

Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode *K-Nearest Neighbor*

Arwa Ulayya Haspriyanti¹, Putri Wahyu Prasetyaningrum²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
Yogyakarta, Indonesia

e-mail: ¹alyalfa96@gmail.com, ²putri@mercubuana-yogya.ac.id

Diajukan: 19 Agustus 2020; Direvisi: 25 Februari 2021; Diterima: 16 Maret 2021

Abstrak

PT. Telkom adalah perusahaan yang bergerak dalam bidang telekomunikasi dengan produk IndiHome yang merupakan gabungan dari jenis produk dan layanan telecommunication, information, media dan edutainment menjadi satu terdiri dari layanan internet, internet voicete dan internet interaktif (UseeTV cable) atau Triple Play. Pengembangan strategi pemasaran yang maksimal memerlukan bauran pemasaran efektif dan efisien. Pada prinsipnya, strategi pemasaran memberikan arah dalam kaitannya dengan variabel-variabel seperti segmentasi pasar, target, positioning, dan elemen bauran pemasaran. Untuk menunjang penelitian ini, maka variabel yang digunakan adalah variabel operasional untuk analisis eksternal dan internal, Dalam penelitian ini metode yang digunakan adalah metode penelitian kualitatif. Sedangkan jenis penelitiannya adalah deskriptif. Data primer diperoleh langsung dari subjek penelitian melalui observasi, wawancara, dan dokumentasi. Dilihat dari banyaknya permintaan konsumen akan produk layanan berdasarkan data penjualan selama 1 tahun terakhir maka dibutuhkan prediksi untuk penjualan produk layanan terlaris, guna untuk mempermudah pihak perusahaan dalam perencanaan penyediaan layanan. Untuk mengetahui penjualan produk layanan terlaris digunakan teknik klasifikasi data mining dan algoritma K-Nearest Neighbor. Hasil dari penelitian ini adalah prediksi penjualan layanan terlaris sebanyak 1 jenis produk dari 3 jenis produk yang terjual yaitu Internet, Internet Voice, Internet Voice Useetv. Berdasarkan nilai akurasi terhadap klasifikasi penjualan produk terlaris sebesar 99,99 %.

Kata kunci: Telekomunikasi, internet, K-Nearest Neighbor

Abstract

PT. Telkom is a company engaged in the field of telecommunication that develops IndiHome product. IndiHome itself is the combination of product and service types including telecommunication, information, media and edutainment becomes one consisting internet service, internet voicete and interactive internet (UseeTV cable) or Triple Play. The development of a maximum marketing strategy requires an effective and efficient marketing mix. In principle, the marketing strategy provides direction in relation with variables such as market segmentation, target, positioning, and marketing mix element. In order to support this study, the variable used is operational variable for external and internal analysis. In this study, the method used is qualitative research method. Meanwhile, the type of the study is descriptive. The primary data is directly obtained from the research subject through observation, interview, and documentation. As seen from the large number of customer request related to the service product based on sales data for the past 1 year, it is important to get the sales prediction of best-selling service product, in order to facilitate the company in developing service provision planning. Furthermore, the use of data mining classification technique and K-Nearest Neighbor algorithm is needed to find out the best-selling service product sales. The result of this study is the prediction of the best-selling service product sales which is 1 from the total of 3 product types that are sold including Internet, Internet Voice, and Internet Voice Useetv. It is based on the accuracy value towards the classification of the best-selling product sales of 99.99%.

Keywords: telecommunications, internet, K-Nearest Neighbor.

1. Pendahuluan

JISAI Vol. 1, No. 2, (Mei) (2021).

Telkom merupakan perusahaan yang menyediakan jasa dalam bidang telekomunikasi terkemuka di Indonesia, dengan demikian Telkom wajib memberikan layanan terbaik kepada pelanggannya. Laju perkembangan teknologi akses jaringan dan teknologi berbasis internet memungkinkan bagi operator layanan untuk dapat menyediakan layanan yang beragam (*multi service*) seperti layanan suara (telepon), data (internet) dan gambar (*IP-TV*) dalam satu jaringan akses dengan layanan *triple-play*. IndiHome merupakan layanan *Triple Play* dari Telkom yang terdiri dari *Internet on Fiber* atau *High Speed Internet*, *Phone (Telepon Rumah)*, dan *IPTV (UseeTV Cable)* (Telkom, 2019) [1].

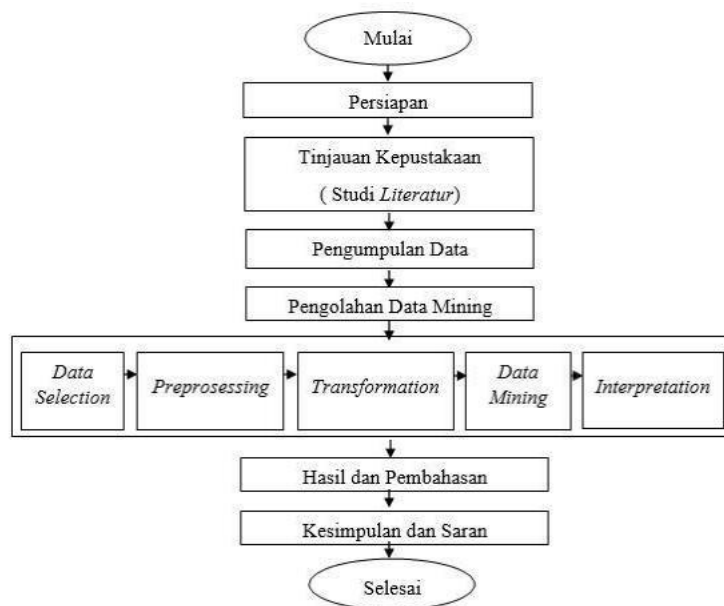
Salah satu permasalahan yang dihadapi oleh perusahaan yang bergerak dibidang penjualan jasa adalah mempertahankan layanan-layanan tersebut. Dengan melakukan prediksi dapat membantu pihak Telkom untuk mengetahui apakah pelanggan Telkom nyaman terhadap layanan produk yang disediakan yaitu Indihome. Layanan dapat dikatakan baik ketika banyak pelanggan yang berminat dan membutuhkan layanan Indihome sesuai dengan kebutuhan masing-masing. Prediksi atau peramalan penjualan (*forecasting*) adalah suatu perhitungan untuk meramalkan keadaan di masa mendatang melalui pengujian keadaan di masa lalu. Meramalkan penjualan di masa mendatang berarti menentukan perkiraan besarnya volume penjualan, bahkan menentukan potensi penjualan dan luas pasar yang dikuasai di masa yang akan datang (Eriyanto, 2012). Salah satu dari kegunaan prediksi adalah untuk membantu pemilik perusahaan dalam mengambil keputusan dalam menentukan jumlah barang yang harus disediakan oleh perusahaan. Selain itu prediksi dapat membantu pihak perusahaan dalam perencanaan penyediaan layanan, karena prediksi ini dapat memberikan *output* terbaik sehingga diharapkan resiko kesalahan yang disebabkan oleh kesalahan perencanaan dapat ditekan seminimal mungkin. Prediksi biasanya digunakan untuk menemukan informasi dari sejumlah data yang besar sehingga diperlukan *data mining*. *K-Nearest Neighbor* adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada KNN. Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. KNN juga merupakan metode yang baik digunakan untuk melakukan suatu prediksi. KNN merupakan metode klasifikasi yang mencari nilai K (cluster) atau tetangga terdekat dalam menentukan kelas pada data baru[2]. Dikarenakan hal tersebut maka penelitian dilakukan dengan metode manual maupun metode *K-Nearest Neighbour* untuk mengetahui hal tersebut[3]. Algoritma *K-Nearest Neighbor* adalah salah satu algoritma yang bisa dimanfaatkan untuk implementasi pengklasifikasiannya[4]. Algoritma *K-Nearest Neighbor (KNN)* yang merupakan salah satu metode klasifikasi pada data mining dan juga menjadi algoritma supervised learning pada machine learning adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut[5]. Untuk mengetahui penjualan produk elektronik terlaris digunakan teknik klasifikasi data mining dan algoritma *K-Nearest Neighbor* [6]. Metode KNN digunakan untuk memprediksi kelulusan seorang mahasiswa di sebuah matakuliah berdasarkan data sebelumnya [7]. Berdasarkan hal tersebut, penelitian ini telah membandingkan metode klasifikasi data mining untuk menganalisis prediksi kelayakan pemberian kredit dengan metode *K-NN*, *Naïve Bayes* dan *Decision Tree*. Data-data calon debitur yang telah melalui tahapan data mining akan diproses menggunakan metode klasifikasi data mining yaitu *K-NN*, *Naïve Bayes* dan *Decision Tree*[8]. Penelitian ini menerapkan algoritma *K-NN* dalam menentukan koperasi yang layak mendapatkan pembinaan. Hasil yang dari penelitian ini adalah klasifikasi koperasi yang layak mendapatkan pembinaan dengan akurasi yang diperoleh sebesar 96,33% [9]. Berdasarkan hasil pengujian, diperoleh rata-rata nilai akurasi metode *K-Nearest Neighbor* sebesar 82,42% dan rata-rata nilai akurasi metode *Naïve Bayes* sebesar 70,32%. Dapat disimpulkan bahwa metode yang paling baik untuk klasifikasi kualitas air bersih adalah metode *K-Nearest Neighbor*[10].

Oleh karena itu, berdasarkan permasalahan diatas, peneliti tertarik untuk melakukan penelitian dengan judul "Penerapan Data Mining Untuk Prediksi Penjualan Produk Indihome Terlaris Menggunakan Metode *K-Nearest Neighbor*" dengan cara melakukan pengolahan data dan penerapan data mining dengan memanfaatkan data-data terdahulu sehingga dapat menghasilkan informasi dan rangkaian pola yang bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya guna memecahkan permasalahan tersebut. Selanjutnya adalah penelitian yang dibuat yaitu untuk memperkirakan produk layanan Indihome terlaris dengan menggunakan metode *K-Nearest Neighbor*. Perbedaan dengan penelitian sebelumnya adalah lokasi penelitian yang akan dikerjakan dan atribut yang digunakan yaitu wilayah pemasangan, package name, jenis layanan, serta rumusan masalah yang berbeda, untuk kasus saat ini di kerjakan dengan tujuan untuk mengetahui produk layanan indihome terlaris. Alasan pemilihan atribut yaitu efisiensi dari data dan hasil akan mempermudah manager dalam mengetahui layanan mana yang dapat di

pertahankan untuk produksi tahun berikutnya Sedangkan metode K-Nearest Neighbor atau k-NN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (*train data sets*), yang diambil dari k tetangga terdekatnya (*nearest neighbors*). Dengan k merupakan banyaknya tetangga terdekat.

2. Metode Penelitian

Tahapan penelitian yang dituangkan dalam diagram alir ini. Menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan. Tahapan yang akan ditempuh yaitu:



Gambar 1. Tahapan Penelitian

Selanjutnya Penentuan nilai *k*. Penentuan nilai *k* yang digunakan dalam klasifikasi tidak memiliki aturan yang baku, namun pada penelitian ini nilai *k* yang digunakan adalah 3, 5, 7, 9, 11, 13, 15.

Perhitungan jarak antar data *training* dan data uji (*test*). Teknik perhitungan jarak yang digunakan dalam metode KNN ini adalah Jarak *Euclidean Distance*. Dengan rumus sebagai berikut

$$\sqrt{\sum_{i=1}^k (X_i - Y_i)^2}$$

Menentukan kelompok data hasil uji berdasarkan label mayoritas dari *k* tetangga terdekat.

3. Hasil dan Pembahasan

Setelah melakukan analisa terhadap perancangan dengan tahapan *data mining* untuk menghasilkan prediksi penjualan produk layanan terlaris pada PT.Telkom Indonesia di Yogyakarta dengan menggunakan Algoritma *K-Nearest Neighbor*, analisa ini berakhir dengan melakukan proses *data mining* yang sesungguhnya, maka hasil yang dicapai oleh peneliti adalah untuk mengetahui prediksi penjualan layanan terlaris berdasarkan data penjualan produk layanan dari data penjualan di tahun 2019 yang ada di PT.Telkom Indonesia di Yogyakarta. Peneliti menggunakan aplikasi *RapidMiner* untuk mempermudah proses *data mining*, yang menghasilkan informasi prediksi penjualan produk indihome terlaris. Dan didapatkan hasil prediksi penjualan produk terlaris sebanyak 3 jenis produk diantaranya Layanan 1P (Internet), Layanan 2P (Internet Voice), Layanan 3P (Internet Voice Usestv).

Data yang digunakan dalam penelitian ini yaitu data penjualan produk layanan berdasarkan penjualan 1 tahun terakhir yaitu tahun 2019 yang berasal dari PT.Telkom Indonesia regional Yogyakarta, dengan data sebagai berikut :

Tabel 1. *Data Selection*

No	Nama Produk	Kuantitas	Last Update
1	INTERNET	1	January 2019
2	INTERNET	1	January 2019
3	INTERNET	1	January 2019
4	INTERNET	1	January 2019
5	INTERNET	1	January 2019
6	INTERNET	1	January 2019
7	INTERNET	1	January 2019
8	INTERNET	1	January 2019
9	INTERNET	1	January 2019
10	INTERNET	1	January 2019

Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang paling jauh (*ascending*).

Tabel 1 Merupakan data yang telah diselection dan siap digunakan untuk proses data mining. Jumlah data yang di proses untuk pengujian adalah 39.631 data.

Tabel 2. *Preprocessing 1*

Jenis Layanan	Januari	Februari	Maret	April
Single Play	0	8	1	4
Single Play	5	4	0	2
Single Play	6	2	0	4
Single Play	5	1	0	5
Single Play	1	1	4	3
Single Play	0	0	3	3
Single Play	5	5	4	2
Single Play	5	0	1	4
Single Play	4	6	2	3

Tabel 2 Merupakan hasil pengelompokan penjualan produk layanan single play untuk setiap bulan dengan total penjualan sebanyak 967 item.

Tabel 3. *Preprocessing 2*

Jenis Layanan	Januari	Februari	Maret	April
Dual Play	12	48	56	57
Dual Play	40	53	56	76
Dual Play	51	33	36	46
Dual Play	58	60	68	67
Dual Play	65	41	77	53
Dual Play	36	69	71	46
Dual Play	76	63	63	28
Dual Play	73	42	61	46

Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearst Neighbor (Arwa)

Jenis Layanan	Januari	Februari	Maret	April
Dual Play	60	52	64	52

Tabel 3 Merupakan hasil pengelompokan penjualan produk LED untuk setiap bulan dan tahun dengan total penjualan sebanyak 19518 item.

Tabel 4. Preprocessing 3

Jenis Layanan	Januari	Februari	Maret	April
Triple Play	19	32	27	26
Triple Play	37	47	39	30
Triple Play	42	22	25	29
Triple Play	41	43	45	50
Triple Play	50	24	34	23
Triple Play	30	52	28	43
Triple Play	37	41	38	30
Triple Play	59	23	27	32

Tabel 4 merupakan hasil pengelompokan penjualan produk Layanan Internet Voice Usee Tv untuk setiap bulan dengan total penjualan sebanyak 19146 item.

Selanjutnya Perhitungan jarak antar data *training* dan data uji (*test*) yang ada pada tahap *transformation* dengan menggunakan perhitungan *Euclidean Distance* sebagai berikut :

$$d1 = \sqrt{(0-3)^2+(8-2)^2+(1-4)^2+(4-2)^2+(2-2)^2} = 1,7222$$

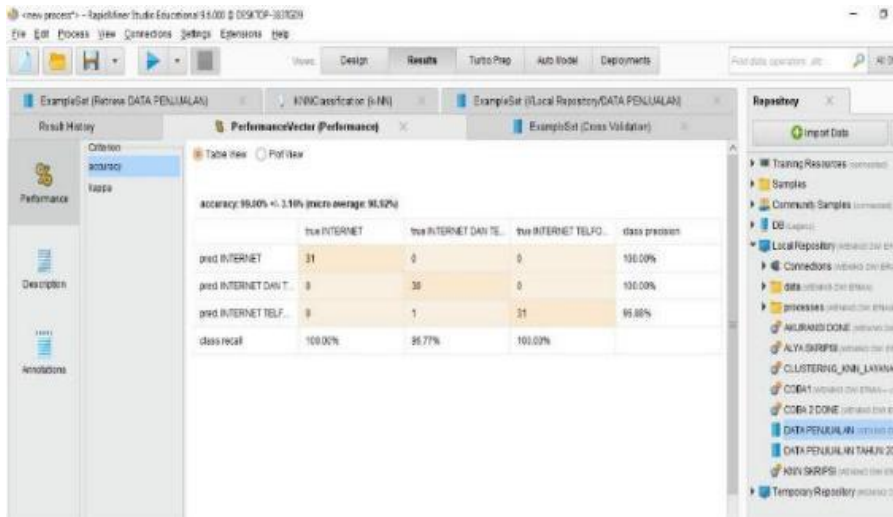
$$d2 = \sqrt{(5-3)^2+(4-2)^2+(0-4)^2+(2-2)^2+(6-2)^2} = 2.0466$$

$$d3 = \sqrt{(6-3)^2+(2-2)^2+(0-4)^2+(4-2)^2+(0-2)^2} = 1,927$$

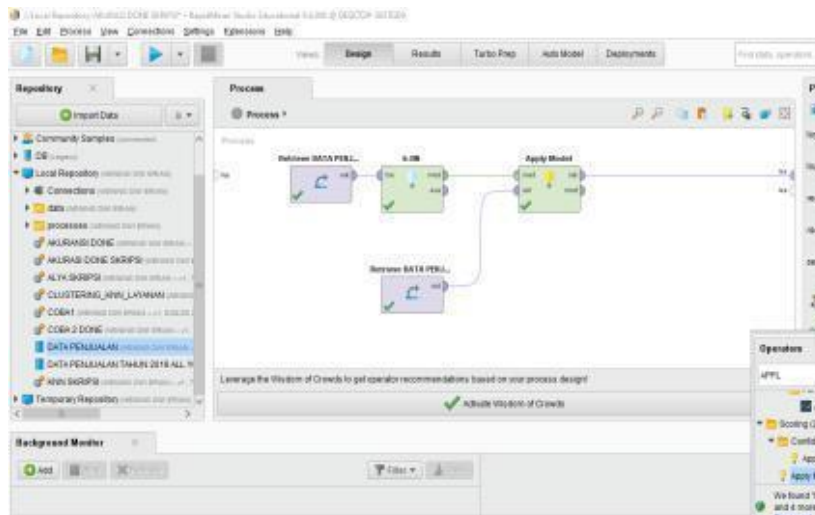
$$d4 = \sqrt{(5-3)^2+(1-2)^2+(0-4)^2+(5-2)^2+(4-2)^2} = 1,9066$$

$$d5 = \sqrt{(1-3)^2+(1-2)^2+(4-4)^2+(3-2)^2+(0-2)^2} = 2.0588$$

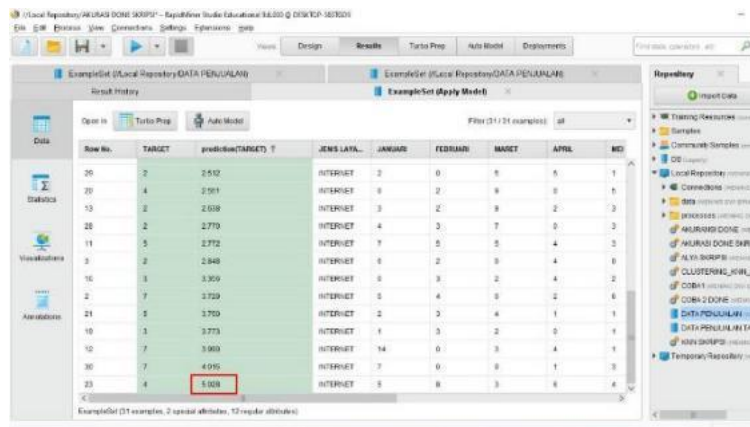
Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang paling jauh (*ascending*). Maka di peroleh hasil sebagai berikut pada implementasi rapid miner :



Gambar 2. Akurasi Data sebesar 99,99



Gambar 3. Susunan operator algoritma k-NN



Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearest Neighbor (Arwa)

Gambar 4. Hasil prediksi layanan single play sebesar 5,028

Row No.	TARGET	predic...	JENIS LAYANAN	JANUARI	FEBRUARI	MARET	APRIL	MEI
18	14	13,981	INTERNET VOICE	86	52	55	38	92
17	13	14,352	INTERNET VOICE	112	39	18	18	95
21	15	15,859	INTERNET VOICE	50	70	47	27	102
14	21	17,005	INTERNET VOICE	51	88	34	30	87
8	10	18,353	INTERNET VOICE	80	52	64	52	43
3	37	22,027	INTERNET VOICE	31	39	26	46	8
10	18	22,451	INTERNET VOICE	77	35	39	58	53
7	49	28,357	INTERNET VOICE	75	83	83	28	59
5	32	29,664	INTERNET VOICE	65	43	77	43	34
4	38	30,735	INTERNET VOICE	58	89	88	67	90
2	43	32,919	INTERNET VOICE	40	53	58	78	49
6	32	32,954	INTERNET VOICE	36	89	71	48	40
1	54	37,427	INTERNET VOICE	12	48	55	57	34

Gambar 5. Hasil prediksi layanan single play sebesar 37,427

Row No.	TARGET	predic...	JENIS LAYANAN	JANUARI	FEBRUARI	MARET	APRIL	MEI	JUN
23	71	79,457	INTERNET V...	49	36	27	45	47	38
19	197	80,426	INTERNET V...	51	38	33	34	0	74
12	77	80,172	INTERNET V...	45	58	25	37	37	59
30	85	83,373	INTERNET V...	35	0	28	47	47	38
17	103	83,603	INTERNET V...	59	19	19	13	0	0
18	90	83,538	INTERNET V...	35	34	29	37	1	0
29	79	85,200	INTERNET V...	33	30	33	37	32	69
28	81	86,284	INTERNET V...	53	38	34	23	45	82
27	80	88,113	INTERNET V...	31	47	26	38	43	83
25	81	90,800	INTERNET V...	40	21	44	48	77	42
24	82	90,187	INTERNET V...	30	18	18	58	38	48
31	104	90,511	INTERNET V...	34	0	22	0	43	0

Gambar 6. Hasil prediksi layanan single play sebesar 90,511

4. Kesimpulan

Pengubahan data primer yang telah diuji menggunakan rapid miner menghasilkan akurasi data sebesar 99,99%, dimana data tersebut dapat digunakan sebagai acuan prediksi pada proses atau tahapan selanjutnya Implementasi data mining pada rapid miner menghasilkan 1 jenis produk layanan terlaris dari 3 produk layanan, dengan perhitungan sebagai berikut : layanan (internet;5,028),(internet voice;37,427), (internet voice usestv;90,511). Pengolahan data sehingga menghasilkan prediksi paling akurat hanya dapat dilakukan dengan menggunakan metode K-Nearest Neighbor karena pengujian telah membuktikan bahwa akurasi data lebih tinggi dibandingkan dengan metode lainnya.

Daftar Pustaka

- [1] Telkom, "Profil PT.Telekomunikasi Indonesia," 2019.
- [2] M. R. Noviansyah, T. Rismawan, and D. M. Midyanti, "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*, vol. 06, no. 2, pp. 48–56, 2018.
- [3] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, pp. 1–13, 2017.
- [4] M. Rivki and A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia," *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017, doi: 10.21609/jsi.v13i1.500.
- [5] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [6] Y. R. Amalia, "Penerapan data mining untuk prediksi penjualan produk elektronik terlaris menggunakan metode k-nearest neighbor," 2018.
- [7] G. A. Pradnyana and A. A. J. Permana, "Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 1, p. 59, 2018, doi: 10.12962/j24068535.v16i1.a696.
- [8] S. Wahyuningsih and D. R. Utari, "Perbandingan Metode K-Nearest Neighbor , Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," *Konf. Nas. Sist. Inf. 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018*, pp. 619–623, 2018.
- [9] Y. A. Setianto, K. Kusri, and H. Henderi, "Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur," *Creat. Inf. Technol. J.*, vol. 5, no. 3, p. 232, 2019, doi: 10.24076/citec.2018v5i3.179.
- [10] M. A. Rahman, N. Hidayat, and A. Afif Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Vol. 2, No. 12, Desember 2018, hlm. 6346-6353 e-ISSN*, vol. 2, no. 12, pp. 925–928, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>