

## PENERAPAN ALGORITMA $K$ – *NEAREST NEIGHBOR* (KNN) UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA JAKARTA

Siti Nurjanah  
Universitas Buana Perjuangan  
Karawang, Indonesia  
If15.Sitinurjanah@mhs.ubpkarawan g.ac.id

Amril Mutoi Siregar Universitas  
Buana Perjuangan  
Karawang, Indonesia  
amrilmutoi@ubpkarawang.ac.id

Dwi Sulistya Kusumaningrum  
Universitas Buana Perjuangan  
Karawang, Indonesia  
dwi.sulistya@ubpkarawang.ac.i

### Abstract—

Pencemaran Udara adalah masuknya atau dimasukkannya zat, energy atau komponen lain ke dalam udara ambien oleh kegiatan manusia. Sehingga mutu udara ambien turun sampai ketinggian tertentu yang menyebabkan udara ambien tidak dapat memenuhi fungsinya. Kondisi udara yang buruk juga dapat mengganggu aktifitas manusia. Dampak pencemaran udara dapat menyebabkan gangguan pernapasan dan bahkan kematian. Sumber pencemaran udara dapat disebabkan dari semakin banyaknya kendaraan bermotor dan pembangunan gedung yang tidak memperhatikan keseimbangan alam. Di Kota Jakarta Indeks Standar Pencemaran Udara (ISPU) yang dilakukan oleh Dinas Lingkungan Hidup melalui stasiun pemantau setiap hari selalu mengalami perubahan kategori pencemaran udara. Data yang ada belum mampu memberikan informasi yang memadai kepada masyarakat Jakarta. Sehingga kesadaran masyarakat terhadap bahaya polusi udara belum maksimal. Salah satu cara untuk mengolah data ISPU harian dikota Jakarta adalah dengan teknik klasifikasi *data mining* menggunakan algoritma *K-Nearest Neighbor* (KNN). Sehingga diharapkan dapat membantu memberikan informasi yang bermanfaat bagi masyarakat dan bagi Dinas Lingkungan Hidup. Berdasarkan hasil perhitungan klasifikasi kategori pencemaran udara dengan algoritma *K-Nearest Neighbor* (KNN) di Kota Jakarta dengan data *training* sebanyak 304 dan satu data uji (*testing*) diperoleh nilai akurasi sebanyak 95.78% dengan menentukan  $K=7$ .

**Kata kunci** — Algoritma *K-Nearest Neighbor* (KNN), Indek Standar Pencemaran Udara (ISPU), Kota Jakarta

### I. PENDAHULUAN

Pencemaran udara merupakan suatu masalah yang berdampak buruk bagi kehidupan makhluk hidup. Udara yang tercemar akan menimbulkan berbagai macam penyakit. Sehingga perlu dilakukan pengamatan tingkat pencemaran udara pada lingkungan masyarakat [1]. Menteri Kesehatan Republik Indonesia, menjelaskan bahwa Pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan atau komponen lain ke dalam udara oleh kegiatan manusia, sehingga mutu udara turun sampai ketinggian tertentu yang menyebabkan atau mempengaruhi kesehatan manusia [2]. Kota Jakarta termasuk dalam kota dengan tingkat pencemaran udara yang buruk di dunia. Pemantauan kualitas udara yang dilakukan oleh Green Peace pada tahun 2017 juga menunjukkan bahwa polusi udara pada wilayah Jabodetabek berada jauh di atas ambang batas kesehatan *World Healty Organizacion* (WHO) dan Baku Mutu Udara Ambien Nasional. Di sisi lain, masyarakat Jabodetabek masih memiliki pemahaman yang terbatas mengenai polusi udara dan bahayanya, terutama karena akses terhadap informasi mengenai kualitas udara yang sangat minim [3]. Sebagai langkah awal, pemerintah harus menyediakan informasi kualitas udara yang memadai dan selanjutnya mengimplementasikan beragam kebijakan untuk memenuhi hak warga terhadap udara yang berkualitas.

Kualitas data harian Indeks Standart Pencemaran Udara (ISPU) yang dilakukan oleh Dinas Lingkungan Hidup selama tahun 2018 melalui stasiun pemantau pada sejumlah titik di Jakarta mengalami perubahan setiap harinya. Data yang ada belum dapat memberikan pemahaman dan informasi yang memadai kepada masyarakat secara terus-menerus. Pemerintah dituntut harus menyediakan informasi kualitas udara lebih memadai dan selanjutnya menerapkannya beragam kebijakan untuk memenuhi hak warga terhadap udara yang berkualitas [4]. Salah satu solusi untuk menyelesaikan adalah dengan teknik komputasi.

Teknik komputasi telah dibuktikan mampu menyelesaikan beberapa masalah. Seperti Penelitian yang dilakukan dengan Algoritma *K-Nearest Neighbor* telah dibuktikan berhasil oleh Dzikrulloh Nafi Nihru, *et al*, tentang topik Penerapan Metode *K-Nearest Neighbor* (KNN) dan Metode *Weighted Product* (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi, bukti bahwa akurasi mencapai 94% [5]. Kemudian Rahman Aditya Maulana, *et a*, tentang topik Komparasi Metode *Data mining K-Nearest Neighbor* Dengan *Naïve Bayes* untuk Klasifikasi Kualitas Air Bersih bukti bahwa akurasi mencapai 82,89% [6]. Penelitian Mustakim, Oktaviani F Giantika tentang topik Algoritma *K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa bukti bahwa akurasi mencapai 82% [7].

Berdasarkan masalah yang telah dipaparkan dan bukti bahwa algoritma KNN mampu menyelesaikan beberapa masalah, maka penelitian ini mengimplementasikan algoritma KNN untuk melakukan

klasifikasi pada data ISPU di wilayah Jabodetabek. Sehingga masyarakat dapat melakukan tindakan pencegahan dan tanggapan terhadap polusi udara.

## II. STUDI LITERATUR

### A. Data mining

Pengertian tentang *Data mining* menurut A.Siregar dan A.Paspabhuana [8] didefinisikan sebagai salah satu teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud disini adalah set data yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam manajemen basis data relasional. Akan tetapi teknik-teknik *data mining* dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spatial, berbasis text, dan multimedia (citra). Sedangkan menurut Aribowo [9], *Data mining* adalah metode untuk menemukan suatu pengetahuan dalam suatu database yang cukup besar. Dalam *data mining* terdapat metode salah satunya adalah metode klasifikasi menggunakan algoritma KNN.

### B. Klasifikasi

Klasifikasi adalah teknik untuk merancang fungsi berdasarkan hasil pengamatan dari data dan atribut. Kemudian dari data tersebut dapat dilakukan pemetaan terhadap data yang belum memiliki kelas kedalam data yang telah terklasifikasi sesuai dengan aturan-aturan yang diberikan [10]:

1. Perancangan model  
Proses membangun solusi untuk menyelesaikan masalah berdasarkan data yang telah terklasifikasi (Data latih)
2. Implementasi Model  
Proses penentuan kelas untuk data uji berdasarkan model fungsi dan parameter-parameter data yang telah ditentukan pada tahap perancangan.
3. Evaluasi Model  
Proses yang bertujuan untuk melakukan evaluasi terhadap hasil implementasi model fungsi dalam mengklasifikasikan data uji berdasarkan parameter-parameter yang telah ditetapkan

### C. Algoritma *K-Nearest Neighbor* (KNN)

KNN adalah suatu metode algoritma *supervised learning*, di mana kelas yang paling banyak muncul (*mayoritas*) yang akan menjadi kelas hasil klasifikasi [11]. Langkah-langkah untuk menghitung metode Algoritma KNN adalah:

- a. Menentukan Parameter K (Jumlah tetangga paling dekat).
- b. Menghitung kuadrat jarak Euclid (*query instance*) masing-masing objek terhadap data sampel yang diberikan.
- c. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
- d. Mengumpulkan kategori YA (Klasifikasi *Nearest Neighbor*).
- e. Dengan menggunakan kategori KNN yang paling mayoritas maka dapat diprediksi nilai *query distance* yang telah dihitung

Persamaan (1) merupakan persamaan untuk Algoritma KNN.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Di mana:

$d(x_i, x_j)$  = Jarak Euclidean

$(x_i)$  = record ke-i

$(x_j)$  = record ke-j

$(a_r)$  = data ke-r

$i, j = 1, 2, 3, \dots, n$

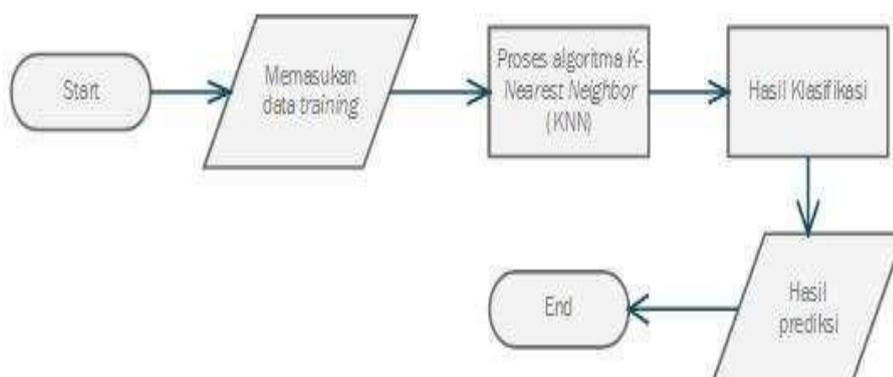
### D. RapidMiner

Rapidminer adalah salah satu software untuk pengolahan *data mining*. Pekerjaan yang dilakukan oleh *RapidMiner* text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan *database* [12].

### III. METODE PENELITIAN

#### A. Gambaran Umum Penelitian

Pada tahap klasifikasi *data mining* dengan algoritma KNN, adapun gambaran penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1 Alur tahapan klasifikasi Algoritma KNN

#### B. Tahapan pencarian dataset

Tahap pencarian dataset dilakukan dengan mencari data website di Data Indonesia dalam satu portal yang bersumber dari Dinas lingkungan Hidup kota Jakarta tahun 2018 Data yang digunakan berjumlah 304 dan 10 atribut sehingga dilakukan pembersihan atribut data [13].

#### C. Pembersihan Atribut data

Dari data yang berjumlah 304 data dan 10 atribut maka dari atribut yang ada akan di lakukan pembersihan dengan cara menghilangkan beberapa atribut yang tidak terpakai dalam proses perhitungan. Seperti atribut Tanggal, Max dan Critical yang terdapat pada metode penelitian. Atribut-atribut tersebut tidak digunakan dalam prediksi karena tidak memiliki pengaruh terhadap penentuan kategori kualitas udara pada bulan November. Sehingga nantinya tidak mengganggu pada saat proses perhitungan klasifikasi. Dari proses pembersihan didapatkan data sebesar 304 data dan atribut menjadi tujuh yaitu Lokasi ISPU, pm10, SO2, CO, SO3, NO2 dan Kategori yang akan digunakan dalam proses klasifikasi. Pada Tabel 1 adalah data atribut yang sudah melalui pembersihan data sehingga menjadi tujuh atribut.

Tabel 1 Data yang akan di proses untuk klasifikasi

NO	LOKASI ISPU	pm10	SO2	CO	O3	NO2	CATEGORY
1	DKI4	76	31	38	35	9	SEDANG
2	DKI3	23	31	24	39	14	BAIK
3	DKI3	53	35	35	101	23	TIDAK SEHAT
4	DKI3	53	49	34	57	15	SEDANG
5	DKI4	44	32	26	28	10	BAIK
6	DKI5	20	30	22	31	5	BAIK
7	DKI5	26	30	14	46	5	BAIK
8	DKI5	62	33	19	73	17	SEDANG
9	DKI3	45	32	16	63	14	SEDANG
10	DKI5	42	34	16	95	17	SEDANG
.							
.							
.							
304	DKI	40	38	18	81	9	SEDANG

#### D. Tahap Implementasi Algoritma K-Nearest Neighbor (KNN)

Sebelum melakukan proses perhitungan algoritma KNN perlu disiapkan data kelas (label), data atribut, data lama sebagai data latih dan data baru sebagai data uji. Kategori pencemaran udara berdasarkan rentang nilai untuk menentukan kualitas pencemaran Udara di Kota Jakarta disajikan pada Tabel 2. Kemudian, memberikan data uji untuk menentukan klasifikasi baru yaitu pm10 = 55, SO2 = 57, CO = 75, O3 = 80, NO2 = 70.

Tabel 2 Kategori pencemaran udara

RENTANG NILAI	KATEGORI
0 – 50	BAIK

RENTANG NILAI	KATEGORI
51 – 100	SEDANG
101 – 199	TIDAK SEHAT
200 – 299	SANGAT TIDAK SEHAT

E. Hasil Klasifikasi

Hasil klasifikasi ini nantinya akan menghasilkan kategori kualitas udara yang baru sesuai data *training* yang ada dan menambahkan data *testing*.

F. Hasil Prediksi

Setelah mengetahui hasil klasifikasi mayoritas dengan menentukan nilai k maka kategori kualitas udara yang baru di kota Jakarta untuk prediksi bulan November akan di diketahui. Selanjutnya dilakukan evaluasi berdasarkan hasil yang di dapat

**IV. HASIL DAN PEMBAHASAN**

Hasil perhitungan manual dilakukan menggunakan Microsoft Excel 2013. Pada perhitungan manual proses dilakukan berdasarkan rumus algoritma KNN dengan menambahkan data uji pada data latih. Kemudian menghitung jarak dan memberikan *ranking* pada hasil perhitungan jarak, sehingga akan dihasilkan kategori baru dengan menentukan nilai k=7. Contoh perhitungan dengan algoritma KNN yang merujuk pada persamaan (1) adalah:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

$$d_1 = \sqrt{(76 - 55)^2 + (31 - 57)^2 + (38 - 75)^2 + (35 - 80)^2 + (9 - 70)^2}$$

$$= \sqrt{(21)^2 + (26)^2 + (37)^2 + (45)^2 + (61)^2}$$

$$= \sqrt{441 + 676 + 1369 + 2025 + 3721}$$

$$= 90.73036978$$

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

$$d_2 = \sqrt{(23 - 55)^2 + (31 - 57)^2 + (23 - 75)^2 + (39 - 80)^2 + (14 - 70)^2}$$

$$= \sqrt{(32)^2 + (26)^2 + (52)^2 + (41)^2 + (56)^2}$$

$$= \sqrt{1024 + 676 + 2704 + 1681 + 3136}$$

$$= 95.48821917$$

(dan seterusnya sampai d ke 304). Kemudian, setelah melakukan perhitungan jarak terhadap data *testing* yang diberikan, berikut pada Tabel 3 adalah hasil perhitungan jarak.

Tabel 3 Hasil perhitungan jarak secara manual

NO	LOKASI ISPU	pm10	SO2	CO	O3	NO2	CATEGORY	DISTANCE
1	DKI4	76	31	38	35	9	SEDANG	90.73036978
2	DKI3	23	31	24	39	14	BAIK	95.48821917
3	DKI3	53	35	35	101	23	TIDAK SEHAT	68.83313156
4	DKI3	53	49	34	57	15	SEDANG	72.82170006
5	DKI4	44	32	26	28	10	BAIK	97.21625379
6	DKI5	20	30	22	31	5	BAIK	106.7192579
7	DKI5	26	30	14	46	5	BAIK	103.3053726
8	DKI5	62	33	19	73	17	SEDANG	81.35723692
9	DKI3	45	32	16	63	14	SEDANG	87.35559513
10	DKI5	42	34	16	95	17	SEDANG	84.92938243
.								
.								
.								
304	DKI5	40	38	18	81	9	SEDANG	80.23091673

Setelah menghitung jarak dengan data *training* dan data *testing* yang diberikan, kemudian memberikan *ranking* pada masing-masing hasil perhitungan jarak sesuai dari data terkecil hingga terbesar. Kemudian kategori baru dapat dihasilkan berdasarkan *ranking*.

Tabel 4 Hasil perhitungan *ranking*

NO	LOKASI ISPU	pm10	SO <sub>2</sub>	CO	O <sub>3</sub>	NO <sub>2</sub>	CATEGORY	DISTANCE	RANGKING
1	DKI4	76	31	38	35	9	SEDANG	90.73036978	116
2	DKI3	23	31	24	39	14	BAIK	95.48821917	152

NO	LOKASI ISPU	pm10	SO <sub>2</sub>	CO	O <sub>3</sub>	NO <sub>2</sub>	CATEGORY	DISTANCE	RANGKING
3	DKI3	53	35	35	101	23	TIDAK SEHAT	68.83313156	16
4	DKI3	53	49	34	57	15	SEDANG	72.82170006	26
5	DKI4	44	32	26	28	10	BAIK	97.21625379	162
6	DKI5	20	30	22	31	5	BAIK	106.7192579	201
7	DKI5	26	30	14	46	5	BAIK	103.3053726	188
8	D	62	33	19	73	17	SEDANG	81.35723692	53
9	DKI3	45	32	16	63	14	SEDANG	87.35559513	93
10	DKI5	42	34	16	95	17	SEDANG	84.92938243	68
.									
.									
.									
304	DKI5	40	38	18	81	19	SEDANG	80.23091673	50

Tabel 5 Hasil Klasifikasi manual dengan nilai k=7

DISTANCE	RANGKING	CATEGORY	KLASIFIKASI
63.85138996	6	SEDANG	YA
56.0178543	3	SEDANG	YA
60.49793385	4	SEDANG	YA
51.16639522	1	SEDANG	YA
64.17943596	7	SEDANG	YA
53.3947563	2	SEDANG	YA
61.98386887	5	SEDANG	YA

Dari hasil perhitungan klasifikasi Tabel 5 berdasarkan data *testing* yang diberikan, maka hasil kategori yang dihasilkan adalah “SEDANG” karena jumlah mayoritas SEDANG = 7.

Hasil klasifikasi yang digunakan merupakan Data ISPU di Kota Jakarta pada bulan Januari sampai dengan Oktober dengan jumlah data sebanyak 304 dan 10 atribut. Proses klasifikasi dilakukan dengan menentukan k=7. Kemudian dihitung jarak dan menentukan *rangking* berdasarkan data uji yang sudah ditentukan. Setelah menghitung jarak dan *rangking* maka dapat menentukan klasifikasi baru sesuai standart kategori yang sudah di tentukan oleh Dinas Lingkungan Hidup. Hasil dari klasifikasi baru dapat ditentukan dengan kategori mayoritas. Setelah menghasilkan klasifikasi baru kemudian dapat menghitung nilai akurasi yaitu sebesar 95.78 %. Pada perhitungan klasifikasi data ISPU menggunakan algoritma KNN terdapat perbedaan hasil yaitu pada saat menentukan prediksi dengan nilai K yang berbeda. Prediksi yang dibuat untuk menentukan nilai K sebaiknya menggunakan jumlah nilai K yang ganjil agar tidak terjadi kesamaan hasil kategori.

## V. KESIMPULAN

Berdasarkan hasil perhitungan yang telah dilakukan dengan teknik klasifikasi *data mining* menggunakan algoritma KNN untuk menentukan data ISPU pada kota Jakarta. Maka, dapat ditarik kesimpulan bahwa data latih sebanyak 304 data dan satu data uji dapat menghasilkan kategori klasifikasi baru untuk memprediksi kualitas udara pada bulan berikutnya. Selanjutnya hasil kategori mayoritas k=7 yang ada menunjukkan bahwa kualitas udara di bulan berikutnya adalah “SEDANG” dan hasil penelitian ini menghasilkan nilai akurasi sebesar 95.78 %.

Hasil klasifikasi pencemaran udara di Kota Jakarta, diharapkan dapat membantu Dinas Lingkungan Hidup Kota Jakarta untuk memberikan informasi yang lebih memadai kepada masyarakat. Sehingga masyarakat dapat melakukan tindakan pencegahan terhadap dampak buruk pencemaran udara ketika beraktifitas diluar ruangan. sebagai saran pada penelitian selanjutnya, sebaiknya dikembangkan aplikasi yang mampu memberikan informasi secara langsung kepada Dinas Lingkungan Hidup di Kota Jakarta.

## PENAKUAN

Naskah ilmiah ini adalah sebagian penelitian Tugas Akhir milik Siti Nurjanah dengan judul Klasifikasi pencemaran udara dengan algoritma K-Nearest Neighbor (KNN) di Kota Jakarta yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulisty Kusumaningrum.

**DAFTAR PUSTAKA**

- [1] Satra, R. and Rachman, A. (2016) 'Pengembangan Sistem Monitoring Pencemaran Udara Berbasis Protokol Zigbee Dengan Sensor Co', 8(April), pp. 17–22.
- [2] Menteri Kesehatan Republik Indonesia (2002) 'Keputusan Menteri Kesehatan Republik Indonesia', Kementerian Kesehatan Republik Indonesia NOMOR 1407, pp. 1–13. doi: 10.1037/teo0000081.
- [3] Green Peace. (2017). Kualitas Udara yang Buruk di Jabodetabek: Dampak Kesehatan dan Pentingnya Pemantauan Kualitas Udara yang Memadai.
- [4] Website Kementerian Lingkungan hidup,2018[online] (<https://lingkunganhidup.jakarta.go.id/>)
- [5] Dzikrulloh, N. N. and Setiawan, B. D. (2017) 'Penerapan Metode K – Nearest Neighbor ( KNN ) dan Metode Weighted Product ( WP ) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi ( Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri )', Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(5), pp. 378–385.
- [6] Aditya, M ., Hidayat N., R.,and Supianto,A,A. (2014) 'Komparasi Metode *Data mining* K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih,"
- [7] Mustakim and Oktaviani F, G. (2016) 'Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa', 13(2), pp. 195–202.
- [8] A.Siregar and A.Paspabhuana,Pengolahan Data menjadi Informasi dengan Rapidminer,Surakarta,2017
- [9] Aribowo, A. S. (2013) 'Metode *Data mining* Untuk Klasifikasi Kesetiaan Pelanggan Terhadap Merek Produk', pp. 2–4. salah
- [10] Wafiyah, F., Hidayat, N. and Perdana, R. S. (2017) 'Implementasi Algoritma Modified K-Nearest Neighbor ( MKNN ) untuk Klasifikasi Penyakit Demam', Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(10), pp. 1210–1219.
- [11] Wafiyah, F., Hidayat, N. and Perdana, R. S. (2017) 'Implementasi Algoritma Modified K-Nearest Neighbor ( MKNN ) untuk Klasifikasi Penyakit Demam', Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(10), pp. 1210–1219.
- [12] M. S. Mustafa and I. W. Simpen, 2014 "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik *Data mining* ( Studi Kasus : Data Akademik Mahasiswa STMIK Dipanegara Makassar )," Citec J., vol. Vol. 1, No, pp. 270–281,.
- [13] *Data mining Tools Rapidminer [online]* (<https://www.softovator.com/eksplorasi-data-mining-menggunakan-rapidminer/>)