

## ***EVALUASI JARAK CENTROID PADA PEMETAAN SEKOLAH MENEGAH ATAS UNTUK PENENTUAN LOKASI PROMOSI MAHASISWA BARU MENGGUNAKAN ALGORITMA K-MEANS***

### ***EVALUATION OF CENTROID DISTANCE ON HIGH SCHOOL MAPPING FOR LOCATION DETERMINATION OF NEW STUDENTS PROMOTION USING K-MEANS ALGORITHM***

Nur Aeni Widiastuti<sup>1\*</sup>, Noor Nailie Azzat<sup>2</sup>

<sup>1,2</sup>Universitas Islam Nahdlatul Ulama Jepara  
Email : <sup>1\*</sup>nuraeniwidiastuti@unisnu.ac.id, <sup>2</sup>nailie@unisnu.ac.id

**Abstrak** - Universitas Islam Nahdlatul Ulama Jepara merupakan salahsatu perguruan tinggi swasta yang ada dikabupaten Jepara yang terletak didaerah pesisir. Berdasarkan data yang diperoleh dari bagian UPT Pusat Data UNISNU, pada tahun 2016 hingga 2018 jumlah mahasiswa baru mengalami peningkatan yang kurang signifikan. Salahsatu permasalahan yang muncul adalah ketidaktepatan wilayah dalam melakukan strategi promosi secara langsung sehingga minat siswa berkurang. Solusi yang ditawarkan dengan pemetaan wilayah/sekolah yang berpotensi untuk dilaksanakannya strategi promosi secara langsung. Hasil dari penelitian ini dihasilkan 3 cluster yang mengarah ke jenjang pendidikan Sekolah Menengah Kejuruan (SMK). Hasil cluster dikatakan akurat dikarenakan jarak data mahasiswa pertama dengan pusat cluster pertama adalah 4, bahwa jarak data mahasiswa pertama dengan pusat cluster kedua adalah 1 dan bahwa jarak data mahasiswa pertama dengan pusat cluster ketiga adalah 1 dengan evaluasi yang digunakan dengan perbandingan nilai  $BCV = 21,09$  nilai  $WCV = 12211,29$  dan rasio keduanya = 0,0017 dengan kriteria Baik.

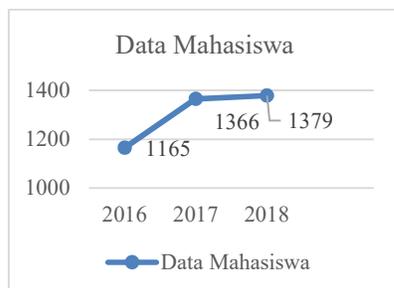
**Kata kunci:** data mining, k-means cluster, strategi promosi, data mahasiswa

**Abstract** - Universitas Islam Nahdlatul Ulama Jepara is one of the private universities in the district of Jepara which is located in the coastal area. Based on data obtained from the UPT section of the UNISNU Data Center, from 2016 to 2018 the number of new students experienced a less significant increase. One of the problems that arise is the inaccuracy of the region in carrying out direct promotion strategies so that student interest is reduced. The solution offered is by mapping areas/schools that have the potential to implement direct promotion strategies. The results of this study produced 3 clusters that lead to the level of Vocational High School (SMK) education. The cluster results are said to be accurate because the distance from the first student data to the first cluster center is 4, that the distance from the first student data to the second cluster center is 1 and that the distance from the first student data to the third cluster center is 1 with the evaluation used with a comparison of the  $BCV$  value = 21, 09  $WCV$  value = 12211.29 and the ratio of both = 0.0017 with Good criteria

**Keywords:** data mining, k-means cluster, promotion strategy, student data

## **1. PENDAHULUAN**

Perkembangan teknologi menghasilkan data yang berlimpah. Salahsatunya dibidang pendidikan yang mempunyai peranan penting dalam keberhasilan membangun suatu bangsa. Data yang berlimpah bisa diselesaikan dengan data mining yaitu pengambilan keputusan di masa yang akan datang berdasarkan informasi yang diperoleh dari masa lalu. Misalnya untuk prediksi, estimasi, asosiasi, *clustering*, dan deskripsi [1]. Universitas Islam Nahdlatul Ulama Jepara merupakan salahsatu perguruan tinggi swasta yang ada dikabupaten Jepara yang terletak didaerah pesisir. Selama ini mahasiswa yang ada berasal dari kabupaten Jepara sendiri sedangkan untuk luar kota sangat jarang. Berdasarkan data yang diperoleh dari bagian UPT Pusat Data UNISNU, pada tahun 2016 jumlah mahasiswa baru berjumlah 1165 mahasiswa, pada tahun 2017 berjumlah 1366 mahasiswa, dan pada tahun 2018 berjumlah 1379. Pada tahun 2016 hingga 2017 jumlah mahasiswa baru mengalami peningkatan sejumlah 201 mahasiswa. Pada tahun 2017 hingga 2018 jumlah mahasiswa baru mengalami peningkatan sejumlah 13 mahasiswa. Namun pada tahun 2018 peningkatan mahasiswa hanya sejumlah 13 mahasiswa lebih turun jika dibandingkan pada tahun 2017 berjumlah 201 mahasiswa. Hal ini sesuai dengan Gambar 1.



Gambar 1 Jumlah Penerimaan Mahasiswa Baru Tahun 2016 s/d 2018

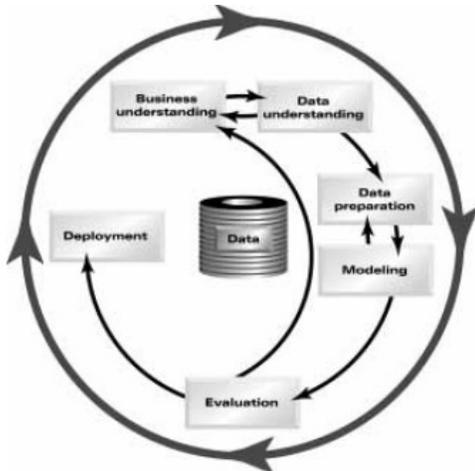
Sesuai pada Gambar 1. Peningkatan jumlah mahasiswa baru kurang signifikan. penentuan lokasi promosi yang tidak tepat sasaran akan membuat masalah di masa yang akan datang baik dari segi efisiensi waktu & biaya serta efektifitas pekerjaan. Dengan memanfaatkan data mahasiswa yang ada dan diolah dengan algoritma k-means cluster dapat digunakan untuk menyelesaikan masalah tersebut dengan melakukan pemetaan daerah asal, asal sekolah dan IPK.

Beberapa penelitian yang pernah dilakukan pertama oleh [2] kemudian dilanjutkan oleh [3][4][5][6][7][8] penelitiannya hampir sama tentang penentuan strategi promosi untuk mahasiswa baru dengan menggunakan algoritma k-means cluster dari hasil penelitian dihasilkan 3 cluster yang nantinya akan ditentukan strategi promosi yang akan digunakan baik itu menggunakan strategi periklanan, personal branding, mengadakan promosi, hubungan masyarakat dan pemasaran langsung. Perbedaan dengan yang lain adalah jumlah data dan parameter untuk hasil hampir sama baik itu terkait cluster dan strategi promosinya. Parameter yang sering digunakan seperti nama mahasiswa, jurusan yang diambil, kota asal mahasiswa dan nilai IPK [2][3] kemudian parameter dengan nama mahasiswa, NIM, kota asal, nilai IPK dan program studi [5][4], program study, gender, origin, religion, school, major, and promotion [6], kota asal mahasiswa, jurusan asal sekolah mahasiswa, jenis sekolah mahasiswa, jenis kelamin mahasiswa, agama mahasiswa dan jurusan sekolah mahasiswa [8]. Dari penelitian terdahulu, penelitian ini merupakan lanjutan dari penelitian [2] dengan menggunakan parameter yang sama tetapi jumlah data yang berbeda dan meneliti tentang kinerja jarak centroidnya terhadap keakuratan data yang akan dihasilkan. Sedangkan untuk strategi promosi yang dilakukan dengan melakukan promosi langsung. Tujuan dari penelitian ini adalah dihasilkan pemetaan lokasi yang tepat sasaran dalam melakukan promosi sehingga dapat meningkatkan jumlah mahasiswa baru di Universitas Islam Nahdlatul Ulama Jepara. Evaluasi kinerja akan diukur dengan menggunakan nilai BCV, MWC dan rasionya.

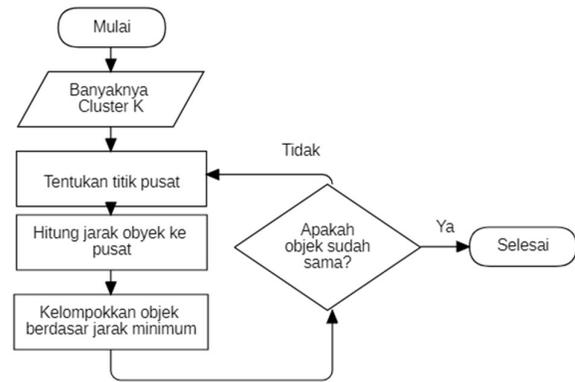
## 2. METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen, yaitu penelitian yang menerapkan algoritma data mining K-Means Clustering pada data mahasiswa UNISNU Jepara dari angkatan 2016 sampai dengan angkatan 2018 yang akan dianalisa dan dikelompokkan sesuai dengan persebaran wilayah dan sesuai asal sekolah berdasarkan indeks prestasi kumulatif selama dua semester awal yaitu pada semester 1 dan semester 2. Yang nantinya diuji atau diolah dengan *microsoft excel* dan *rapidminer studio*. Kemudian hasilnya akan dibandingkan berdasarkan kategori atau pengklasteran yang sudah ditentukan, yaitu tiga kategori: pertama kategori wilayah terbanyak, kedua kategori wilayah sedang, dan ketiga wilayah kategori sedikit yang nantinya dijadikan bahan pertimbangan dalam lokasi/ wilayah yang strategis untuk promosi Universitas Islam Nahdlatul Ulama Jepara. Parameter yang digunakan adalah kecamatan asal, asal SMA dan IPK. Tahapan dalam penelitian ini menggunakan metode CRISP-DM [9], yaitu:

- 1) Business Understanding,
- 2) Data Understanding,
- 3) Data Preparation,
- 4) Modelling,
- 5) Evaluation,
- 6) Deployment.



Gambar 2 Tahapan CRISP-DM [9]



Gambar 3 . Flowchart Algoritma K-Means Clustering [10]

Untuk tahapan penerapan algoritma K-Means clusternya langkah-langkahnya sebagai berikut [11]:

- a) Pilih secara acak k buah data sebagai pusat cluster.
- b) Jarak antara data dan pusat cluster dihitung menggunakan Euclidian Distance. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i,j)$  = Jarak data ke  $i$  ke pusat cluster  $j$

$X_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

- c) Data ditempatkan dalam cluster yang terdekat, dihitung dari tengah cluster.
- d) Pusat cluster baru akan ditentukan bila semua data telah ditetapkan dalam cluster terdekat.
- e) Proses penentuan pusat cluster dan penempatan data dalam cluster diulangi sampai nilai centroid tidak berubah lagi.

Setelah penerapan algoritma k-means kemudian tahap selanjutnya adalah evaluasi. Evaluasinya menggunakan metode *Between-Class Variation* (BCV) dan *Within-Class Variation* (WCV) pada iterasi terakhir yang sering disebut dengan rasio. Apabila hasil perhitungan pengujian yang diperoleh besar, maka semakin bagus tingkat kualitas *clustering* tersebut. BCV merupakan rata-rata dari *centroid*, sedangkan WCV adalah nilai keseluruhan dari jarak minimum yang telah dijumlahkan. Rumus perhitungannya adalah sebagai berikut [12]:

$$BCV = \frac{1}{Nk} \sum_i^k = 1 d(m_i, m_i) \tag{1}$$

Dimana:

- $k$  = Jumlah cluster
- $m_i$  = Jumlah anggota dari cluster ke- $i$
- $i$  = Nama yang mewakili cluster yang dibentuk
- $m_i$  = Jumlah anggota dari cluster ke- $i$

$$WCV = \sum_{j=i}^n \sum_{p \in c_i} d(p, m_i)^2 \tag{2}$$

Dimana:

- $p \in c_i$  = Jumlah semua data
- $k$  = Jumlah cluster
- $p$  = Cluster jarak terdekat
- $m_i$  = Jumlah anggota dari cluster ke- $i$

$$\text{Rasio} = \frac{BCV}{WCV} \tag{3}$$

Apabila nilai rasio yang didapat semakin kecil maka semakin bagus pula tingkat hasil dari akurasi cluster  $p$ , kriteria hasil ukuran rasio dapat dilihat pada tabel 1 [13].

Tabel 1 Kriteria Pengukuran Rasio

Nilai Rasio	Kriteria
$\leq 0,25$	Sangat baik
0,25- 0,50	Baik
0,50- 0,75	Kurang baik
0,75- 1,00	Buruk

### 3. HASIL DAN PEMBAHASAN

Tahapan dalam penelitian ini menggunakan metode CRISP-DM [9], yaitu: 1) *Business Understanding*, 2) *Data Understanding*, 3) *Data Preparation*, 4) *Modelling*, 5) *Evaluation*, 6) *Deployment*. Untuk rincian tahapan sebagai berikut:

#### 1) Business Understanding

Pada tahap ini fokus pada analisis kebutuhan bisnis. Dengan menentukan permasalahan dalam data mining kemudian merumuskan tujuan dan mempersiapkan strategi untuk mencapai tujuan tersebut. Permasalahan dalam penelitian ini adalah ketidaktepatan lokasi promosi dengan strategi promosi secara langsung sehingga menjadi salahsatu faktor menurunnya jumlah mahasiswa baru di lingkungan UNISNU Jepara. Sedangkan tujuannya adalah diharapkan dengan menerapkan algoritma k-means cluster dapat memetakan lokasi/dearah yang potensial untuk promosi sehingga lebih efektif dan efisien serta mengetahui tips and trik untuk mendukung strategi promosi tersebut.

#### 2) Data Understanding

Tahap ini merupakan tahapan pengumpulan data. Metode yang digunakan adalah studi literatur, studi pustaka, wawancara, dan dokumentasi. Data yang diperoleh dari bagian UPT Pusat Data UNISNU Jepara adalah data mahasiswa tahun 2016 s/d tahun 2018 dengan jumlah data 3910.

#### 3) Data Preparation

Tahap yang pertama yaitu pre-processing data adalah tahap dimana data yang sudah didapatkan, dipilah, dan dipisahkan agar mendapatkan data yang dibutuhkan untuk proses selanjutnya. Proses dalam tahapan pre-processing adalah sebagai berikut[14]:

##### a. Data Reduction

proses untuk mengurangi atau mereduksi sejumlah data yang tidak dibutuhkan. Data reduction sangat berguna untuk mendapatkan atribut dan data yang akan digunakan dalam penelitian ini. Atribut yang digunakan adalah kecamatan asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa semester pertama dan kedua pada angkatan 2016 sampai dengan angkatan 2018.

NO	NIM	NAMA	ALAMAT	Kecamatan	TAHU	Sekolah	Asal Sekol	Jenis	Prodi	IPK	IPK	Rata-rata
1	161120001961	ALIF MIFFAKHUR RIZQIYYA	PELEMKEREK MAYONG	Mayong	2016	SMA NEGERI 1 MAYONG	SMA	IPS	Akuntansi	3,30	3,30	3,30
2	161230000141	DIYAH AYU HERWATI	GLUPANGAN BANGSRI	Bangsri	2016	MA HASYIM ASY'ARI BANGSRI	MA	IPA	Teknik Sipil	3,48	3,74	3,61
3	161230000173	AHMAD NUR RIYAN SATRIA	BATEALIT RT 34/07 BATEALIT JEPARA	Batealit	2016	MA NEGERI 01 JEPARA	MA	IPA	Sistem Informatika	3,73	3,73	3,73
4	161310003556	AHMAD SYAFI I	JAMBU TIMUR	Mlonggo	2016	MA HASYIM ASY'ARI BANGSRI	MA	Keagamaan	Pendidikan Agt	3,53	3,64	3,58
5	161310003592	ROFIATUZ ZAHRO	BATEALIT	Batealit	2016	SMA/MA/SMK (Bagi Sekolah Y.SMA)	SMA	IPS	Pendidikan Agt	3,55	3,67	3,61
6	161310003650	HARIYONO	KANCILAN	Kembang	2016	PKBM Anyelir	PKBM	IPS	Pendidikan Agt	2,91	3,14	3,02
7	161110001799	NUR QOMARIAH	TELUKAWUR	Tahunan	2016	SMK NEGERI 3 JEPARA	SMK	Teknik Penyiaran Rasi	Manajemen	3,39	3,44	3,41
8	161110001800	RYKI BUDI SANTO	BANTRUNG	Batealit	2016	MA MATHOLUL HUDA BUGEI	MA	Keagamaan	Manajemen	2,75	2,78	2,77
9	161110001801	KURNIA INDRYANI	PANGGANG	Jepara	2016	SMA PGRI JEPARA	SMA	IPA	Manajemen	3,23	3,34	3,28
10	161110001802	DIFA ABIAS APRILINO	KRASAK,BANGSRI,JEPARA	Bangsri	2016	SMA NEGERI 1 BANGSRI	SMA	IPS	Manajemen	3,16	3,26	3,21
11	161110001803	IRIMA SALMA	ROBAYAN	Kalinyamatan	2016	SMK ISLAM AL HIKMAH MAYON	SMK	Administrasi Perkant	Manajemen	2,89	3,01	2,95
12	161110001804	ASMA KARIMAH	DESA SENGONBUGEL RT 05 RW 03 KI	Mayong	2016	SMA NEGERI 1 PECANGAAN	SMA	IPA	Manajemen	3,64	3,71	3,67
13	161110001805	VIVI SUSI KURNIAWATI	MAINTANGAN	Tahunan	2016	PKBM WIDYATAMA WELAHAN	PKBM	IPS	Manajemen	3,05	3,16	3,10
14	161110001806	WAHYU SETIYAWAN	DESA PULODARAT	Pecangaan	2016	SMA ISLAM SULTAN AGUNG 2 K	SMA	IPS	Manajemen	2,82	2,74	2,78
15	161110001807	VITA ARDIANSARI	DAREN	Nalumsari	2016	MA NURUS SALAM	MA	IPS	Manajemen	3,09	3,12	3,11
16	161110001808	SITI ISTIQOMAH	DEMANGAN	Tahunan	2016	MA WALISONGO PECANGAAN	MA	IPS	Manajemen	3,48	3,59	3,53
17	161110001809	HAMIDATUN NASIHATI	SUKODOONO	Tahunan	2016	SMK NEGERI 3 JEPARA	SMK	Tata Niaga	Manajemen	3,25	3,19	3,22
18	161110001810	KHODIRUDIN ANJUN NAJIB	DESA BANGSRI	Bangsri	2016	SMK NEGERI 1 BANGSRI	SMK	Administrasi Perkant	Manajemen	3,30	3,13	3,21
19	161110001811	SEPTIAN YUDI PRADANA	BUGEI	Kedung	2016	SMK NEGERI 1 KEDUNG	SMK	Manajemen	Manajemen	2,27	2,49	2,38
20	161110001812	NOVAL FEBY KURNIAWAN	KECAPI NGESONG	Tahunan	2016	MA AL-MAARIF SARIPAN	MA	IPS	Manajemen	2,77	2,99	2,88
21	161110001813	ISNA MARLIA ULFAH	JL JENDRAL SUDIRMAN, NO.33 DEM	Jepara	2016	SMK NEGERI 3 JEPARA	SMK	Administrasi Perkant	Manajemen	3,34	3,54	3,44
22	161110001814	INNAYATUL USRAK	JL KH.WACHID HASYIM BAPANGAN K	Jepara	2016	SMA NEGERI 1 TAHUNAN	SMA	IPA	Manajemen	3,75	3,83	3,79
23	161110001815	SITI LUTFIYAH	SUKOSOONO	Kedung	2016	SMK NEGERI 3 JEPARA	SMK	Administrasi Perkant	Manajemen	3,82	3,82	3,82
24	161110001816	LATIFATUL KHIIDMAH	MAINTANGAN	Tahunan	2016	SMK NEGERI 3 JEPARA	SMK	Administrasi Perkant	Manajemen	3,50	3,61	3,56
25	161110001817	ELIYANA WAHYUNINGSIH	MULYO HARJO	Jepara	2016	SMK NEGERI 1 JEPARA	SMK	Teknologi Pengolahan	Manajemen	3,48	3,62	3,55
26	161110001818	DEWI AISYAH	SURODADI KEDUNG JEPARA	Kedung	2016	MA NU MUALLIMAT KUDUS	MA	IPS	Manajemen	3,16	3,23	3,20
27	161110001819	A'RIYAH DWI RAHMAWATI	SLAGI	Pakis Aji	2016	SMK NEGERI 1 PAKIS AJI JEPARA	SMK	Administrasi Perkant	Manajemen	3,05	3,03	3,04
28	161110001820	MAHFUDIN DWIKA NURBENI	KELET RT 18 RW 02 KELING JEPARA	Keling	2016	MA NEGERI 2 JEPARA	MA	IPA	Manajemen	3,09	3,24	3,17
29	161110001821	DITA ANJIN PUTRI	TAHUNAN	Tahunan	2016	SMK TERPADU HADZIQIYAH B/SHK	SMK	Administrasi Perkant	Manajemen	3,43	3,62	3,53
30	161110001822	MUHAMMAD JAMALUL INSANI	SINANGGUL SIDANG RT39/RW08 ML	Mlonggo	2016	SMK NU MA ARIF 2 KUDUS	SMK	Teknik Otomotif Sepel	Manajemen	3,11	3,28	3,20
31	161110001823	NHURUL HIDAYAH	DERMOLO	Kembang	2016	SMA NEGERI 1 KEMBANG	SMA	IPA	Manajemen	3,36	3,48	3,42
32	161110001824	AHMAD NAITUL MUQID	TROSO	Pecangaan	2016	SMA NEGERI 1 TAHUNAN	SMA	IPS	Manajemen	2,86	2,65	2,75

Gambar 4 Tahapan Preprocessing Data

Pada file pertama yaitu file data mahasiswa semua angkatan dari tahun 2016 sampai tahun 2018. Peneliti mereduksi atribut Nama, Program Studi, Tahun Akademik, Jenis Sekolah, dan Jurusan Sekolah pada atribut tersebut. Dalam penelitian ini data yang tidak direduksi peneliti adalah NIM mahasiswa, Alamat, dan asal sekolah mahasiswa. Sehingga proses reduksi mendapatkan data mahasiswa dengan angkatan 2016 sampai dengan angkatan 2018 berjumlah 3910 data mahasiswa. Pada file kedua yaitu adalah file

data Indeks Prestasi Kumulatif (IPK) mahasiswa. File kedua ini berisikan beberapa kolom yaitu kolom NIM, tahun semester, dan IPK. Setelah peneliti menganalisis file data tersebut, peneliti mereduksi sejumlah data mahasiswa angkatan tersebut yang tidak dibutuhkan dalam penelitian ini.

NO	Asal Kecamatan	Asal Sekolah	IPK 1	IPK 2
1	Mayong	SMA	3.30	3.30
2	Bangsri	MA	3.48	3.74
3	Batealit	MA	3.73	3.73
4	Mlonggo	MA	3.53	3.64
5	Batealit	SMA	3.55	3.67
6	Kembang	PKBM	2.91	3.14
7	Tahunan	SMK	3.39	3.44
8	Batealit	MA	2.75	2.78
9	Jepara	SMA	3.23	3.34
10	Bangsri	SMA	3.16	3.26
11	Kalinyamatan	SMK	2.89	3.01
12	Mayong	SMA	3.64	3.71
13	Tahunan	PKBM	3.05	3.16
14	Pecangaan	SMA	2.82	2.74
15	Nalumsari	MA	3.09	3.12
16	Tahunan	MA	3.48	3.59
17	Tahunan	SMK	3.25	3.19

Gambar 1 Tahapan Data Reduction

b. Data Cleaning

Data cleaning adalah proses untuk mengisi *missing value* atau isian yang hilang atau kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redundansi data yaitu penumpukan data atau duplikasi data. Proses data cleaning pada penelitian ini menggunakan aplikasi Microsoft Excel dengan cara memfilter setiap kolom dan mencari data yang kosong atau *missing*.

Pada data awal yang berjumlah 3910 ada beberapa kolom yang kosong, peneliti tambah agar data bisa lengkap. Selain terdapat data yang kosong atau blank terdapat juga data yang kurang sesuai dalam penulisan, maka peneliti mengubah data yang tidak konsisten tersebut menjadi data yang konsisten dan menghasilkan data yang berkualitas. Setelah dianalisa terdapat data angka yaitu IPK mahasiswa dengan format yang kurang pas yaitu dengan format general maka peneliti mengganti format tersebut ke dalam format angka desimal. Selain data IPK, peneliti juga mengganti format pada data NIM Mahasiswa UNISNU ke format number karena data asli yang didapatkan menggunakan format general.

c. Data Transformation

Transformasi data dilakukan agar data dapat diolah dengan menggunakan metode K-Means Clustering. Data yang berjenis nominal seperti Kecamatan Asal dan Asal Sekolah harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk angka/numerikal.

Tabel 2 Inialisasi Data Kecamatan Asal

Kecamatan Asal	Frekuensi	Inisial
Jepara	8	1
Batealit	4	2
Bangsri	2	3
Tahunan	2	4
Donorojo	1	5
Mijen	1	6
Mlonggo	1	7
Pakis Aji	1	8
Pecangaan	1	9

Tabel 3 Inialisasi Data Asal Sekolah

Asal Sekolah	Frekuensi	Inisial
SMK	14	1
SMA	4	2
MA	3	3

## d. Data Integration

Data integration atau integrasi data adalah suatu proses untuk menggabungkan data dari beberapa file sumber. Data integration ini hanya dilakukan apabila data yang akan diolah bersumber dari beberapa file sumber. Pada penelitian ini peneliti menggunakan 2 file excel yang digabungkan sesuai dengan kebutuhan penelitian menjadi satu tabel dan satu file agar mudah untuk diolah kedalam metode K-Means clustering.

Setelah kedua data yang didapat melewati proses *data reduction*, *data cleaning*, dan *data transformation* maka langkah terakhir adalah menggabungkan atau mengintegrasikan semua data tersebut menjadi satu dataset ke dalam suatu tabel.

Tabel 4 Proses Data integration Mahasiswa UNISNU Jepara

NO	Kecamatan 2	Asal Sekolah	IPK
1	9	1	3,298
2	1	2	3,609
3	2	2	3,728
4	10	2	3,583
5	2	1	3,612
6	8	4	3,024
7	14	3	3,413
8	2	2	2,765
9	4	1	3,284
10	1	1	3,207
11	5	3	2,949
...	.....	....	.....
3908	14	2	3,477
3909	1	3	3,477
3910	14	1	2,898

## 4) Modelling

Penerapan algoritma K-Means Clustering menggunakan Microsoft Excel 2019 membutuhkan beberapa tahapan proses sebagai berikut:

**1. Menentukan Titik Pusat Cluster**

Titik pusat cluster atau bisa disebut centroid digunakan sebagai nilai pengurang untuk perhitungan jarak antara data ke setiap cluster atau disebut *distance*. Dalam proses ini penentuan nilai titik pusat cluster bisa ditentukan dengan metode acak sesuai keinginan peneliti dengan syarat nilai centroid masih termasuk dalam range nilai data pada setiap atribut. Selain itu penentuan nilai titik pusat cluster bisa ditentukan dengan menggunakan nilai rata-rata disetiap atribut.

Dalam proses ini peneliti menentukan nilai titik pusat cluster menggunakan nilai rata-rata dari setiap atribut dataset yaitu kecamatan asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa. Nilai rata-rata pada setiap atribut tersebut dipakai sebagai nilai centroid pada cluster 2.

Nilai rata-rata didapatkan dengan menjumlahkan nilai data pada setiap atribut kemudian dibagi dengan jumlah data yang ada. Nilai yang didapatkan adalah pada atribut kecamatan asal mahasiswa adalah 7,264, asal sekolah mahasiswa 2,187, dan IPK mahasiswa adalah 3,34. Lalu untuk mengisi nilai centroid pada cluster satu dan cluster tiga peneliti menentukan nilai centroid tersebut berdasarkan nilai centroid pada cluster dua yaitu nilai cluster satu kurang dari nilai cluster dua sedangkan nilai cluster tiga lebih tinggi dari nilai cluster dua. Sehingga peneliti mendapatkan nilai titik pusat setiap cluster / nilai centroid adalah sebagai berikut:

Tabel 5 Titik Pusat Awal Setiap Cluster

Titik Pusat	Kecamatan Asal	Asal Sekolah	IPK
Cluster 1	15	2	3,674
Cluster 2	16	2	3,227
Cluster 3	4	3	3,262

**2. Menghitung Jarak Data ke Setiap Cluster**

Setelah mendapatkan nilai titik pusat setiap cluster proses selanjutnya adalah menghitung jarak data ke setiap cluster atau bisa disebut *distance*. Cara untuk menghitung nilai jarak atau distance tersebut peneliti menggunakan sebuah rumus *Euclidean Distance*:

$$D(1,1)=\sqrt{(2 - 6)^2 + (1 - 1)^2 + (3663 - 2956)^2} = 4062$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat cluster pertama adalah 4,062. Jarak data mahasiswa pertama ke pusat cluster kedua:

$$D(1,2)=\sqrt{(2 - 3)^2 + (1 - 1)^2 + (3663 - 2077)^2} = 1875$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat cluster kedua adalah 1,875. Jarak data mahasiswa pertama ke pusat cluster ketiga:

$$D(1,3)=\sqrt{(2 - 1)^2 + (1 - 2)^2 + (3663 - 3554)^2} = 1418$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat cluster ketiga adalah 1,418. Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data mahasiswa pertama yang paling dekat adalah dengan cluster 3, sehingga data mahasiswa pertama dimasukkan ke dalam cluster 3. Hasil perhitungan selengkapnya untuk 21 data mahasiswa pertama dapat di lihat pada tabel

*Tabel 6 Contoh Hasil Perhitungan Setiap Data ke Setiap Cluster*

No	Kecamatan Asal	Asal Sekolah	IPK	Jarak Ke			Jarak terdekat ke Cluster
				C1	C2	C3	
1	2	1	3,66	5,51	2,19	0,88	3
2	7	1	3,52	0,58	3,40	5,69	1
3	8	1	3,54	0,58	4,36	6,68	1
...	...	...	...	...	...	...	...
19	9	2	3,51	1,68	5,27	7,69	1
20	1	1	3,40	6,50	2,99	0,55	3
21	1	1	3,82	6,52	3,12	0,76	3

**3. Mengalokasikan Data ke dalam Cluster**

Setelah proses perhitungan jarak data ke setiap cluster selesai maka proses berikutnya adalah mengalokasikan data kedalam setiap cluster yang terbentuk. Pengalokasian data tersebut berdasarkan hasil jarak antara data ke setiap cluster, apabila nilai jarak antara data pertama ke cluster 1 lebih kecil daripada nilai jarak antara data pertama ke cluster 2 ataupun cluster 3 maka data pertama masuk kedalam cluster 1. Pengalokasian data ini dilakukan bertujuan agar bisa menentukan titik pusat cluster baru pada proses selanjutnya.

**4. Menentukan Titik Pusat Cluster Baru**

Setelah proses pengalokasian data sudah selesai maka tahapan selanjutnya adalah menentukan titik cluster baru. Penentuan centroid baru ini menggunakan metode yang hampir sama dengan menentukan titik pusat cluster pada tahapan sebelumnya, yang membedakan dari kedua tahapan tersebut adalah dari jumlah data yang digunakan. Dalam tahapan ini centroid dapat ditentukan dengan cara menghitung nilai rata-rata sesuai dengan data yang sudah dikelompokkan kedalam setiap cluster masing- masing (tahapan pengalokasian data).

Dalam proses ini peneliti mencari nilai rata-rata berdasarkan data yang sudah dikelompokkan ke setiap cluster masing-masing dan mendapatkan hasil sebagai berikut:

Tabel 7 Pada cluster 1 dengan jumlah data 1407 data

<b>TITIK PUSAT CLUSTER BARU</b>					
Kecamatan Asal	12,466	Asal Sekolah	2,181	IPK	3,342

Tabel 8 Pada cluster 2 dengan jumlah data 118 data

<b>TITIK PUSAT CLUSTER BARU</b>					
Kecamatan Asal	12,466	Asal Sekolah	2,181	IPK	3,342

Tabel 9 Pada cluster 3 dengan jumlah data 2385 data

<b>TITIK PUSAT CLUSTER BARU</b>					
Kecamatan Asal	12,466	Asal Sekolah	2,181	IPK	3,342

### 5. Memverifikasi Titik Pusat Cluster

Proses ini adalah tahap untuk memverifikasi titik pusat cluster. Proses tersebut adalah proses untuk memverifikasi antara titik pusat cluster baru dengan titik pusat cluster yang lama, apabila kedua titik pusat cluster tersebut mempunyai nilai centroid yang berbeda maka proses K-Means Clustering masih berlanjut dan akan dimulai lagi pada proses ke-2 yaitu (menghitung jarak data ke setiap cluster) dengan menggunakan titik pusat cluster yang baru. Sedangkan jika nilai kedua centroid tersebut bernilai sama maka proses K-Means Clustering berhenti sampai di tahap tersebut. Proses verifikasi ini dilakukan dengan tujuan untuk menentukan apakah proses K-Means Clustering sudah selesai atau masih perlu dilakukan perulangan proses kembali. Dalam proses ini peneliti melalui 7 tahapan proses pengulangan untuk mendapatkan nilai titik pusat cluster atau centroid yang sudah tidak berubah lagi.

Setelah melewati tahapan proses K-Means Clustering menggunakan aplikasi Microsoft Excel 2019 maka proses berikutnya adalah menampilkan hasil dari penerapan tersebut. Untuk mendapatkan hasil yang diolah menggunakan algoritma K-Means Clustering peneliti melewati perulangan tahapan proses sebanyak 7 kali proses. Hasil dari penerapan ini adalah sebagai berikut:

Tabel 10 Hasil Cluster 1 Penerapan Algoritma K-Means Clustering

<b>CLUSTER 1 TERDIRI DARI 709 MAHASISWA YANG BERASAL DARI</b>			
<b>Kecamatan Asal</b>		<b>Asal Sekolah</b>	
Keling	94	SMA	181
Kembang	116	MA	250
Mayong	90	SMK	256
Mlonggo	361	PKBM	22
Nalumsari	48		
<b>TOTAL</b>	<b>709</b>		<b>709</b>

Berdasarkan hasil cluster 1 dapat diketahui terdapat 709 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal maupun asal sekolah mahasiswa beserta dengan jumlah dari masing-masing kecamatan asal dan asal sekolah tersebut. Jumlah data mahasiswa pada cluster 1 adalah jumlah data yang paling sedikit dibandingkan dengan cluster yang lain. Informasi yang menonjol dari cluster 1 adalah informasi mengenai persebaran wilayah kecamatan asal mahasiswa. Berdasarkan hal tersebut dapat

diketahui bahwa kecamatan asal dengan frekuensi paling sedikit adalah kecamatan yang kurang diminati untuk masuk di Universitas Islam Nahdlatul Ulama Jepara.

*Tabel 11 Hasil Cluster 2 Penerapan Algoritma K-Means Clustering*

<b>CLUSTER 2 TERDIRI DARI 1116 MAHASISWA YANG BERASAL DARI</b>			
<b>Kecamatan Asal</b>		<b>Asal Sekolah</b>	
Bangsri	1	SMA	268
Luar Kecamatan Jepara	115	MA	401
Pakis Aji	212	SMK	410
Pecangaan	228	PKBM	37
Tahunan	476		
Welahan	84		
<b>TOTAL</b>	<b>1116</b>		<b>1116</b>

Berdasarkan hasil cluster 2 dapat diketahui bahwa terdapat 1116 data mahasiswa. Jumlah data mahasiswa pada cluster 2 adalah jumlah data yang kurang dari cluster 3 tetapi lebih dari cluster 1. Informasi yang menonjol dari cluster 2 adalah informasi mengenai persebaran asal sekolah mahasiswa yaitu banyaknya jumlah jurusan asal sekolah yang berbeda beda dengan frekuensi yang berbeda pula. Berdasarkan hal tersebut dapat diketahui bahwa jurusan asal sekolah dengan frekuensi paling sedikit adalah asal sekolah yang kurang diminati untuk masuk di Universitas Islam Nahdlatul Ulama Jepara.

*Tabel 12 Hasil Cluster 3 Penerapan Algoritma K-Means Clustering*

<b>CLUSTER 3 TERDIRI DARI 2085 MAHASISWA YANG BERASAL DARI</b>			
<b>Kecamatan Asal</b>		<b>Asal Sekolah</b>	
Bangsri	490	SMA	493
Batealit	446	MA	746
Donorojo	94	SMK	804
Jepara	643	PKBM	40
Kalinyamatan	129	SKB	1
Kedung	283	DLL	1
<b>TOTAL</b>	<b>2085</b>		<b>2085</b>

Berdasarkan hasil cluster 3 dapat diketahui terdapat 2085 data mahasiswa. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kecamatan asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah dari masing-masing kecamatan asal dan asal sekolah tersebut. Jumlah data mahasiswa pada cluster 1 adalah jumlah data yang paling banyak diantara cluster yang lain sehingga dapat dikatakan bahwa mayoritas mahasiswa UNISNU paling banyak masuk ke dalam kelompok cluster 1. Berdasarkan hal tersebut pihak admisi dapat mempertimbangkan strategi promosi di tahun depan menggunakan hasil persebaran wilayah dan asal sekolah pada cluster 3.

##### 5) Evaluation

Dalam eksperimen data UNISNU Jepara, data mahasiswa angkatan 2016 sampai angkatan 2018, pasti memiliki nilai error, semakin kecil nilai error yang dimiliki pada hasil perhitungan maka semakin bagus pula

hasil yang akan didapatkan. Pada penelitian ini menghitung nilai error menggunakan persamaan yang pada sebelumnya sudah dijelaskan, perhitungan nilai error terdapat pada proses berikut ini:

#### Pengujian Metode BCV dan WCV

1. Menentukan iterasi beberapa akan dihitung

Untuk menentukan iterasi diambil pada iterasi terakhir karena iterasi terakhir memiliki kualitas centeroid yang lebih baik dari sebelumnya, pada penelitian ini menggunakan data perhitungan dari data mahasiswa UNISNU angkatan 2016 sampai angkatan 2018, untuk lebih jelasnya dapat Tabel di bawah ini.

Tabel 13 Nilai Centroid pada iterasi terakhir

Titik Pusat	Kecamatan Asal	Asal Sekolah	IPK
Cluster 1	9,22	2,02	3,32
Cluster 2	13,70	2,03	3,33
Cluster 3	3,16	1,99	3,35

Kemudian hitung nilai *Centroid* dengan persamaan *Between-Class Variation* (BCV).

$$\begin{aligned} \text{BCV} &= \sqrt{(9,22 - 13,70)^2 + (2,02 - 2,03)^2 + (3,32 - 3,33)^2 +} \\ &= \sqrt{(9,22 - 3,16)^2 + (2,02 - 1,99)^2 + (3,32 - 3,35)^2 +} \\ &= \sqrt{(13,70 - 3,16)^2 + (2,03 - 1,99)^2 + (3,33 - 3,35)^2} \\ &= 21,09 \end{aligned}$$

2. Menentukan jarak minimum centeroid

Pada proses ini menggunakan jarak minimum pusat centeroid yang didapat pada iterasi terakhir, dapat dilihat pada tabel 4.10 Setelah mendapatkan jarak minimum dengan nilai pusat centroid maka langkah selanjutnya menghitung seluruh jarak minimum dengan persamaan *Within-Class Variation* (WCV) sebagai berikut:

$$\text{WCV} = 0,22^2 + 2,39^2 + 1,58^2 + 1,28^2 + \dots + 2,38^2 + 0,52^2$$

Sehingga hasil yang didapat adalah  $\text{WCV} = 12211,29$

3. Menghitung perbandingan BCV dengan WCV

Pada langkah terakhir adalah menghitung nilai perbandingan BCV dengan WCV sehingga menghasilkan nilai error hitung dengan persamaan Rasio seperti terlihat pada hasil dibawah ini.

$$\text{Rasio} = \frac{21,09}{12211,29} = 0,0017$$

Hasil pengujian menggunakan perbandingan *Between-Class Variation* (BCV) dan *Within-Class Variation* (WCV) mendapatkan nilai rasio yang tidak tinggi yaitu 0,0017 dan artinya tingkat penggunaan nilai sample data *Centroid* memiliki kualitas yang baik.

#### 6) Deployment

Pada tahap ini sesuai dengan hasil penelitian yang diperoleh 3 cluster sesuai dengan Tabel 9, Tabel 10, dan Tabel 11.

Tabel 14 Hasil Cluster

Jenis Cluster	Asal Kecamatan	Jumlah Mahasiswa	Asal Sekolah	Jumlah Cluster Mahasiswa
Cluster 1	Mlonggo	361	SMK	256
Cluster 2	Tahunan	476	SMK	410
Cluster 3	Jepara	643	SMK	804

Dari hasil penelitian *Cluster 1* adalah kelompok dengan cukup peminat yaitu berasal dari kecamatan Mlonggo berjumlah 361 mahasiswa dan untuk asal sekolah berasal dari SMK dengan jumlah 256 mahasiswa. Pada *cluster 2* kelompok dengan cukup peminat yaitu berasal dari kecamatan Tahunan berjumlah 476 mahasiswa dan untuk asal sekolah berasal dari SMK dengan jumlah 410 mahasiswa. Dan *cluster 3* kelompok dengan banyak peminat yaitu berasal dari kecamatan Jepara berjumlah 643 mahasiswa dan untuk asal sekolah berasal dari SMK dengan jumlah 804 mahasiswa. Maka, kelompok yang perlu mendapatkan prioritas utama dari pihak Universitas adalah kelompok yang berada pada *cluster 3*. Oleh sebab itu evaluasi kinerja dari jarak centroid sangat baik hal ini dapat dilihat dari nilai  $BCV = 21,09$   $WCV = 12211,29$  dan nilai rasionya = 0,0017 dengan kriteria Baik. Untuk strategi promosi secara langsung bisa melakukan pembinaan terhadap wilayah tersebut untuk meningkatkan animo masyarakat terkait pentingnya untuk melanjutkan pendidikan ke jenjang perguruan tinggi terutama untuk kuliah di UNISNU Jepara.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian analisis data mahasiswa dari tahun 2016 sampai dengan tahun 2018 digunakan untuk evaluasi kinerja jarak centroid yang menggunakan algoritma k-means. Terdapat 3 cluster yang terbentuk. Dengan bahwa jarak data mahasiswa pertama dengan pusat cluster pertama adalah 4, bahwa jarak data mahasiswa pertama dengan pusat cluster kedua adalah 1 dan bahwa jarak data mahasiswa pertama dengan pusat cluster ketiga adalah 1 dengan evaluasi yang digunakan dengan perbandingan nilai  $BCV = 21,09$  nilai  $WCV = 12211,29$  dan rasio keduanya = 0,0017 dengan kriteria Baik.

Diharapkan untuk penelitian selanjutnya dengan menggunakan teknologi machine learning sehingga memudahkan melakukan promosi mahasiswa baru.

#### DAFTAR PUSTAKA

- [1] S. C. Widiastuti Nur Aeni, Santosa Stefanus, "ALGORITMA KLASIFIKASI DATA MINING NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK DETEKSI PENYAKIT JANTUNG," *Pseudocode*, vol. 1, no. 1, pp. 11–14, 2014.
- [2] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *Jurnal Ilmiah teknik Industri*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [3] R. D. Ramadhani, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro," *Industrial Marketing Management*, vol. 1, no. 1, pp. 1–9, 2014.
- [4] M. Tonggiroh and M. T. Jufri, "Data Mining Strategi Promosi Pada Universitas Yapis Papua Menggunakan Algoritma K-Means Clustering," *Seminar Nasional Teknologi Informasi Dan Komunikasi (SEMNASITIK) X*, vol. 8, pp. 587–594, 2018.
- [5] I. Kurniawati, R. E. Indrajit, and M. Fauzi, "Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru," vol. 1, no. 2, pp. 70–79, 2017.
- [6] G. A. Sandag, E. Yahuda Putra, R. L. Wurangian, and N. Believer Tulangow, "Analysis of Strategy for Targeted New Student Using K-Means Algorithm," *2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019*, vol. 1, no. August, pp. 94–99, 2019.
- [7] R. Budiman and R. Anto, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)," *ProTekInfo(Pengembangan Riset dan Observasi Teknik Informatika)*, vol. 6, no. 1, p. 6, 2019.
- [8] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 1–14, 2020.
- [9] M. Boussaa, I. Atouf, M. Atibi, and A. Bennis, "Journal Of Data Warehousing," *Proceedings of 2016 International Conference on Electrical and Information Technologies, ICEIT 2016*, vol. 5, no. 4, pp. 480–484, 2016.
- [10] M. A. Nur, "PENERAPAN ALGORITMA K-MEANS DALAM PENENTUAN SISWA BERMASALAH DI SMK WALISONGO PECANGAAN BERDASARKAN ANECDOTAL RECORD," Universitas Islam Nahdlatul Ulama Jepara, 2020.
- [11] R. K. Daniel and S. Eko, "Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis," *JdC*, vol. 2, no. 2, pp. 18–25, 2013.
- [12] S. M. Amril, "PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DAERAH RAWAN BENCANA DI INDONESIA," *Information System Journal*, pp. 1–10, 2010.
- [13] J. E. Gentle, L. Kaufman, and P. J. Rousseuw, "Finding Groups in Data: An Introduction to Cluster Analysis.," *Biometrics*, vol. 47, no. 2, p. 788, 1991.
- [14] D. Fahmi and Assena, "MODEL ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN

STRATEGI PROMOSI UNIVERSITAS ISLAM NAHDLATUL ULAMA JEPARA,” Universitas Islam Nahdlatul Ulama Jepara, 2020.

- [15] Kusrini and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: Andi, 2009.