

Analisis Pengelompokan Komoditas Impor Indonesia dengan Teknik *K-Means Clustering*

Nur Anisah¹, Harmayani²

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Asahan
e-mail: anisahnur166@gmail.com¹, harmayani@una.ac.id²

(Received: Mei 2025, Revised: Agustus 2025, Accepted: Oktober 2025)

Abstract—The growth of Indonesia's import volume each year highlights the importance of understanding import segmentation patterns to support more efficient economic strategies. However, the large and varied nature of import data makes manual analysis challenging. This study aims to apply the *K-Means Clustering* method to segment Indonesia's import data from 2021 to 2023 based on variables such as country of origin, type of goods, and import quantity. The research utilizes RapidMiner software to cluster the data into several groups that represent specific import patterns. The clustering results reveal the division of Indonesian imports into three main segments, providing valuable insights for both policymakers and economic actors in developing appropriate import management strategies.

Keywords: Data Mining, Indonesian Imports, Clustering, *K-Means*, RapidMiner.

Intisari—Pertumbuhan volume impor Indonesia setiap tahun menunjukkan pentingnya pemahaman terhadap pola-pola segmentasi impor guna mendukung strategi ekonomi yang lebih efisien. Namun, data impor yang besar dan bervariasi menyulitkan analisis manual. Penelitian ini bertujuan untuk menerapkan metode *K-Means Clustering* dalam segmentasi data impor Indonesia periode 2021–2023 berdasarkan beberapa variabel seperti negara asal, jenis barang, dan jumlah impor. Penelitian dilakukan menggunakan perangkat lunak RapidMiner untuk mengelompokkan data ke dalam beberapa kluster yang merepresentasikan pola impor tertentu. Hasil clustering menunjukkan pembagian impor Indonesia dalam tiga segmen utama, yang dapat memberikan wawasan bagi pemerintah maupun pelaku ekonomi untuk menyusun kebijakan dan strategi yang tepat dalam mengelola impor nasional.

Kata Kunci: Data Mining, Impor Indonesia, Clustering, *K-Means*, RapidMiner.

I. PENDAHULUAN

Dilihat dari aspek impor dan pendapatan, jumlah konsumsi yang sebanding dengan pendapatan menunjukkan bahwa penduduk suatu negara sangat bergantung pada barang dan jasa yang mereka butuhkan karena mereka tidak mampu menghasilkan dan menyediakan kebutuhan mereka sendiri. Nilai impor menunjukkan kemampuan masyarakat atau negara untuk mendatangkan barang dan jasa (Tri Puspendari et al., 2022).

Setiap tahun, Badan Pusat Statistik mencatat volume impor Indonesia yang sangat besar dan beragam. Data yang kompleks ini sebenarnya menyimpan banyak informasi penting yang bisa membantu pemerintah merumuskan strategi ekonomi yang lebih efisien dan tepat sasaran. Sayangnya, karena ukurannya yang sangat besar dan

ragamnya yang tinggi, data impor ini sulit dianalisis secara manual. Hal ini membuat kita kesulitan memahami pola atau tren yang sebenarnya. Akibatnya, pemerintah bisa mengalami kendala dalam menentukan komoditas atau negara mitra impor mana yang harus diprioritaskan, mengenali potensi ketergantungan pada negara tertentu, atau merancang strategi substitusi impor yang lebih baik.

Jika tidak ditangani dengan baik, masalah ini bisa berdampak pada kebijakan ekonomi yang kurang efektif dan mengurangi ketahanan ekonomi nasional. Untuk itu, diperlukan pendekatan berbasis data yang mampu mengelompokkan data impor dengan lebih terstruktur dan otomatis. Salah satu solusi yang bisa diterapkan adalah penggunaan metode data mining, khususnya teknik *clustering*. Dengan pendekatan ini, kita bisa menemukan pola-pola tersembunyi dalam data impor sehingga hasil analisisnya dapat mendukung pengambilan keputusan yang lebih akurat, efisien, dan strategis bagi pembangunan ekonomi Indonesia.

Penelitian sebelumnya oleh Ramadhana (2023), yang berjudul “Penerapan Data Mining Menggunakan Metode *K-means Clustering* Pada Data Ekspor Batubara” dalam penelitian ini peneliti menggunakan algoritma *K-means* untuk menganalisis data ekspor batu bara. Selain itu, Asri Samsiar Ilmananda, (2023) yang berjudul “Klasterisasi Negara Pengekspor Beras ke Indonesia Menggunakan Algoritma *K-means Clustering*” dalam studi peneliti menerapkan metode *k-means* untuk *clustering* negara pengekspor beras.

Penelitian-penelitian ini memberikan kontribusi penting dalam pengelompokan ekspor Indonesia. Namun, meskipun demikian penelitian ini masih terdapat beberapa kekurangan yaitu hanya fokus dalam ekspor saja dan cuman 1 jenis ekspor yaitu batubara dan beras.

Berdasarkan penelitian sebelumnya peneliti menyimpulkan dengan metode *k-means* dapat menyelesaikan penelitian yang berjudul “**Penerapan Data Mining untuk Clustering Segmentasi Impor Indonesia Menggunakan K-means**”.

Penelitian ini bertujuan untuk mendapatkan model dengan tujuan mengelompokkan *import* di Indonesia dan penelitian ini dapat memberikan kontribusi dalam analisa *import* di Indonesia dengan memanfaatkan sistem untuk menerapkan metode *k-means*.

II. TINJAUAN PUSTAKA

Data mining adalah proses mengekstraksi informasi berharga dari kumpulan data yang besar melalui serangkaian tahapan analisis kompleks. Menurut Rachmawaty & Febriani (2022), data mining merupakan tahapan analisis yang bertujuan menggali sejumlah besar data untuk menemukan informasi yang berharga, konsisten, dan tersembunyi. Proses ini melibatkan berbagai teknik seperti kecerdasan buatan, statistik, matematika, dan pembelajaran mesin untuk mengamati sekaligus mengekstrak informasi berguna dari data dalam jumlah besar (Sudarsono et al., 2021).

Clustering adalah metode dalam machine learning yang berfungsi mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan karakteristiknya. Menurut Nahjan et al. (2023), tujuan utama clustering adalah mengidentifikasi pola atau struktur tersembunyi dalam data yang sebelumnya tidak diketahui. Teknik ini sangat efektif untuk menemukan kelompok-kelompok yang belum diketahui sebelumnya dalam data (Prastiwi et al., 2022).

Algoritma K-Means merupakan salah satu metode clustering yang paling umum digunakan karena kesederhanaan dan kecepatannya. Rahmadayanti et al. (2023) menjelaskan bahwa K-Means bekerja dengan meminimalkan fungsi objektif selama proses clustering, yaitu dengan mengurangi variasi di dalam setiap kelompok sekaligus meningkatkan perbedaan antar kelompok. Algoritma ini termasuk dalam kategori unsupervised learning yang tidak memerlukan tahap pelatihan sebelumnya (Nugroho et al., 2022).

Impor merupakan kegiatan membawa masuk barang atau komoditas dari luar negeri ke dalam negeri sebagai bentuk transaksi pembelian (Jalunggono, 2020). Aktivitas impor berperan penting dalam perdagangan internasional karena dapat menciptakan kemakmuran dan pemerataan terhadap permintaan serta penawaran barang dan jasa (Dewi & Nawawi, 2022). Namun, impor yang berlebihan dapat mengurangi cadangan devisa negara akibat kewajiban pembayaran ke luar negeri.

RapidMiner menawarkan antarmuka drag-and-drop yang memungkinkan pengguna membangun alur kerja untuk memproses dan menganalisis data tanpa memerlukan pengalaman pemrograman yang mendalam (Nahjan et al., 2023).

III. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis, dimulai dari pengumpulan data yang bersumber dari kajian literatur dan observasi tidak langsung terkait jumlah impor berdasarkan negara asal. Selanjutnya, peneliti menentukan variabel yang digunakan dalam proses *clustering*, yaitu tahun dan jenis impor, sehingga data dapat dikelompokkan secara lebih terarah. Pengolahan data dilakukan menggunakan algoritma *K-means* dengan langkah-langkah utama, yakni menetapkan jumlah *cluster* ($K=3$), memilih titik acak sebagai centroid awal, mengelompokkan data sesuai centroid terdekat menggunakan Euclidean distance, hingga proses iterasi berhenti ketika tidak terjadi perpindahan centroid lagi. Implementasi dan pengujian model dilakukan dengan bantuan software RapidMiner untuk memastikan metode yang diterapkan dapat memberikan hasil yang bermanfaat dalam menentukan *clustering* data impor di Indonesia. Analisis data dalam penelitian ini menggunakan metode deskriptif yang bertujuan menggambarkan keadaan objek secara jelas, dengan objek penelitian berupa data impor di Indonesia yang dijadikan fokus utama kajian. Untuk mendukung pelaksanaan penelitian, diperlukan perangkat keras berupa PC dengan spesifikasi Acer Aspire 3, prosesor AMD A4, memori 8 GB DDR4 L, dan penyimpanan 500 GB HDD, serta perangkat lunak seperti aplikasi RapidMiner dan sistem operasi Windows 10.

IV. HASIL DAN PEMBAHASAN

Hasil Penelitian

Analisa Masalah

Permasalahan utama dalam penelitian ini adalah bagaimana mengelompokkan data impor di Indonesia berdasarkan negara asalnya. Dikumpulkan dari sumber resmi yaitu Badan Pusat Statistik (BPS), melalui situs resmi mereka di (<https://www.bps.go.id/id>). Data yang digunakan terdiri dari 102 data impor berdasarkan negara asal mulai dari tahun 2021 hingga 2023. Tantangan utama adalah menentukan jumlah impor yang banyak sedang dan sedikit. Untuk itu, metode *K-Means Clustering* digunakan untuk membagi data menjadi 3 *cluster* yang berbeda, berdasarkan jumlah impor pada tiap negara. Dengan metode ini, data dapat dikelompokkan secara otomatis, sehingga jumlah impor yang sesuai ketentuan masuk ke dalam satu *cluster* yang sama. Penggunaan 10 sampel dari dataset membantu untuk pengujian metode ini dan memberikan hasil pengelompokan yang diharapkan mampu mengidentifikasi jumlah impor pada Indonesia.

Analisa Kebutuhan Sistem

Analisa kebutuhan dilakukan untuk mengidentifikasi kebutuhan-kebutuhan dan komponen-komponen yang akan digunakan pada aplikasi rapiminer untuk menganalisis data impor menggunakan data mining metode *K-Means*.

Analisa Metode K-means

K-Means clustering merupakan algoritma efektif untuk menentukan *cluster* dalam sekumpulan data, dimana pada algoritma tersebut dilakukan analisis kelompok yang mengacu pada pemartisan N objek ke dalam K kelompok (*cluster*) berdasarkan nilai rata-rata (*means*) terdekat, adapun tahapan algoritma ini adalah sebagai berikut:

Pertama, tentukan berapa banyak jumlah k (*cluster*), kedua, secara acak tentukan record yang menjadi lokasi pusat *cluster*. Ketiga, tentukan pusat *cluster* terdekat untuk setiap *record*. Adapun persamaan yang sering digunakan dalam pemecahan masalah dalam menentukan jarak terdekat adalah persamaan *Euclidean* berikut:

Dalam pengujian ini di berikan data impor Indonesia dari BPS untuk sampel sebagai berikut

Tabel 1. Jumlah Impor Indonesia

Jenis Impor	Negara Asal	2021	2022	2023
Biji Gandum dan Meslin	Australia	4626,4	4192	4239,6
Biji Gandum dan Meslin	Ukraina	2833,8	166,8	652,4
Biji Gandum dan Meslin	Kanada	1919,1	1322,4	2378
Biji Gandum dan Meslin	Argentina	606,8	1469,7	198,8
Biji Gandum dan Meslin	Amerika Serikat	447,9	392,4	390,9
Biji Gandum dan Meslin	India	318,5	908,1	0
Biji Gandum dan Meslin	Bulgaria	227,6	167,5	867,9
Biji Gandum dan Meslin	Brasil	123	641,6	828,1
Biji Gandum dan Meslin	Rep. Moldova	65,9	30,4	61,9
Biji Gandum dan Meslin	Fed. Rusia	3	0	909,4
Biji Gandum dan Meslin	Lainnya	0	59,5	59,6
Impor Sayuran	Tiongkok	645580,3	622788,5	641132,9
Impor Sayuran	Myanmar	90962,5	63748,4	13135,9
Impor Sayuran	Selandia Baru	50133,3	34356,8	32884,5
Impor Sayuran	India	55643,5	131422,6	128965,9
Impor Sayuran	Australia	26551,1	26030,2	25099,7
Impor Sayuran	Ethiopia	12933,9	9600	12765,9
Impor Sayuran	Amerika Serikat	24559,2	3780,1	4235,4
Impor Sayuran	Belanda	17207,4	35014,4	20247,3
Impor Sayuran	Jerman	12507,9	21974,3	17208,4
Impor Sayuran	Kanada	9479,8	9200,3	17391,4

Impor Sayuran	Lainnya	23944,2	43535,2	87422,9
Impor Garam	Australia	2108345	1998382,5	2158030,68
Impor Garam	India	715506	751398	641037,01
Impor Garam	Selandia Baru	3487,7	4382,1	5138,37
Impor Garam	Tiongkok	2470,1	1377	1506,05
Impor Garam	Denmark	448,5	194,9	484,33
Impor Garam	Jerman	201,8	286	304,02
Impor Garam	Thailand	375	425	1028,06
Impor Garam	Lainnya	247,5	180,5	328,74
Impor Daging Sejenis Lembu	India	84219,3	77515,6	104204,1
Impor Daging Sejenis Lembu	Australia	84954,8	105756,3	112601
Impor Daging Sejenis Lembu	Brasil	12903,9	13693,1	12303,2
Impor Daging Sejenis Lembu	Amerika Serikat	9945,6	7414,8	6361,8
Impor Daging Sejenis Lembu	Selandia Baru	3355,7	1114,5	525,7
Impor Daging Sejenis Lembu	Spanyol	131,6	35,6	48,3
Impor Daging Sejenis Lembu	Lainnya	15918,7	20120,2	2389,5
Impor Beras	India	215386,46	178533,57	69715,7
Impor Beras	Thailand	69360,04	80182,51	1381921,2
Impor Beras	Vietnam	65692,87	81828,04	1147705,3
Impor Beras	Pakistan	52479,01	84407	309309,7
Impor Beras	Myanmar	3790	3830	141204
Impor Beras	Jepang	230,29	56,09	61,5
Impor Beras	Tiongkok	42,6	6	7
Impor Beras	Lainnya	760,15	364,07	12933,3
Impor Pupuk	Kanada	1804,2	1766,5	625,2
Impor Pupuk	Tiongkok	1784,3	1066,2	1147
Impor Pupuk	Rusia	974,6	1030	1078,7

Impor Pupuk	Belarusia	905,5	133,4	144,8
Impor Pupuk	Vietnam	182,4	126,8	146,8
Impor Pupuk	Mesir	1141,2	801,4	486,6
Impor Pupuk	Yordania	243,7	562,7	360,6
Impor Pupuk	Jerman	151,5	53,8	50,5
Impor Pupuk	Norwegia	73,5	37,2	106,4
Impor Pupuk	Australia	286,6	213,3	368,8
Impor Pupuk	Malaysia	83,8	72,4	46,8
Impor Pupuk	Laos	96,4	223,6	193,4
Impor Pupuk	Taiwan	43	15,5	45,1
Impor Pupuk	Lainnya	349,4	309,5	567,2
Impor Semen	Tiongkok	2911,9	1880,8	2960,3
Impor Semen	Korea Selatan	641,6	656,8	695,2
Impor Semen	Inggris	693,6	743,2	501,5
Impor Semen	Perancis	742,2	0	329,3
Impor Semen	Belanda	523,3	294,8	333,2
Impor Semen	Kroasia	223,3	310,2	310,1
Impor Semen	Singapura	469,9	517,1	225,6
Impor Semen	Jepang	32,7	596,9	420,4
Impor Semen	Lainnya	27,2	50,4	26,5
Impor Besi dan Baja	Tiongkok	2518,3	2711,4	3711,1
Impor Besi dan Baja	Jepang	2383,6	2397,9	2398,6
Impor Besi dan Baja	India	1438	514,3	400,9
Impor Besi dan Baja	Korea Selatan	936	933,9	906,4
Impor Besi dan Baja	Vietnam	552,1	584,5	687,8
Impor Besi dan Baja	Singapura	315,3	392,3	315,7
Impor Besi dan Baja	Thailand	206,4	203,3	103,9
Impor Besi dan Baja	Amerika Serikat	71,7	11,1	10,5
Impor Besi dan Baja	Jerman	9	4	2,9

Impor Besi dan Baja	Italia	3,4	5,2	1,8
Impor Besi dan Baja	Lainnya	4604	6331,6	5273,5
Impor Aluminium	Tiongkok	210937,7	268900,1	271456,8
Impor Aluminium	Malaysia	55273,8	39954	61239,4
Impor Aluminium	Australia	56851,4	56368,2	62216,4
Impor Aluminium	Amerika Serikat	80019,9	68310,4	43203,5
Impor Aluminium	Korea Selatan	37916,3	33563,3	22420,4
Impor Aluminium	Uni Emirat Arab	44933,6	50581	45904,3
Impor Aluminium	Thailand	21429,5	15279,3	21628,3
Impor Aluminium	Singapura	19247,4	20159	13233,3
Impor Aluminium	India	23173	26721,2	24197
Impor Aluminium	Jepang	5793,4	5741,7	4704,8
Impor Aluminium	Lainnya	167135,9	139168,1	137795,3
Impor Tembaga	Tiongkok	55399,48	49447,3	72775,8
Impor Tembaga	Jepang	41084,88	39903,7	52908,6
Impor Tembaga	Korea Selatan	12297,8	12938,8	5976,5
Impor Tembaga	Thailand	11571,14	9030,6	5404,5
Impor Tembaga	Malaysia	10201,5	10568,8	6720,3
Impor Tembaga	Filipina	7176,57	13279,4	15225,5
Impor Tembaga	Taiwan	7660,31	6834,9	6288,8
Impor Tembaga	Singapura	6528,88	8214	8032,2
Impor Tembaga	Cile	5570,08	11310,5	9024,1
Impor Tembaga	Australia	2651,7	2539,5	9483
Impor Tembaga	Lainnya	15882,53	10899,4	10812,5

Dalam proses ini diambil 10 sampel data khusus untuk proses pengujian *clustering* data yaitu data yang di ambil tahun 2021,2022,2023. Kemudian ke 10 data sampel ini

diubah menjadi data berikut dalam tabel di bawah ini untuk proses *clustering*.

Tabel 2. Sampel Data

Sampel data	X1	X2	X3
1	4626,4	4192	4239,6
2	2833,8	166,8	652,4
3	1919,1	1322,4	2378
4	606,8	1469,7	198,8
5	447,9	392,4	390,9
6	318,5	908,1	0
7	227,6	167,5	867,9
8	123	641,6	828,1
9	65,9	30,4	61,9
10	3	0	909,4

Keterangan:

1. Sampel data 1-10 adalah urutan negaras asal
2. X1 adalah tahun 2021
3. X2 adalah tahun 2022
4. X3 adalah tahun 2023

Tahap 1: Normalisasi Data

$$Normalisasi = \frac{x - \min}{\max - \min}$$

1. Australisa

$$x1 = \frac{4,626.4 - 3.0}{4,263.4} = \frac{4,263.4}{4,263.4} = 1$$

$$x2 = \frac{4,192.0 - 30.4}{4,161.6} = \frac{4,161.6}{4,161.6} = 1$$

$$x3 = \frac{4,239.6 - 61.9}{4,177.7} = \frac{4,177.7}{4,177.7} = 1$$

2. Ukraina

$$x1 = \frac{2,833.8 - 3.0}{4,623.4} = \frac{2,830.8}{4,623.4} = 0.612$$

$$x2 = \frac{166.8 - 30.4}{4,161.6} = \frac{136.4}{4,161.6} = 0.033$$

$$x3 = \frac{652.4 - 61.9}{4,177.7} = \frac{590.5}{4,177.7} = 0.141$$

3. Kanada

$$x1 = \frac{1,919.1 - 3.0}{4,623.4} = \frac{1,916.1}{4,623.4} = 0.414$$

$$x2 = \frac{1,322.4 - 30.4}{4,161.6} = \frac{1,292.0}{4,161.6} = 0.311$$

$$x3 = \frac{2,378.0 - 61.9}{4,177.7} = \frac{2,316.1}{4,177.7} = 0.555$$

4. Argentina

$$x1 = \frac{606.8 - 3.0}{4,623.4} = \frac{603.8}{4,623.4} = 0.131$$

$$x2 = \frac{1,469.7 - 30.4}{4,161.6} = \frac{1,439.3}{4,161.6} = 0.346$$

$$x3 = \frac{198.8 - 61.9}{4,177.7} = \frac{136.9}{4,177.7} = 0.033$$

5. Amerika Serikat

$$x1 = \frac{447.9 - 3.0}{4,623.4} = \frac{444.9}{4,623.4} = 0.096$$

$$x2 = \frac{392.4 - 30.4}{4,161.6} = \frac{362.0}{4,161.6} = 0.087$$

$$x3 = \frac{390.9 - 61.9}{4,177.7} = \frac{329.0}{4,177.7} = 0.079$$

6. India

$$x1 = \frac{318.5 - 3.0}{4,623.4} = \frac{315.5}{4,623.4} = 0.068$$

$$x2 = \frac{908.1 - 30.4}{4,161.6} = \frac{877.7}{4,161.6} = 0.211$$

$$x3 = \frac{613.3 - 61.9}{4,177.7} = \frac{551.4}{4,177.7} = 0.132$$

7. Bulgaria

$$x1 = \frac{227.6 - 3.0}{4,623.4} = \frac{224.6}{4,623.4} = 0.049$$

$$x2 = \frac{167.5 - 30.4}{4,161.6} = \frac{137.1}{4,161.6} = 0.033$$

$$x3 = \frac{867.9 - 61.9}{4,177.7} = \frac{806.0}{4,177.7} = 0.193$$

8. Brasil

$$x1 = \frac{123.0 - 3.0}{4,623.4} = \frac{120.0}{4,623.4} = 0.026$$

$$x2 = \frac{641.6 - 30.4}{4,161.6} = \frac{611.2}{4,161.6} = 0.147$$

$$x3 = \frac{828.1 - 61.9}{4,177.7} = \frac{766.2}{4,177.7} = 0.184$$

9. Rep. Moldova

$$x1 = \frac{65.9 - 3.0}{4,623.4} = \frac{62.9}{4,623.4} = 0.014$$

$$x2 = \frac{30.4 - 30.4}{4,161.6} = \frac{0}{4,161.6} = 0$$

$$x3 = \frac{61.9 - 61.9}{4,177.7} = \frac{0}{4,177.7} = 0$$

10. Fed. Rusia

$$x1 = \frac{3.0 - 3.0}{4,623.4} = \frac{0}{4,623.4} = 0$$

$$x2 = \frac{456.2 - 30.4}{4,161.6} = \frac{425.8}{4,161.6} = 0.102$$

$$x3 = \frac{909.4 - 61.9}{4,177.7} = \frac{847.5}{4,177.7} = 0.202$$

Tahap 2: Inisialisasi Centroid (K=3)

Kita pilih 3 *cluster* secara random:

Centroid Awal:

$$C1 = (0.4, 0.3, 0.4)$$

$$C2 = (0.7, 0.6, 0.2)$$

$$C3 = (0.1, 0.1, 0.8)$$

Tahap 3: Iterasi 1

Langkah 3.1: Hitung Jarak Euclidean (10 negara)

$$\text{Formula: } d = \sqrt{[(x1-c1)^2 + (x2-c2)^2 + (x3-c3)^2]}$$

Untuk Australia (1.000, 1.000, 1.000):

$$\text{Jarak ke C1: } \sqrt{[(1-0.4)^2 + (1-0.3)^2 + (1-0.4)^2]} = \sqrt{[0.36 + 0.49 + 0.36]} = 1.135$$

$$\text{Jarak ke C2: } \sqrt{[(1-0.7)^2 + (1-0.6)^2 + (1-0.2)^2]} = \sqrt{[0.09 + 0.16 + 0.64]} = 0.943$$

$$\text{Jarak ke C3: } \sqrt{[(1-0.1)^2 + (1-0.1)^2 + (1-0.8)^2]} = \sqrt{[0.81 + 0.81 + 0.04]} = 1.291$$

Australia → *Cluster 2* (jarak terkecil: 0.943)

Untuk Ukraina (0.612, 0.033, 0.141):

$$\text{Jarak ke C1: } \sqrt{[(0.612-0.4)^2 + (0.033-0.3)^2 + (0.141-0.4)^2]} = 0.393$$

$$\text{Jarak ke C2: } \sqrt{[(0.612-0.7)^2 + (0.033-0.6)^2 + (0.141-0.2)^2]} = 0.587$$

$$\text{Jarak ke C3: } \sqrt{[(0.612-0.1)^2 + (0.033-0.1)^2 + (0.141-0.8)^2]} = 0.831$$

Ukraina *Cluster 1* (jarak terkecil: 0.393)

Untuk Kanada (0.414, 0.311, 0.555):

$$\text{Jarak ke C1: } \sqrt{[(0.414-0.4)^2 + (0.311-0.3)^2 + (0.555-0.4)^2]} = 0.158$$

$$\text{Jarak ke C2: } \sqrt{[(0.414-0.7)^2 + (0.311-0.6)^2 + (0.555-0.2)^2]} = 0.533$$

Jarak ke C3: $\sqrt{[(0.414-0.1)^2 + (0.311-0.1)^2 + (0.555-0.8)^2]}$
 = 0.441

Kanada *Cluster 1* (jarak terkecil: 0.158)

Untuk Argentina (0.131, 0.346, 0.033):

Jarak ke C1: $\sqrt{[(0.131-0.4)^2 + (0.346-0.3)^2 + (0.033-0.4)^2]}$
 = 0.490

Jarak ke C2: $\sqrt{[(0.131-0.7)^2 + (0.346-0.6)^2 + (0.033-0.2)^2]}$
 = 0.648

Jarak ke C3: $\sqrt{[(0.131-0.1)^2 + (0.346-0.1)^2 + (0.033-0.8)^2]}$
 = 0.798

Argentina *Cluster 1* (jarak terkecil: 0.490)

Untuk Amerika Serikat (0.096, 0.087, 0.079):

Jarak ke C1: $\sqrt{[(0.096-0.4)^2 + (0.087-0.3)^2 + (0.079-0.4)^2]}$
 = 0.516

Jarak ke C2: $\sqrt{[(0.096-0.7)^2 + (0.087-0.6)^2 + (0.079-0.2)^2]}$
 = 0.785

Jarak ke C3: $\sqrt{[(0.096-0.1)^2 + (0.087-0.1)^2 + (0.079-0.8)^2]}$
 = 0.722

Amerika Serikat → *Cluster 1* (jarak terkecil: 0.516)

Untuk India (0.068, 0.211, 0.132):

Jarak ke C1: $\sqrt{[(0.068-0.4)^2 + (0.211-0.3)^2 + (0.132-0.4)^2]}$
 = 0.469

Jarak ke C2: $\sqrt{[(0.068-0.7)^2 + (0.211-0.6)^2 + (0.132-0.2)^2]}$
 = 0.759

Jarak ke C3: $\sqrt{[(0.068-0.1)^2 + (0.211-0.1)^2 + (0.132-0.8)^2]}$
 = 0.698

India *Cluster 1* (jarak terkecil: 0.469)

Untuk Bulgaria (0.049, 0.033, 0.193):

Jarak ke C1: $\sqrt{[(0.049-0.4)^2 + (0.033-0.3)^2 + (0.193-0.4)^2]}$
 = 0.511

Jarak ke C2: $\sqrt{[(0.049-0.7)^2 + (0.033-0.6)^2 + (0.193-0.2)^2]}$
 = 0.891

Jarak ke C3: $\sqrt{[(0.049-0.1)^2 + (0.033-0.1)^2 + (0.193-0.8)^2]}$
 = 0.615

Bulgaria *Cluster 1* (jarak terkecil: 0.511)

Untuk Brasil (0.026, 0.147, 0.184):

Jarak ke C1: $\sqrt{[(0.026-0.4)^2 + (0.147-0.3)^2 + (0.184-0.4)^2]}$
 = 0.507

Jarak ke C2: $\sqrt{[(0.026-0.7)^2 + (0.147-0.6)^2 + (0.184-0.2)^2]}$
 = 0.825

Jarak ke C3: $\sqrt{[(0.026-0.1)^2 + (0.147-0.1)^2 + (0.184-0.8)^2]}$
 = 0.628

Analisis Pengelompokan Komoditas Impor Indonesia dengan Teknik *K-Means Clustering*

Brasil *Cluster 1* (jarak terkecil: 0.507)

Untuk Rep. Moldova (0.014, 0.000, 0.000):

Jarak ke C1: $\sqrt{[(0.014-0.4)^2 + (0.000-0.3)^2 + (0.000-0.4)^2]}$
 = 0.671

Jarak ke C2: $\sqrt{[(0.014-0.7)^2 + (0.000-0.6)^2 + (0.000-0.2)^2]}$
 = 0.942

Jarak ke C3: $\sqrt{[(0.014-0.1)^2 + (0.000-0.1)^2 + (0.000-0.8)^2]}$
 = 0.806

Rep. Moldova *Cluster 1* (jarak terkecil: 0.671)

Untuk Fed. Rusia (0.000, 0.102, 0.202):

Jarak ke C1: $\sqrt{[(0.000-0.4)^2 + (0.102-0.3)^2 + (0.202-0.4)^2]}$
 = 0.539

Jarak ke C2: $\sqrt{[(0.000-0.7)^2 + (0.102-0.6)^2 + (0.202-0.2)^2]}$
 = 0.860

Jarak ke C3: $\sqrt{[(0.000-0.1)^2 + (0.102-0.1)^2 + (0.202-0.8)^2]}$
 = 0.598

Fed. Rusia *Cluster 1* (jarak terkecil: 0.539)

Semua data telah kita dapatkan, dan kita masing – masing data sudah dikelompokkan berdasarkan *cluster 1*, *cluster 2* dan *cluster 3*. Selanjutnya kita input pengelompokan pada tabel berikut.

Tabel 3. Pengelompokan Data

No	Sampel Data	2021 (Ton)	2022 (Ton)	2023 (Ton)	Cluster
1	Australia	4626,4	4192	4239,6	1
2	Kanada	1919,1	1322,4	2378	1
3	Ukraina	2833,8	166,8	652,4	2
4	Argentina	606,8	1469,7	198,8	2
5	India	318,5	908,1	613,3	2
6	Brasil	123	641,6	828,1	2
7	Bulgaria	227,6	167,5	867,9	2
8	Fed. Rusia	3	456,2	909,4	2
9	Amerika Serikat	447,9	392,4	390,9	3
10	Rep. Moldova	65,9	30,4	61,9	3

Dari tabel diatas, sebagian besar data impor berada di *cluster 2* (sedang). Hal ini menunjukkan bahwa pada data yang di uji di dominasi oleh tarif sedang.

Sementara itu, terdapat beberapa data yang tergolong dalam *Cluster 1*(tinggi). Ini menunjukkan ada tarif impor yang tinggi dan ada yang masuk ke dalam *Cluster 3*(rendah.

Berdasarkan hasil *clustering*, dapat disimpulkan bahwa impor di indonesia berdasarkan negara asal lebih banya

diisi oleh impor dengan kategori sedang dan dengan beberapa tinggi serta rendah.

K-means adalah algoritma yang lebih mudah digunakan dalam pengelompokan data impor Indonesia berdasarkan negara asal dibandingkan dengan metode lain seperti *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) Karena beberapa alasan:

1. Tidak memerlukan data berlabel

K-means adalah algoritma *unsupervised learning* yang tidak memerlukan data berlabel untuk pengelompokan. Anda dapat langsung mengelompokkan data impor tanpa perlu proses pelabelan yang rumit. Sebaliknya, *Naïve Bayes* dan KNN membutuhkan data yang sudah memiliki label yang jelas, yang bisa memakan waktu dan usaha ekstra untuk memperoleh dan mempersiapkan data.

2. Efisien dalam Pengelolaan Data Besar:

K-means sangat efisien dalam hal komputasi, terutama pada dataset yang besar, karena dapat mengelompokkan data dalam jumlah besar dengan cepat. Sementara KNN membutuhkan banyak memori dan komputasi untuk menghitung jarak ke setiap titik data, yang membuatnya kurang efisien ketika ukuran dataset sangat besar. *Naïve Bayes* lebih efisien dari KNN, namun terbatas pada klasifikasi, dan tidak bisa digunakan untuk pengelompokan data yang tidak memiliki label.

Secara keseluruhan, *K-means* adalah pilihan yang lebih praktis dan mudah digunakan untuk mengelompokkan data cuaca, terutama jika data tidak memiliki label dan anda hanya ingin menemukan pola atau *cluster* berdasarkan kesamaan fitur, *Naïve Bayes* dan KNN lebih kompleks dan memerlukan lebih banyak persiapan, terutama ketika data impor Indonesia yang dianalisis sudah memiliki label atau memiliki ketergantungan antar fitur yang harus diperhitungkan.

Penerapan Hasil Clustering dalam Kebijakan Perdagangan Indonesia

Berdasarkan hasil *clustering* data impor Indonesia menggunakan metode *K-means* yang telah dilakukan, ditemukan tiga kategori utama negara pengimpor berdasarkan volume impor: tinggi (*Cluster 1*), sedang

(*Cluster 2*), dan rendah (*Cluster 3*). Hasil pengelompokan ini memiliki implikasi praktis yang signifikan untuk perumusan kebijakan perdagangan dan strategi ekonomi nasional.

1. Pengembangan Perjanjian Perdagangan Jangka Panjang: Mengingat volume impor yang stabil dan tinggi, Indonesia perlu memperkuat hubungan bilateral dengan negara-negara ini melalui perjanjian perdagangan yang memberikan kepastian pasokan dan stabilitas harga.
2. Negosiasi Tarif Preferensial: Dengan volume impor yang besar, Indonesia memiliki posisi tawar yang kuat untuk menegosiasikan tarif yang lebih menguntungkan, terutama untuk komoditas strategis seperti gandum dan bahan pangan lainnya.
3. Pengembangan Infrastruktur Logistik: Investasi dalam infrastruktur pelabuhan dan fasilitas penyimpanan untuk mengoptimalkan rantai pasokan dari negara-negara dengan volume impor tinggi.
4. Program Stabilisasi Impor: Mengembangkan mekanisme hedging dan kontrak berjangka untuk mengurangi volatilitas impor dari negara-negara ini, terutama mengingat dampak geopolitik seperti yang terjadi pada Ukraina dan Rusia.

Evaluasi Model dengan Confusion Matrix

Untuk mengukur akurasi dan performa model *K-means* yang telah dibuat, dilakukan evaluasi menggunakan confusion matrix. Evaluasi ini dilakukan dengan membandingkan hasil *clustering* dengan pengelompokan manual berdasarkan domain knowledge tentang volume impor.

Berdasarkan analisis domain dan rata-rata volume impor, ditetapkan ground truth sebagai berikut:

1. Tinggi: Volume > 3000 ton (Australia)
2. Sedang: Volume 500-3000 ton (Ukraina, Kanada, Argentina, Brasil, Bulgaria, Fed. Rusia)
3. Rendah: Volume < 500 ton (Amerika Serikat, India, Rep. Moldova)

Tabel 4. Perbandingan Hasil *Clustering* dengan *Ground Truth*

Negara	Prediksi Cluster	Actual Class	Status
--------	------------------	--------------	--------

Australia	Tinggi	Tinggi	✓
Kanada	Sedang	Sedang	✓
Ukraina	Sedang	Sedang	✓
Argentina	Sedang	Sedang	✓
India	Sedang	Rendah	✗
Brasil	Sedang	Sedang	✓
Bulgaria	Sedang	Sedang	✓
Fed. Rusia	Sedang	Sedang	✓
Amerika Serikat	Sedang	Rendah	✗

Tabel 5. Confusion Matrix

	Prediksi Tinggi	Prediksi Sedang	Prediksi Rendah	Total
Actual Tinggi	1	0	0	1
Actual Sedang	0	6	0	6
Actual Rendah	0	3	3	6
Total	1	9	3	10

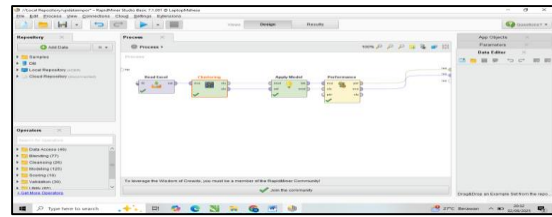
Perhitungan Metrik Evaluasi:

1. Akurasi Akurasi = $(TP + TN) / \text{Total Data}$ Akurasi = $(1 + 6 + 0) / 10 = 7/10 = 70\%$
2. Precision untuk setiap kelas
 Precision Tinggi = $1/1 = 100\%$
 Precision Sedang = $6/9 = 66.7\%$
 Precision Rendah = $0/0 = 0\%$ (tidak ada prediksi rendah)
3. Recall untuk setiap kelas
 Recall Tinggi = $1/1 = 100\%$
 Recall Sedang = $6/6 = 100\%$
 Recall Rendah = $0/3 = 0\%$
4. F1-Score untuk setiap kelas
 F1-Score Tinggi = $2 \times (100\% \times 100\%) / (100\% + 100\%) = 100\%$
 F1-Score Sedang = $2 \times (66.7\% \times 100\%) / (66.7\% + 100\%) = 80\%$
 F1-Score Rendah = $2 \times (0\% \times 0\%) / (0\% + 0\%) = 0\%$

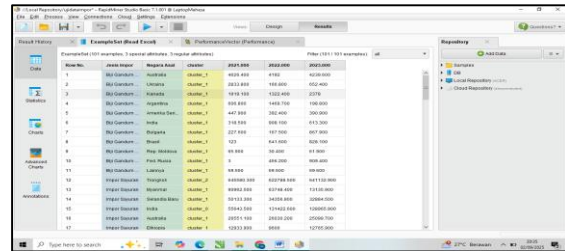
Pengujian dengan RapidMiner

Untuk memastikan keakuratan hasil perhitungan yang dilakukan, pada penelitian ini juga dilakukan proses perhitungan menggunakan software *RapidMiner* dengan data yang sama sebagai pembandingan. Gambar 4.1 menunjukkan proses *clustering* dengan menggunakan algoritma *K-Means* di *RapidMiner*, dengan jumlah centroid

yang digunakan sebanyak 3 (tiga). Sementara itu, Gambar 4.2 menampilkan data impor yang sama, yang juga digunakan pada proses perhitungan manual.

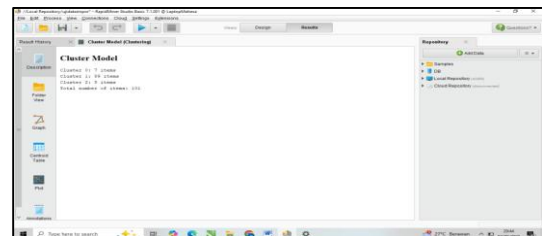


Gambar 1. Proses Clustering pada RapidMiner

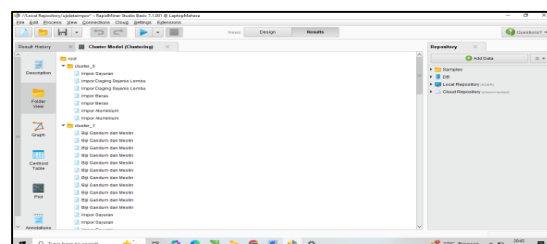
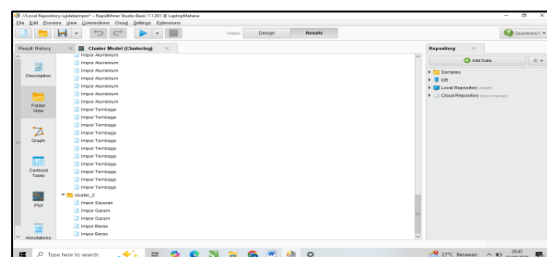


Gambar 2. Data Impor pada RapidMiner

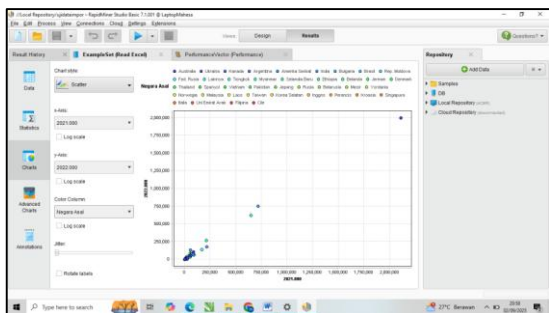
Setelah proses dijalankan, hasil menunjukkan bahwa jumlah *cluster* sebanyak 3 (tiga) diperoleh dengan total data item sebanyak 101 record. Hasil *cluster* model pada *RapidMiner* ditunjukkan pada gambar 4. Sedangkan pada gambar 5 menunjukkan pengelompokan data yang dihasilkan dimana pada *cluster_0*.



Gambar 3. Hasil Cluster Model Pada RapidMiner



Gambar 4. Hasil Pengelompokan Data Pada RapidMiner



Gambar 5. Visualisasi *Cluster Impor* Pada *RapidMiner*
Analisis Hasil

Hasil dari pengelompokan dengan algoritma *K-Means Clustering* berdasarkan data impor dari tahun 2021, 2022, dan 2023 berhenti pada iterasi pertama dan menghasilkan 3 (tiga) *cluster* yang mengelompokkan negara-negara berdasarkan

Cluster 1 (Kategori Tinggi)

Cluster ini terdiri dari 2 negara dengan karakteristik volume impor yang tinggi dan konsisten:

1. Australia: Dengan volume impor 4.626,4 ton (2021), 4.192 ton (2022), dan 4.239,6 ton (2023)
2. Kanada: Dengan volume impor 1.919,1 ton (2021), 1.322,4 ton (2022), dan 2.378 ton (2023)

Kedua negara ini menunjukkan pola impor yang stabil dengan volume yang tinggi sepanjang periode pengamatan. Australia mempertahankan posisi sebagai importir terbesar dengan volume yang konsisten di atas 4.000 ton per tahun.

Cluster 2 (Kategori Sedang)

Cluster ini merupakan kelompok terbesar dengan 6 negara yang memiliki karakteristik volume impor sedang dengan variasi yang cukup signifikan antar tahun:

1. Ukraina: 2.833,8 ton (2021), 166,8 ton (2022), 652,4 ton (2023)
2. Argentina: 606,8 ton (2021), 1.469,7 ton (2022), 198,8 ton (2023)
3. India: 318,5 ton (2021), 908,1 ton (2022), 613,3 ton (2023)
4. Brasil: 123 ton (2021), 641,6 ton (2022), 828,1 ton (2023)
5. Bulgaria: 227,6 ton (2021), 167,5 ton (2022), 867,9 ton (2023)

6. Fed. Rusia: 3 ton (2021), 456,2 ton (2022), 909,4 ton (2023)

Negara-negara dalam *cluster* ini menunjukkan pola fluktuasi yang tinggi, dengan beberapa negara mengalami penurunan drastis pada tahun tertentu (seperti Ukraina di tahun 2022) dan peningkatan signifikan di tahun lain (seperti Fed. Rusia yang meningkat drastis dari 2021 ke 2023).

Cluster 3 (Kategori Rendah)

Cluster ini terdiri dari 2 negara dengan karakteristik volume impor yang rendah dan relatif stabil:

1. Amerika Serikat: 447,9 ton (2021), 392,4 ton (2022), 390,9 ton (2023)
2. Rep. Moldova: 65,9 ton (2021), 30,4 ton (2022), 61,9 ton (2023)

Kedua negara ini menunjukkan volume impor yang konsisten rendah sepanjang periode pengamatan, dengan Amerika Serikat mempertahankan volume sekitar 400 ton per tahun dan Rep. Moldova di bawah 100 ton per tahun.

Evaluasi Performa Model

Berdasarkan hasil evaluasi menggunakan confusion matrix, model *K-means* menunjukkan performa yang cukup baik dengan akurasi 70%. Namun, terdapat beberapa catatan penting:

Kekuatan Model

1. Model mampu mengidentifikasi *cluster* tinggi dengan sempurna (precision dan recall 100%)
2. Model menunjukkan performa yang baik untuk *cluster* sedang (F1-score 80%)
3. Hasil *clustering* konsisten antara perhitungan manual dan RapidMiner

Kelemahan Model

1. Model gagal mengidentifikasi *cluster* rendah (recall 0%)
2. Terdapat misklasifikasi pada negara-negara dengan volume rendah yang masuk ke *cluster* sedang
3. Distribusi data yang tidak seimbang mempengaruhi performa model

Faktor Penyebab Misklasifikasi:

1. Variabilitas data yang tinggi antar tahun pada beberapa negara

2. Pengaruh outlier seperti penurunan drastis impor Ukraina tahun 2022
3. Distribusi data yang tidak seimbang antar *cluster*

V. PENUTUP

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai pengelompokan data impor Indonesia berdasarkan negara asal menggunakan metode *K-means Clustering*, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Metode *K-means Clustering* efektif untuk mengelompokkan volume impor. Algoritma ini berhasil membagi 102 data impor dari berbagai negara ke dalam tiga *cluster* yang jelas, yaitu *Cluster 1* (Volume Tinggi), *Cluster 2* (Volume Sedang), dan *Cluster 3* (Volume Rendah). Proses ini membuktikan bahwa *K-means* merupakan alat yang efisien dan powerful untuk analisis data perdagangan tanpa pengawasan (unsupervised).
2. Sebagian besar negara asal impor Indonesia masuk dalam kategori volume sedang. Hasil *clustering*, baik secara manual maupun yang divalidasi dengan RapidMiner, menunjukkan bahwa mayoritas negara pemasok produk impor Indonesia (seperti Ukraina, Argentina, India, Brasil, Bulgaria, dan Rusia) tergolong dalam *Cluster 2*. Hal ini mengindikasikan bahwa struktur impor Indonesia tidak bergantung secara berlebihan pada satu atau dua negara saja, tetapi telah memiliki diversifikasi yang cukup baik dengan banyak mitra dagang menengah.
3. Teridentifikasi mitra dagang strategis dengan volume tinggi dan rendah. Australia dan Kanada konsisten masuk dalam *Cluster 1* (Tinggi) untuk komoditas Biji Gandum, menandakan mereka adalah pemasok utama yang kritis. Sementara itu, negara seperti Amerika Serikat dan Republik Moldova masuk dalam *Cluster 3* (Rendah) untuk komoditas yang sama, menunjukkan potensi untuk ditingkatkan atau dipertimbangkan kembali dalam strategi procurement.
4. Hasil *clustering* memberikan landasan data yang kuat untuk perumusan kebijakan. Pengelompokan ini bukan hanya bersifat statistik, tetapi memiliki implikasi praktis yang signifikan. Tiap *cluster* memerlukan pendekatan kebijakan yang berbeda, mulai dari penguatan kemitraan jangka panjang dengan negara di *Cluster 1*, diversifikasi dan optimasi dengan negara di *Cluster 2*, hingga eksplorasi potensi dan transfer teknologi dengan negara di *Cluster 3*.

DAFTAR PUSTAKA

- [1] Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus: Makan Barbeque Sepuasnya. *Jurnal Ilmiah Informatika*, 9(02), 75-81.
- [2] Bui, M. A., & Bahtiar, A. (2024). Implementasi Metode Algoritma *K-means Clustering* Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1451-1456.
- [3] Buulolo, E. (2020). *Data mining untuk perguruan tinggi*. Deepublish.
- [4] Erfina, A. (2021). *Buku Ajar Data Mining*. Nusa Putra Press.
- [5] Firdaus, E. A., Maulani, S., & Dharmawan, A. B. (2021). Pengukuran Minat Baca Mahasiswa Dengan Metode *Clustering* Di Perpustakaan Akademi Keperawatan Rs. Dustira Cimahi Menggunakan Data Mining. *Nuansa Informatika*, 15(1), 32-40.
- [6] Latifah, U. W., Bahri, S., & Satriandhini, M. (2024). Implementasi algoritma *K-means clustering* untuk strategi promosi kampus IBISA. *JIKO (Jurnal Informatika dan Komputer)*, 8(2), 292-300.
- [7] Marisa, F., Maukar, A. L., & Akhriza, T. M. (2021). *Data mining konsep dan penerapannya*. Deepublish.
- [8] Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode *Clustering K-means* Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101-104.
- [9] Noviantoro, A., Silviana, A. B., Fitriani, R. R., & Permatasari, H. P. (2022). Rancangan Dan Implementasi Aplikasi Sewa Lapangan Badminton Wilayah Depok Berbasis Web. *Jurnal Teknik Dan Science*, 1(2), 88-103.
- [10] Nugroho, M. R., Hendrawan, I. E., & Purwantoro, P. P. (2022). Penerapan Algoritma *K-means* Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI. *Nuansa Informatika*, 16(1), 125-133.
- [11] Pamungkas, P. A., Indrawati, L. R., & Jalunggono, G. (2020). Analisis Pengaruh Ekspor, Impor, Inflasi, Kurs Rupiah, dan Utang Luar Negeri terhadap Cadangan Devisa Indonesia Tahun 1999 –2018. *DINAMIC: Directory Journal of Economic*, 2(3), 659–674.
<http://jom.untidar.ac.id/index.php/dinamic/article/view/1416>.
- [12] Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US

- Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- [13] Prastiwi, H., Pricilia, J., & Rasywir, E. (2022). Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode *K-means Clustering*. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(1), 141-148.
- [14] Puspandari, T., Priyatno, S. H., Novialumi, A., & Herwanti, L. (2022). Pengaruh Ekspor Dan Impor Terhadap Pertumbuhan Ekonomi Di Indonesia. *JIIP- Jurnal Ilmiah Ilmu Pendidikan*, 5(11), 4968-4971.
- [15] Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., ... & Andiyana, A. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.
- [16] Rachmawaty, D., & Febriani, J. M. A. (2022). Penerapan Metode Klasifikasi Decision Tree Untuk Memprediksi Kelulusan Tepat Waktu. *Journal of Industrial Engineering and Technology*, 2(1), 61–74. <https://doi.org/10.24176/jointtech.v2i1.7432>
- [17] Rahmadayanti, F., Anggraini, I., & Susanti, T. (2023). Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode *K-means*. *J. Inf. Syst. Res*, 4(2), 737-741.
- [18] Silalahi, F. D. (2022). Manajemen Database MySQL (Structured Query Language). *Penerbit Yayasan Prima Agus Teknik*, 1-158.
- [19] Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE- Journal of Business and Audit Information Systems*, 4(1).
- [20] Sulianta, F. (2023). *Basic Data Mining from A to Z*. Feri Sulianta.
- [21] Tuasamu, Z., Lewaru, N. A. I. M., Idris, M. R., Syafaat, A. B. N., Faradilla, F., Fadlan, M., ... & Efendi, R. (2023). Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan DFD Dan Flowchart Pada Bisnis Porobico. *Jurnal Bisnis dan Manajemen (JURBISMAN)*, 1(2), 495-510.
- [22] Dewi, R., & Nawawi, Z. M. (2022). Analisis perkembangan ekspor impor kelompok barang ekonomi di Provinsi Sumatera Utara tahun 2016-2020. *Jurnal Pamator Jurnal Ilmiah Universitas Trunojoyo*, 15(1), 137–145. <https://doi.org/10.21107/pamator.v15i1.14283>
- [23] Londjo, M. F. (2021, December 9). IMPLEMENTASI WHITE BOX TESTING DENGAN TEKNIK BASIS PATH PADA PENGUJIAN FORM LOGIN. Londjo | Jurnal Siliwangi Seri Sains Dan Teknologi. <https://jurnal.unsil.ac.id/index.php/jssainstek/article/view/4086/1929>
- [24] Nahjan, M. R., Heryana, N. N., & Voutama, N. A. (2023). IMPLEMENTASI RAPIDMINER DENGAN METODE *CLUSTERING K-MEANS* UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104. <https://doi.org/10.36040/jati.v7i1.6094>