining: Introductory and Advanced Topics 1st Edition. Engineering*, 1–89.
- Haryadi, T. S. (2012). *Penerapan Algoritma K-Means Untuk Pengelompokan Data Nilai Siswa*. 1–9.
- Lubis, M. I. Y., & Zufria, I. (2022). Perancangan Aplikasi Pendataan Valins Yang Tervalidasi Oleh PT. Telkom Witel Medan Berbasis Web. *Aisyah Journal of Informatics and Electrical Engineering*, 4(1), 33–41.
<http://jti.aisyahuniversity.ac.id/index.php/AJIEE%0Ahttps://doi.org/10.30604/jti.v4i1.100>
- Marks Hall, G. H. (2007). *WEKA: Practical Machine Learning Tools and Techniques with Java Implementations*.
<https://researchcommons.waikato.ac.nz/bitstream/handle/10289/1040/uow-cs-wp-1999-11.pdf?sequence=1&isAllowed=y>
- Muslim Rejeki, S., & Tarmuji, A. (2018). *Memebangun Aplikasi Autogenerate Script ke Flowchart Untuk Mendukung Business Process Reengineering*. 1(2), 448–456.
- Ong, J. O. (2013). Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university. *Jurnal Ilmiah Teknik Industri*, vol.12, no(juni), 10–20.
- Pujiono, S., Amborowati, A., Suyanto, M., & Kunci, K. (2013). Analisis kepuasan publik menggunakan weka dalam mewujudkan. *Jurnal DASI*, 14(2), 45–55.
- Reza Noviansyah, M., Rismawan, T., & Marisa Midyanti, D. (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). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48–56.
- Syukriyawati, G., Fidana, G., Dwi, R., Nilasari, A., & Risky, D. (2013). *Data Mining Penerapan Metode K-Mean Clustering Untuk Pengelompokkan Mahasiswa Universitas Brawijaya Tahun 2012 Disusun Oleh: Ulya ' Darojatun Nisa ' Reysia Dwi Angga Nilasari Dika Risky*.