

**Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika
Menggunakan Metode K-Means Di Institut Asia Malang**

***Determination System for Department Concentration for Informatics Students Using the
K-Means Method at the Institute of Asia Malang***

Puji Subekti¹, Titania Dwi Andini², Mufidatul Islamiyah³

^{1,2,3}Institut Teknologi dan Bisnis Asia, Malang, Kota, Indonesia,

E-mail: ²titania@asia.ac.id

Abstrak

Peminatan pada konsentrasi sebuah jurusan merupakan fokus mahasiswa terhadap suatu bidang studi tertentu sesuai dengan minatnya. Tujuan peminatan yaitu untuk lebih memfokuskan mahasiswa terhadap ilmu yang didapat dari matakuliah sebelumnya, agar dapat lebih terarah. Pada prodi Teknik Informatika di Institut Asia Malang terdapat 3 konsentrasi yaitu : Sistem Cerdas; Multimedia dan Game dan Administrasi Jaringan Komputer. Belum adanya sistem yang membantu mahasiswa memilih konsentrasi jurusan cukup menyulitkan mahasiswa untuk mengetahui kemampuan akademik yang dimiliki. Dengan melihat permasalahan tersebut penelitian ini bertujuan membangun sistem pengambilan keputusan dalam memilih konsentrasi jurusan menggunakan metode K-Means Clustering. Dimana data nilai mahasiswa semester 1 sampai semester 4 digunakan sebagai variabel dalam perhitungan metode k-means clustering yang di implementasikan menggunakan Bahasa pemrograman PHP dan MySQL. Berdasarkan hasil pengujian, diperoleh hasil bahwa cluster 1 terdapat 3 mahasiswa yang masuk di konsentrasi sistem cerdas. Cluster ke 2 terdapat 20 mahasiswa yang masuk di konsentrasi Multimedia dan Game, cluster ke 3 terdapat 56 mahasiswa yang masuk di konsentrasi jaringan.

Kata kunci: Data Mining, Konsentrasi Jurusan, K-Means, Clustering.

Abstract

Specialization in the concentration of a major is a student's focus on a particular field of study according to his interests. The purpose of specialization is to focus students more on the knowledge gained from previous courses so that they can be more focused. In the Informatics Engineering study program at the Institute of Asia Malang, there are 3 concentrations, namely: Intelligent Systems; Multimedia and Games and, Computer Network Administration. The absence of a system that helps students choose the concentration of majors is quite difficult for students to know their academic abilities. By looking at these problems, this study aims to build a decision-making system in choosing the concentration of majors using the K-Means Clustering method. Where the data value of semester 1 to semester 4 students is used as a variable in the calculation of the k-means clustering method which is implemented using the PHP and MySQL programming languages. Based on the test results, it was found that in cluster 1 there were 3 students who entered the intelligent system concentration. In the second cluster, there are 20 students who are in the Multimedia and Game concentration, in the third cluster there are 56 students who are in the network concentration.

Keywords: data mining, the concentration of major, K-Means, clustering.

Naskah diterima 2 Jan. 2022; direvisi 1 Feb. 2022; dipublikasikan 1 Apr. 2022.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



I. PENDAHULUAN

Pemilihan konsentrasi dalam kegiatan akademik mahasiswa memang bukan hal yang mudah karena tergantung pada minat, bagat dan keinginan, oleh karena itu perlu pertimbangan yang matang supaya mahasiswa tidak salah dalam memilih konsentrasi yang diinginkan [1], [2]. Hal ini sering terjadi ketika mahasiswa semester akhir mengerjakan tugas akhir namun tidak sesuai dengan kemampuan bidang yang dimiliki. Pemilihan konsentrasi yang asal-asalan tanpa pertimbangan yang matang, menyebabkan dampak negatif pada mahasiswa, yaitu kesulitan dalam penyerapan materi-materi perkuliahan [3], [4]. Oleh sebab itu perlu metode khusus yang dapat digunakan mahasiswa dalam menentukan konsentrasi mahasiswa. Salah satu metode yang digunakan adalah metode K-Means.

Penentuan konsentrasi mahasiswa tidak mudah [5], [6]. Hal ini disebabkan akan berpengaruh terhadap kegiatan akademik mahasiswa. Tidak jarang mahasiswa yang mengambil konsentrasi karena terpengaruh oleh teman tanpa melihat kemampuannya. Sehingga pada akhirnya tidak jarang yang berpindah konsentrasi jurusan jika tidak sesuai kemampuan. Proses berpindahnya konsentrasi cukup memakan waktu dikarenakan kegiatan

belajar mengajar telah dimulai dan proses batal tambah mata kuliah yang telah dipilih memiliki *range* waktu tidak banyak. Waktu yang tersedia hanya sekitar satu minggu bagi mahasiswa untuk merubah mata kuliah konsentrasi yang telah dipilih untuk diganti dengan mata kuliah konsentrasi yang lain.

Statistika dan *machine learning* merupakan algoritma K-means clustering [7], [8], dimana metode K-means ini merupakan metode dalam fungsi pengelompokan (*clustering*) sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam cluster yang sama [9], [10]. *Clustering* merupakan teknik yang digunakan untuk fungsionalitas *data mining* [11], [12]. Algoritma clustering merupakan pengelompokan data menjadi kelompok data tertentu (*cluster*).

Clustering mempunyai banyak manfaat diantaranya penerapan algoritma K-Means *clustering analysis* pada penyakit menular manusia [13], pengelompokan pemilihan produk dan pelanggan potensial [14], dan juga penerapan penyebaran diare di Kabupaten Langkat. Selain itu clustering ini juga digunakan untuk pengelompokan penjualan produk [15], pemanfaatan fitur analisis data yang menggunakan K-Means cluster dalam point of sales [16]. Banyak peneliti juga menggunakan metode ini untuk pengelompokan penggunaan daya di desa/ kelurahan [17], penerapan pada aplikasi android pada tanaman obat herbal [18], penerapan dalam rekomendasi kerja sampingan [19], penerapan dalam status ekonomi desa [20] dan baru baru ini juga digunakan untuk penerapan dalam clustering data Covid-19. Penelitian ini berbeda dari penelitian pendukung sebelumnya dalam hal manfaat dari pengelompokan jenis konsentrasi. Penelitian ini tidak hanya berfungsi memberikan penggambaran tentang konsentrasi terhadap mahasiswa namun juga memberikan penggambaran penentuan kebijakan bagi kampus. Sehingga dengan adanya pengelompokan data ini pihak kampus juga dapat mengetahui jurusan yang paling banyak peminatnya, dari penelitian ini output 2 yang dihasilkan adalah jenis konsentrasi yang banyak di minati oleh mahasiswa Institut Teknologi dan Bisnis Asia Malang. Tujuan pada penelitian ini adalah menghasilkan sebuah sistem pendukung keputusan dalam menentukan konsentrasi jurusan bagi mahasiswa, sehingga dengan adanya sistem ini dapat membantu mahasiswa Institut Teknologi dan Bisnis Asia Malang dalam menentukan konsentrasi jurusan yang ada di program Teknik Informatika.

II. METODE PENELITIAN

Institut Teknologi dan Bisnis Asia Malang (lebih dikenal Institut Asia Malang) memiliki 5 program studi yang tersebar dalam dua Fakultas, yaitu Fakultas Ekonomi dan Bisnis (FEB) serta Fakultas Teknologi dan Desain (FTD). Program Studi Teknik Informatika merupakan bagian dari FTD, memiliki tiga konsentrasi yaitu : Sistem Cerdas untuk Optimasi Aplikasi Bisnis; Multimedia dan Game dan Sistem Administrasi Jaringan dan Keamanan Komputer dimana pemilihan konsentrasi bisa dilakukan mahasiswa saat melakukan proses perancangan Kartu Rencana Studi (KRS) mulai semester 4.

Pada saat ini untuk menentukan konsentrasi jurusan hanya berdasarkan pada keinginan mahasiswa saja tanpa adanya suatu sistem yang bisa memberikan bahan pertimbangan dalam menentukan konsentrasi jurusan. Akibatnya banyak mahasiswa yang mengalami kesulitan saat mengambil konsentrasi yang mereka pilih. Dengan adanya salah satu teknik data mining clustering K-means yang akan memberikan kelompok-kelompok data yang bisa dijadikan bahan pertimbangan mahasiswa untuk menentukan konsentrasi berdasarkan nilai akademik dari semester satu hingga semester tiga yang berhubungan dengan konsentrasi yang ditawarkan.

Aplikasi ini menggunakan teknik clustering dengan metode K-Means dalam mengolah data, menghasilkan 3 cluster kelompok. Cluster pertama (C1), yaitu tingkat konsentrasi tinggi, cluster kedua (C2), yaitu tingkat konsentrasi sedang, cluster ketiga (C3), yaitu tingkat konsentrasi rendah. Hasil akhir dari aplikasi ini berupa pengelompokan konsentrasi jurusan mahasiswa berdasarkan nilai-nilai mata kuliah dengan centroid yang sudah ditentukan.

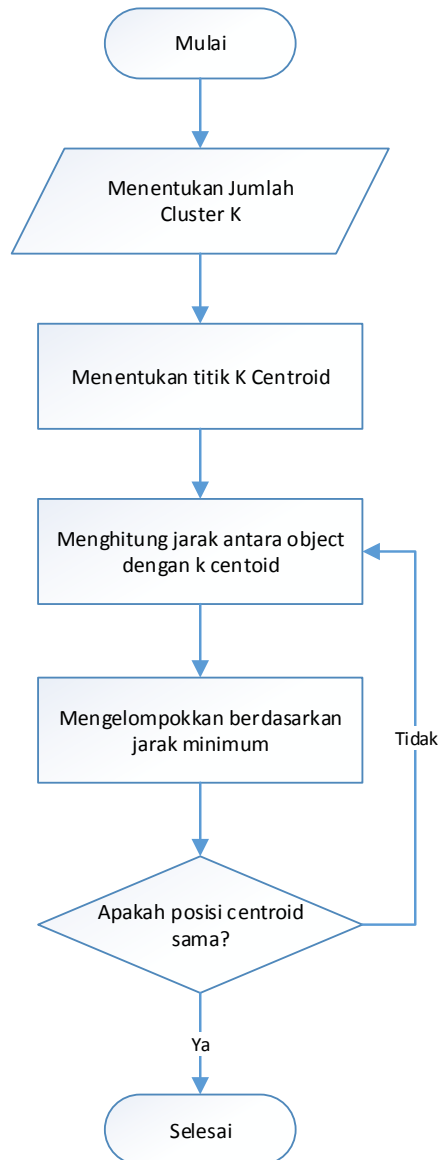
Berdasarkan analisa permasalahan yang ada, ditawarkan solusi pada Institut Asia Malang. Pembuatan aplikasi yang merupakan sebuah solusi dari permasalahan ini. Sistem melakukan proses berdasarkan data yang menjadi kriteria yang di tentukan dan menghasilkan kelompok mahasiswa yang sudah terbagi sesuai cluster. Data yang di gunakan untuk pengujian adalah data nilai mahasiswa. Dengan adanya sistem ini diharapkan mampu bekerja dengan baik dan memiliki kelebihan sebagai berikut:

1. Proses perhitungan yang cepat
2. Mempermudah proses pengelompokan
3. Data hasil pengelompokan bisa di simpan

Sistem penerapan data mining konsentrasi jurusan menggunakan algoritma K-means merupakan suatu sistem yang melakukan pengklasteran atau mengelompokan data konsentrasi jurusan sesuai nilai matakuliah menjadi 3 kelompok, yaitu data nilai mahasiswa tinggi, sedang dan rendah. Proses pengelompokan data dilakukan dengan teknik clustering, dengan menggunakan algoritma K-means, yaitu sebagai berikut :

1. Mencari nilai k sebagai jumlah cluster
2. Bangkitkan k centroid (titik pusat cluster) awal secara random

3. Hitung jarak setiap data ke masing-masing centroid.
4. Setiap data memilih centroid yang terdekat.
5. Tentukan posisi centroid baru dengan cara menghitung nilai rata-rata, dari rata-rata dari data yang memilih pada centroid yang sama.
6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.



Gambar 1. Flowchart K-Means

Gambar 1 merupakan flowchart pengelompokan data menggunakan algoritma K-Means dari penelitian terhadap konsentrasi jurusan mahasiswa. Dimulai dengan menentukan jumlah cluster kemudian dilanjutkan dengan titik centroidnya. Setelah itu, langkah yang selanjutnya dilakukan adalah menghitung jarak antara objek dengan centroidnya. Kemudian, penelitian dilanjutkan dengan mengelompokkan berdasarkan jarak minimumnya. Dari hasil pengelompokan jarak dapat dilihat posisi centroidnya. Jika posisi centroid sama maka pengelompokan selesai. Sebaliknya, jika posisi centroid tidak sama maka akan sistem akan kembali menghitung jarak antara objek dengan centroidnya.

Kebutuhan Input

Penelitian dilakukan pada tahun 2020, dilakukan tracing data mahasiswa yang sedang berjalan pada semester 3 yang akan melakukan perencanaan pemilihan konsentrasi di semester empat, sehingga data yang diambil adalah data mahasiswa prodi TI angkatan 2017. Data masukan yang digunakan terdiri atas :

- a. NIM atau Nama Mahasiswa.
- b. Data Nilai mahasiswa periode 2017 semester satu hingga semester tiga.
- c. Data Matakuliah semester satu hingga semester tiga.

Kebutuhan Proses

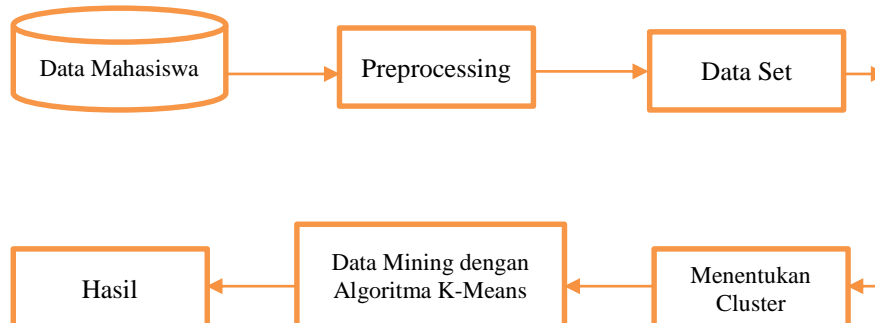
Proses awal yang dilakukan adalah dengan mengolah data mahasiswa Prodi Teknik Informatika semester tiga hingga didapatkan dataset yang kemudian akan di lakukan proses data mining dengan menggunakan algoritma K-Means.

Kebutuhan Output

Keluaran yang diharapkan dari pengolahan data ini yaitu dapat memberikan hasil berupa pengelompokan data konsentrasi yang dapat digunakan untuk membantu pihak kampus dalam mengelompokkan konsentrasi mahasiswa secara tepat. Sistem *clustering* untuk mengelompokkan konsentrasi dengan nilai yang dimiliki masing-masing mahasiswa dan menghasilkan keluaran yang terdiri dari cluster konsentrasi sistem cerdas (C1), cluster konsentrasi multimedia (C2), cluster konsentrasi jaringan(C3).

Pemodelan Input, Proses, dan Output

Pemodelan input, proses dan output difokuskan pada proses data mining untuk menentukan pengelompokkan konsentrasi jurusan berdasarkan cluster yang telah ditentukan dengan menggunakan algoritma K-Means yang ditunjukkan pada gambar 2.



Gambar 1. Pemodelan Sistem

Dari gambar 2 di atas dapat dijelaskan dari sebuah database mahasiswa (angkatan 2017) akan melalui proses Preprocessing dimana pemrosesan pendahuluan seperti sinkronisasi data mahasiswa agar sesuai dengan studi kasus. Dari proses tersebut akan tampil beberapa Data Set yang nantinya akan digunakan dalam proses mining data. Selanjutnya menentukan cluster dimana jumlah cluster akan ditentukan dan digunakan dalam pengelompokan data pada metode K-Means. Proses yang terakhir adalah Mining data menggunakan algoritma K-Means dimana merupakan proses mencari pola atau informasi yang menarik dalam data yang sebelumnya tidak diketahui dengan menggunakan algoritma.

Pre-processing

Preprocessing melibatkan validasi dan imputasi data [15], [16]. Tujuan dari validasi adalah untuk menilai tingkat kelengkapan dan akurasi data yang tersaring. Di sisi lain, tujuan imputasi adalah untuk memperbaiki kesalahan dan memasukkan nilai yang hilang, baik secara manual atau otomatis melalui program *business process automation* (BPA). *Data preprocessing* biasanya digunakan pada aplikasi *rule-based* dan yang berbasis *database*. Tahapan preprocessing pada penelitian ini dilakukan agar mendapatkan dataset yang sesuai dengan kebutuhan pada sistem. Selain itu juga untuk meminimalisir kesalahan yang ada pada data yang didapat.

Data Cleaning

Data preprocessing dibagi menjadi beberapa langkah, yaitu data cleaning, data transformation, dan data reduction [17], [18]. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Dalam data tersebut, terdapat record mahasiswa mencakup NIM (Nomor Induk Mahasiswa) dari setiap mahasiswa dan nilai kriteria matakuliah konsentrasi pada tahun 2017. Untuk detail data dapat dilihat pada Tabel 1.

TABEL 1
 CONTOH CLEANING DATA

No	NIM	Nama Mahasiswa	Id Matkul	Mata Kuliah	SKS	Nilai	Huruf	Bobot	Kelas	Sms
1	17201002	Dimas Edy Prasetyo	13	Komputer Desain	2	85	A	4	EKSA3	2
2	17201002	Dimas Edy Prasetyo	14	Animasi 2D	3	80	B+	3,5	EKSC1	3
3	17201002	Dimas Edy Prasetyo	21	Jaringan Komputer	3	85	A	4	EKSC1	4
4	17201002	Dimas Edy Prasetyo	1	Kecerdasan Buatan	4	70	B	3	EKSC1	4
5	17201003	Ega Radiansyah	13	Komputer Desain	2	80	B+	3,5	A3	2
6	17201003	Ega Radiansyah	14	Animasi 2D	3	85	A	4	C2	3
7	17201003	Ega Radiansyah	21	Jaringan Komputer	3	85	A	4	C1	4
8	17201003	Ega Radiansyah	1	Kecerdasan Buatan	4	70	B	3	C1	4
9	17201004	David Wahyu Pradana	13	Komputer Desain	2	70	B	3	A3	2
10	17201004	David Wahyu Pradana	14	Animasi 2D	3	80	B+	3,5	C2	3
11	17201004	David Wahyu Pradana	21	Jaringan Komputer	3	80	B+	3,5	C1	4
12	17201004	David Wahyu Pradana	1	Kecerdasan Buatan	4	70	B	3	C1	4
13	17201010	Siti Robiyatul Andhawiyah	13	Komputer Desain	2	0	E	0	B3	2
14	17201010	Siti Robiyatul Andhawiyah	14	Animasi 2D	3	65	C+	2,5	CW1	3
15	17201010	Siti Robiyatul Andhawiyah	21	Jaringan Komputer	3	0	C	2	C2	4
16	17201010	Siti Robiyatul Andhawiyah	1	Kecerdasan Buatan	4	70	B	3	C2	4
17	17201013	Alvin Fernando	13	Komputer Desain	2	85	A	4	B3	2
18	17201013	Alvin Fernando	14	Animasi 2D	3	85	A	4	CW1	3
19	17201013	Alvin Fernando	21	Jaringan Komputer	3	80	B+	3,5	C1	4
20	17201013	Alvin Fernando	1	Kecerdasan Buatan	4	80	B+	3,5	C1	4
21	17201017P	Veri Andriyawan	14	Animasi 2D	3	80	B+	3,5	EKSA3	1
22	17201017P	Veri Andriyawan	24	Keamanan Data	3	85	A	4	EKSA3	1
23	17201017P	Veri Andriyawan	28	Proyek Administrasi Jaringan	3	80	B+	3,5	EKSC1	3

No	NIM	Nama Mahasiswa	Id Matkul	Mata Kuliah	SKS	Nilai	Huruf	Bobot	Kelas	Sms
24	17201018	Bagas Dwi Pangga Romadlona	13	Komputer Desain	2	80	B+	3,5	B3	2
25	17201018	Bagas Dwi Pangga Romadlona	14	Animasi 2D	3	85	A	4	C2	3
26	17201018	Bagas Dwi Pangga Romadlona	21	Jaringan Komputer	3	70	B	3	C2	4
27	17201018	Bagas Dwi Pangga Romadlona	1	Kecerdasan Buatan	4	70	B	3	C2	4
28	17201021	Ricco Agustino	13	Komputer Desain	2	70	B	3	A3	2
29	17201021	Ricco Agustino	14	Animasi 2D	3	70	B	3	CW1	3
30	17201021	Ricco Agustino	21	Jaringan Komputer	3	80	B+	3,5	C1	4
31	17201021	Ricco Agustino	1	Kecerdasan Buatan	4	70	B	3	C1	4
32	17201023	Fidya Arifia Anggraini	21	Jaringan Komputer	3	85	A	4	A7	2
33	17201023	Fidya Arifia Anggraini	13	Komputer Desain	2	85	A	4	B3	2
34	17201023	Fidya Arifia Anggraini	14	Animasi 2D	3	85	A	4	C2	3
35	17201023	Fidya Arifia Anggraini	1	Kecerdasan Buatan	4	85	A	4	C1	4
36	17201024	Shahrul Abadi	13	Komputer Desain	2	85	A	4	A3	2
37	17201024	Shahrul Abadi	14	Animasi 2D	3	80	B+	3,5	C2	3
38	17201024	Shahrul Abadi	21	Jaringan Komputer	3	80	B+	3,5	C2	4
39	17201024	Shahrul Abadi	1	Kecerdasan Buatan	4	80	B+	3,5	C2	4

Simulasi Perhitungan

Simulasi perhitungan dengan metode K-Means dengan 79 data mahasiswa adalah sebagai berikut:

TABEL 2
 CONTOH DATA SET

No	NIM	Nama Mahasiswa	Sistem Cerdas	Multimedia dan Game	Administrasi Jaringan
1	17201002	Dimas Edy Prasetyo	70	82,5	85
2	17201003	Ega Radiansyah	70	82,5	85
3	17201004	David Wahyu Pradana	70	75	80
4	17201010	Siti Robiyatul Andhawayah	70	32,5	0
5	17201013	Alvin Fernando	80	85	80
6	17201017	Veri Andriyawan		80	82,5
7	17201018	Bagas Dwi Pangga Romadlona	70	82,5	70
8	17201021	Ricco Agustino	70	70	80
9	17201023	Fidya Arifia Anggraini	85	85	85
10	17201024	Shahrul Abadi	80	82,5	80

Tabel 2 menjelaskan dari contoh sampel 10 data mahasiswa angkatan 2017 memiliki tiga data set sesuai dengan konsentrasi pada jurusan Teknik Informatika. Sesuai studi Kasus ada tiga konsentrasi, yaitu Sistem cerdas (lengkapnya : Sistem Cerdas untuk Optimasi Aplikasi Bisnis), Multimedia dan Game, dan Administrasi Jaringan (lengkapnya: Sistem Administrasi Jaringan dan Keamanan Komputer). Berikut mata kuliah yang ditawarkan dan beberapa mata kuliah yang merupakan matakuliah konsentrasi yang dapat dilihat pada Tabel 3 berikut ini :

TABEL 3
 MATA KULIAH DAN KONSENTRASI

No	Mata Kuliah	Konsentrasi Jurusan
1	Kecerdasan Buatan	
2	Sistem Pakar	
3	Sistem Informasi Geografis	
4	Sistem Pendukung Keputusan	Sistem Cerdas Untuk Optimasi Aplikasi Bisnis
5	Data Mining	
6	Web/Text Mining	
7	Pengolahan Citra	
8	Big Data	
9	Pengembangan Game 1	
10	Pengembangan Game 2	
11	Komputer Desain	
12	Animasi 2D	
13	Produksi Musik Digital	Multimedia dan Game
14	Animasi 3D 1	
15	Animasi 3D2	
16	IMM1	
17	IMM2	
18	Jaringan Komputer	
19	Administrasi Jaringan	
20	Keamanan Jaringan	
21	Keamanan Data	
22	Sistem Multimedia	Sistem Administrasi Jaringan dan Keamanan Komputer
23	Pemrograman Jaringan	
24	Komunikasi Data	
25	Jaringan Komputer Lanjut	

Berdasarkan Gambar 1, proses algoritma K-Means dapat dijelaskan sebagai berikut :

Menentukan jumlah kelompok atau cluster

Data akan dikelompokkan kedalam 3 cluster, yaitu nilai konsentrasi sistem cerdas optimasi aplikasi bisnis, nilai konsentrasi multimedia dan game, dan nilai konsentrasi sistem administrasi jaringan dan keamanan komputer, seperti yang ditunjukkan pada Tabel 4.

TABEL 4
 CLUSTER

CLUSTER	STATUS
C1	Nilai Konsentrasi Sistem Cerdas Untuk Optimasi Aplikasi Bisnis
C2	Nilai Konsentrasi Multimedia dan Game
C3	Nilai Konsentrasi Sistem Administrasi Jaringan dan Keamanan Komputer

Menentukan pusat cluster (Centroid)

Penentuan iterasi 1 dengan menentukan centroid awal, dimana centroid awal dilakukan dengan merandom dari dataset. centroid awal diambil dari data ke-1 sebagai pusat cluster 1, data ke-3 sebagai pusat cluster 2 dan data ke-5 sebagai pusat cluster 3. Berikut centroid awal setiap cluster terpapar pada Tabel 5.

TABEL 5
 CENTROID AWAL

Centroid awal	A	B	C
C1	70	82,5	85
C2	70	75	80
C3	80	85	80

Menghitung iterasi pertama

- a. Selanjutnya proses menghitung jarak antara tiap-tiap data dengan pusat cluster (centroid). Untuk menghitung jarak tiap-tiap data terhadap setiap pusat cluster (Centroid) dapat menggunakan rumus pada persamaan (1) :

$$D(data, centroid) = \sqrt{\sum_j^n (data_i - centroid_i)^2} \tag{1}$$

Berikut contoh perhitungan jarak data mahasiswa pertama dengan titik pusat (centroid) sebagai berikut (2):

$$D_{1,1} = \sqrt{(70 - 70)^2 + (82,5 - 82,5)^2 + (85 - 85)^2} = 0 \tag{2}$$

Perhitungan jarak mahasiswa pertama dengan centroid cluster kedua, seperti berikut (3):

$$D_{1,2} = \sqrt{(70 - 70)^2 + (82,5 - 75)^2 + (85 - 80)^2} = 9,013878189 \tag{3}$$

Perhitungan jarak mahasiswa pertama dengan centroid pada cluster ketiga, seperti berikut (4):

$$D_{1,3} = \sqrt{(70 - 80)^2 + (82,5 - 85)^2 + (85 - 80)^2} = 11,45643924 \tag{4}$$

Pada Tabel 6 berikut ini, dijelaskan hasil perhitungan jarak antar pusat cluster secara lengkap sesuai perhitungan.

TABEL 6
 PERHITUNGAN JARAK ANTAR PUSAT CLUSTER

C1	C2	C3	JT
0	9,013878189	11,45643924	0
0	9,013878189	11,45643924	0
9,013878189	0	14,14213562	0
98,61541462	90,58835466	96,20940702	90,58835466
11,45643924	14,14213562	0	0
70,08922884	70,22285953	80,19507466	70,22285953
15	12,5	14,36140662	12,5
13,46291202	5	18,02775638	5
15,20690633	18,70828693	7,071067812	7,071067812
11,18033989	12,5	2,5	2,5
10,30776406	15	5	5
5,590169944	10	10	5,590169944
10,30776406	15	5	5

	C1	C2	C3	JT
0		9,013878189	11,45643924	0
15		12,5	14,36140662	12,5
...	
59,86675018		53,8000413	55,62773089	53,8000413
85,14693183		80,03905297	80,9706737	80,03905297
85,23659882		80,83333333	80,10843346	80,83333333
82,94492818		75,37867367	85,04083986	82,94492818
51,25		42,95710069	53,80810813	42,95710069

Mengelompokkan tiap-tiap data ke centroid terdekat

Pada tabel untuk data pertama, jarak terkecil terdapat pada cluster 1, data kedua terdapat pada cluster 1 seperti ditunjukkan pada Tabel 7.

TABEL 7
 KELOMPOK CLUSTER ITERASI 1

NIM	Cluster 1	Cluster 2	Cluster 3
17201002	1		
17201003	1		
17201004		1	
17201010		1	
17201013			1
17201017P		1	
17201018		1	
17201021		1	
17201023			1
17201024			1
17201029			1
17201032	1		
17201035			1
...
17201552P		1	
17201555		1	
17201557		1	
17201559P	1		
17201560		1	
Total	15	45	19

Menghitung pusat cluster (centroid) baru

1. Cluster 1

Tabel 7 menunjukkan hasil Cluster 1 dengan jumlah data 15, dengan rata-rata C1,a : 70,33333333; C1,b: 72,5; sampai dengan C1,c : 83,05555556 seperti pada persamaan (5);(6);(7).

$$C1, a = \frac{70+70+70+\dots}{15} = 70,33333333 \quad (5)$$

$$C1, b = \frac{82,5+82,5+85+\dots}{15} = 72,5 \quad (6)$$

$$C1, c = \frac{85+85+80+\dots}{15} = 83,05555556 \quad (7)$$

2. Cluster 2

Tabel 8 menunjukkan hasil Cluster 2 dengan jumlah data 45, dengan rata-rata C2,a : 61,835097; C2,b: 73,04907407; sampai C2,c : 42,09656085 seperti pada persamaan (8);(9);(10).

$$C2, a = \frac{70+70+0+\dots}{45} = 61,835097 \quad (8)$$

$$C2, b = \frac{75+32,5+80+\dots}{45} = 73,04907407 \quad (9)$$

$$C2, c = \frac{80+0+82,5+\dots}{45} = 42,09656085 \quad (10)$$

3. Cluster 3

Tabel 8 menunjukkan hasil Cluster 3 dengan jumlah data 19, dengan rata-rata C3,a : 81,05263158; C3,b: 83,55263158; sampai dengan C3,c: 83,15789474 seperti pada persamaan (11);(12);(13).

$$C3, a = \frac{80+85+80+\dots}{19} = 81,05263158 \quad (11)$$

$$C3, b = \frac{85+85+82,5+\dots}{19} = 83,55263158 \quad (12)$$

$$C3, c = \frac{80+85+80+\dots}{19} = 83,15789474 \quad (13)$$

TABEL 8
 CENTROID BARU

Centroid Baru	A	B	C
C1	70,33333333	72,5	83,05555556
C2	61,835097	73,04907407	42,09656085
C3	81,05263158	83,55263158	83,15789474

Hasil dari centroid baru ditunjukkan pada tabel 9. Iterasi berhenti ketika kelompok data terakhir sama dengan kelompok data sebelumnya sama, dengan kata lain iterasi akan berhenti ketika suatu centroid tidak berpindah tempat, dari data diatas dapat disimpulkan bahwa iterasi berhenti pada iterasi 8, karena data pusat centroid baru pada iterasi 7 sama dengan iterasi 8.

TABEL 9
 CENTROID BARU

	A	B	C
C1	71,66666667	10,83333333	74,86111111
C2	65,14880952	74,73541667	3
C3	70,33730159	80,66964286	79,20705782

III. HASIL DAN PEMBAHASAN

Implementasi program yang sudah dihasilkan melalui analisis diterapkan menggunakan bahasa pemrograman web menggunakan PHP (Native) serta Database menggunakan MySQL dengan algoritma clustering, yang digunakan untuk menentukan konsentrasi pada prodi Teknik Informatika di Institut Asia Malang. Menu aplikasi yang ditawarkan adalah : Nilai Mahasiswa (Nilai); Setting untuk menentukan Centriod Awal (di menu Profile user); Menjalankan proses Clustering (Clustering); Melihat hasil Clustering (Hasil Clustering); Memasukkan NIM dan Mahasiswa (Mahasiswa); Memasukkan Mata Kuliah yang nilainya akan diolah (Mata Kuliah).

Form List Nilai Mahasiswa

Halaman pertama yang muncul setelah user berhasil melakukan login, yaitu form List Nilai Mahasiswa. Didalam form ini terdapat *button new* yang digunakan untuk menginputkan data secara manual, *import by excel* menginputkan data secara keseluruhan dan *clear data* nilai digunakan untuk menghapus semua data yang ada di sistem. Dataset yang dimasukkan harus menggunakan file berupa excel 2016 (*.xls) karena data yang didapat dari Institut Asia Malang berupa excel 2016 dengan format (*.xls). Hasil *import* bisa dilihat pada Gambar 3. Pemanfaatan format Excel sangat diperlukan mengingat data mahasiswa untuk angkatan 2017 memiliki jumlah tidak sedikit. Jika input data mahasiswa yang akan diolah dilakukan secara manual (satu persatu), akan memakan waktu lama dan kesalahan inputan bisa saja terjadi. Untuk pengujian sistem dilakukan pada 100 data mahasiswa angkatan 2017 sebagai contoh data inputan.

NO	NIM	NAMA	MATA KULIAH	SEMESTER	NILAI
1	17201002	DIMAS EDY PRASETYO	Algoritma Pemrograman	1	80
2	17201002	DIMAS EDY PRASETYO	Arsitektur dan Organisasi Komputer	1	80
3	17201002	DIMAS EDY PRASETYO	Basic English	1	80
4	17201002	DIMAS EDY PRASETYO	Matematika Dasar	1	85

Gambar 3. Daftar Nilai Mahasiswa

Form Setting

Setelah selesai import nilai mahasiswa, selanjutnya menentukan pusat cluster awal secara random. Jadi sebelum ke perhitungan harus dilakukan *setting* terlebih dahulu pusat cluster awal seperti gambar 4. Pada form ini akan menentukan centroid awal secara random yang akan di gunakan dalam perhitungan iterasi pertama, Untuk perhitungan iterasi berikutnya akan otomatis terhitung sesuai dengan rumus penentuan centroid awal di iterasi berikutnya seperti pada perhitungan di atas. Untuk perhitungan jarak antar centroid akan mengikuti dari setting awal ini. Oleh sebab itu, Form Setting merupakan proses awal sebelum melakukan proses klusterisasi.

	A	B	C
C1	70	82.5	85
C2	70	75	80
C3	80	85	80

Gambar Error! No text of specified style in document.. Form Centroid Awal

Gambar 4. merupakan tampilan data yang disimpan pada database dari kriteria yang mempengaruhi dan alternatif yang ada. Setelah pengaturan menentukan cluster dan iterasi, data kemudian dapat diproses.

Form Matakuliah

Pada form ini terdapat beberapa mata kuliah yaitu mata kuliah umum dan mata kuliah konsentrasi. Mata kuliah umum merupakan mata kuliah dasar program studi. Untuk mata kuliah konsentrasi terbagi lagi menjadi mata kuliah pilihan konsentrasi dan mata kuliah wajib konsentrasi. Untuk total SKS yang wajib ditempuh oleh seorang mahasiswa adalah 145 SKS (secara keseluruhan termasuk total SKS mata kuliah konsentrasi program studi). Untuk mata kuliah wajib konsentrasi, setiap konsentrasi masing-masing memiliki 12 SKS dengan 18 SKS mata kuliah pilihan konsentrasi. Semester satu hingga semester tiga, mata kuliah yang wajib ditempuh oleh mahasiswa adalah 62 SKS yang terbagi dalam 21 mata kuliah, ke 21 mata kuliah tersebut yang akan diolah nilai tiap mahasiswanya sesuai pengelompokan yang telah ditentukan. Pada Gambar 5 akan ditampilkan antarmuka untuk pengisian data mata kuliah.

NO	MATA KULIAH	KONSENTRASI
1	Advanced English	Mata Kuliah Umum
2	Agama Islam	Mata Kuliah Umum

Gambar 5. Form Matakuliah

Form Klastering

Pada Form Klastering ini data yang tersimpan akan diproses. Pusat cluster awal menggunakan random pusat cluster awal pada *mining data* yang telah disimpan. Setelah pengaturan menentukan cluster, data kemudian dapat diproses. Gambar 6 menunjukkan hasil perhitungan yang telah dikelompokkan sesuai cluster yang telah ditentukan.

NIM	NAMA	SISTEM CERDAS	MULTIMEDIA	JARINGAN
17201002	DIMAS EDY PRASETYO	70	82.5	85
17201003	EGA RADIANSYAH	70	82.5	85
17201004	DAVID WAHYU PRADANA	70	75	80
17201010	SITI ROBIYATUL ANDHAWIYAH	70	32.5	0
17201013	ALVIN FERNANDO	80	85	80
17201017P	Veri Andriyawan	0	80	82.5
17201018	BAGAS DWI PANGGA ROMADLONA	70	82.5	70

Gambar 6. Form Klastering

Pada form ini berisi List Klastering Mahasiswa yang sudah bisa diproses. Untuk mengetahui secara rinci perhitungan dapat ditampilkan pada *button star clustering*. Selanjutnya hasil clustering bisa dilihat pada Gambar 7

Form Hasil Klastering

Tampilan hasil dalam sistem pengelompokan data mahasiswa ini, dapat dilihat pada gambar berikut ini :

NIM	NAMA	HASIL KLASTER
17201002	DIMAS EDY PRASETYO	Sistem Administrasi Jaringan dan Keamanan Komputer
17201003	EGA RADIANSYAH	Sistem Administrasi Jaringan dan Keamanan Komputer
17201004	DAVID WAHYU PRADANA	Sistem Administrasi Jaringan dan Keamanan Komputer
17201010	SITI ROBIYATUL ANDHAMIYAH	Multimedia dan Game
17201013	ALVIN FERNANDO	Sistem Administrasi Jaringan dan Keamanan Komputer
17201017P	Veri Andriyawan	Sistem Administrasi Jaringan dan Keamanan Komputer
17201018	BAGAS DWI PANGGA ROMADLONA	Sistem Administrasi Jaringan dan Keamanan Komputer
17201021	RICCO AGUSTINO	Sistem Administrasi Jaringan dan Keamanan Komputer
17201023	FIDYA ARIFIA ANGGRAINI	Sistem Administrasi Jaringan dan Keamanan Komputer
17201024	SHAHRUL ABADI	Sistem Administrasi Jaringan dan Keamanan Komputer

Gambar 7. Form Hasil Klustering

Dari gambar 7 dapat dijelaskan bahwa di kelas konsentrasi Sistem administrasi jaringan dan keamanan komputer terdapat 9 mahasiswa, yaitu: Dimas Edy Prasetyo, Ega Radiansyah, David Wahyu Pradana, Alvin Vernando, Very Andriyawan, Bagas Dwi Pangga, Rico Agustino, Fidya Arifia Anggraini, dan Shahrul Abadi. Sedangkan kelas konsentrasi Sistem Multimedia dan Game ada 1 mahasiswa yaitu Siti Robiyatul. Mahasiswa yang sudah tertera namanya dalam pengklasteran sistem, maka semester selanjutnya wajib mengambil konsentrasi tersebut dan mengikuti aturan yang sudah kaprodi tentukan. Bagi mahasiswa yang tidak memenuhi syarat nilai maka wajib mengulang matakuliah prasyarat masuk konsentrasi program studi.

Untuk mengetahui secara rinci perhitungan dapat ditampilkan pada button History Iterasi. Dalam menu tersebut hasil iterasi awal akan ditampilkan hingga iterasi terakhir yang menjadi hasil akhir dalam perhitungan. Pada Gambar 8 di bawah, ditunjukkan tampilan aplikasi untuk Iterasi klustering mahasiswa terhadap konsentrasi jurusan.

Literasi 1	Literasi 2	Literasi 3	Literasi 4	Literasi 5	Literasi 6	Literasi 7	Literasi 8
Perhitungan				Centroid awal			
NIM	NAMA	CERDAS	MULTIME	A	B	C	
17201002	DIMAS EDY PRASETYO	70	82.5	C1	70	82.5	85
17201003	EGA RADIANSYAH	70	82.5	C2	70	75	80
17201004	DAVID WAHYU PRADANA	70	75	C3	80	85	80
17201010	SITI ROBIYATUL ANDHAMIYAH	70	32.5				
17201013	ALVIN FERNANDO	80	85				
17201017P	Veri Andriyawan	0	80				
				Centroid baru			
				A	B	C	
				C1	70.313	73.125	83.177

Gambar 8 Form Iterasi Klustering Mahasiswa

Pengujian Validasi

Validasi merupakan proses membandingkan rata-rata pusat kluster awal dengan pusat kluster akhir. Pada perhitungan validasi pada data bergantung pada nilai K terdekat dengan tujuan perhitungan ini adalah untuk menghitung jumlah titik dengan label sama pada data latih.[15]. Dalam pemrosesan data seringkali terjadi bias. Bias merupakan suatu faktor pengaruh yang dapat berkontribusi dalam pembentukan hasil studi yang error atau tidak tepat. Untuk mengurangi bias yang disebabkan oleh sampel tertentu, seluruh proses training dan testing diulangi beberapa kali dengan sampel yang berbeda [19], [20]. Hasil validasi dapat dilihat pada tabel 10 dengan persamaan (14)

$$Bias = \frac{x-\mu}{\mu} \times 100\% \quad (14)$$

Ket: \times : Rerata hasil pengulangan pengujian

μ : nilai benar atau acuan

TABEL 10
 HASIL PENGUJIAN VALIDASI

Centroid	C1	C2	C3
Awal	73,33333333	80,83333333	81,66666667
Akhir	69,05092593	55,41279762	52,35605631
Error	5,839646465	31,44808542	35,89054329

Tabel 10 menunjukkan bahwa tingkat error dari hasil pengujian validasi adalah centroid awal dan akhir atribut pertama : 5,83% , atribut kedua : 31,44% , atribut ketiga : 35,89%.

IV. KESIMPULAN

Hasil pengujian program terhadap data dengan jumlah record data 100 record, menggunakan pusat cluster C1 = 70, 82.5, 85, C2 = 70, 75, 80 dan C3= 80, 85, 80 menghasilkan 6 iterasi dengan hasil Cluster 1 Mahasiswa direkomendasikan masuk Konsentrasi sistem cerdas. Dalam perhitungan di atas, ada 3 mahasiswa yang masuk dalam cluster 1. Cluster 2 Mahasiswa direkomendasikan masuk konsentrasi multimedia. Dalam perhitungan di atas, ada 20 mahasiswa yang masuk dalam cluster 2. Cluster 3 Mahasiswa direkomendasikan masuk Konsentrasi jaringan. Dalam perhitungan di atas, ada 56 mahasiswa yang masuk dalam cluster 3. Dari pengujian validasi diperoleh : centroid awal dan akhir atribut pertama : 5,83% , atribut kedua : 31,44% , atribut ketiga : 35,89%.

Diharapkan untuk mengembangkan pengklasteran konsentrasi jurusan Informatika menggunakan metode lain tidak hanya dengan metode K-Means dan penentuan jurusan konsentrasi menggunakan variabel lain selain nilai akademik seperti nilai prestasi non akademik yang linier dengan program studi. Ke depannya sistem penentuan konsentrasi yang telah dilakukan pada prodi Teknik Informatika di Institut Asia Malang dapat diterapkan pada prodi lain yang memiliki konsentrasi prodi masing-masing. Aplikasi ini dengan metode K-Means ini juga bisa diterapkan pada program studi diperguruan tinggi lain yang memiliki pola dan variabel yang sama dengan (atau mendekati sama) pemilihan konsentrasi di prodi Teknik Informatika di Institut Asia Malang.

DAFTAR PUSTAKA

- [1] A. P. Fadillah and M. R. Fachrizal, "Sistem Pendukung Keputusan Untuk Pemilihan Konsentrasi Mata Kuliah (Studi Kasus Program Studi Sistem Informasi Unikom)," *J. Manaj. Inform.*, vol. 8, no. 2, 2018, doi: 10.34010/jamika.v8i2.1029.
- [2] Hartatik, "Klasifikasi Konsentrasi Penjurusan Mahasiswa Universitas Amikom Yogyakarta," *Intecno J.*, vol. 1, pp. 15–18, 2019.
- [3] F. K. Wattimury and J. R. Utara, "PENENTUAN PEMINATAN MAHASISWA PRODI INFORMATIKA DI UNIVERSITAS AMIKOM YOGYAKARTA MENGGUNAKAN SVM," vol. 1, no. 4, pp. 1–4, 2019.
- [4] Y. Religia and A. Susanto, "PENERAPAN ALGORITMA K-MEANS DALAM PENGELOMPOKAN CURAH HUJAN DI DAERAH JABODETABEK," vol. 8, no. 2, pp. 217–224, 2018.
- [5] A. U. Fitriyadi, "Algoritma K-Means dan K-Medoids Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021, doi: 10.33322/kilat.v10i1.1174.
- [6] S. Setiawan, "Pemanfaatan Metode K-Means Dalam Penentuan Persediaan Barang," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 6, no. 1, pp. 41–48, 2018, doi: 10.33558/piksel.v6i1.1398.
- [7] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," *Reg. Dev. Ind. Heal. Sci. Technol. Art Life*, pp. 394–403, 2019.
- [8] W. A. Wahyuni and S. Saepudin, "Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci," in *Seminar Nasional Sistem ...*, 2021, pp. 306–313, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.i>

- d/index.php/sismatik/article/download/35/31.
- [9] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [10] I. Sumadikarta and E. Abeiza, "PENERAPAN ALGORITMA K-MEANS PADA DATA MINING UNTUK MEMILIH PRODUK DAN PELANGGAN POTENSIAL (Studi Kasus : PT Mega Arvia Utama)," *J. Satya Inform.*, vol. 1, no. 1, pp. 1–12, 2018.
- [11] Supangat and A. R. Amna, "Pemanfaatan Fitur Analisis Data Menggunakan K-Means Cluster Dalam Point of Sales (POS)," *Teknika*, vol. 8, no. 2, pp. 97–102, 2019, doi: 10.34148/teknika.v8i2.157.
- [12] Sadewo, "Algoritma K-Means Dalam Mengelompokkan Desa / Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," *Semin. Nas. Teknol. Comput. Sains SAINTEKS 2019*, pp. 754–761, 2018.
- [13] L. Zahrotun and U. Linarti, "Analisis Data Mining Pengelompokan Data Warga Menggunakan Metode Statistik K-Means," *J. Teknol. AKPRIND*, vol. 9, no. 1, pp. 18–23, 2016, [Online]. Available: <https://ejournal.akprind.ac.id/index.php/jurtek/article/view/1134/917>.
- [14] B. Parlambang and Fauziah, "Implementasi Algoritma K-Means Dalam Proses Penilaian Kuesioner Kepada Dosen Guna Mendukung Kepuasan Mahasiswa Terhadap Dosen," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 2, pp. 161–173, 2020, doi: 10.35760/tr.2020.v25i2.2719.
- [15] N. W. Saputra, A. P. Wibawa, U. Pujianto, and P. Anugrah, "Pengembangan Bahan Ajar Data Mining Menggunakan Four-D Model dalam Kerangka Kerja CDIO," *Belantika Pendidik.*, vol. 3, no. 2, pp. 48–58, 2020, doi: 10.47213/bp.v3i2.92.
- [16] B. K. Khotimah *et al.*, "Optimasi Bobot K-Means Clustering Untuk Mengatasi Optimization Weight of K-Means Clustering," vol. 8, no. 4, 2021, doi: 10.25126/jtiik.202184912.
- [17] Jasmir, "Implementasi Teknik Data Cleaning dan Teknik Roughset pada Data Tidak Lengkap dalam Data Mining," *Semin. Nas. APTIKOM*, pp. 99–106, 2016.
- [18] F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 91–93, 2013.
- [19] I. Menarianti, "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10, 2015, [Online]. Available: <http://ejournal.upgrismg.ac.id/index.php/JITEK/article/view/836>.
- [20] A. Ilham, "Level Data Untuk Menangani Data Kelas Tidak Seimbang," *J. Ilm. Ilmu Komput.*, vol. 3, no. 1, pp. 9–14, 2017.