

Implementasi Metode K-Means Dalam Klasifikasi Desa/Kelurahan Menurut Jenis Industri Kecil Mikro

Muhammad Arifullah*, Jaya Tata Hardinata, Yuegilion Pranayama Purba

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar

Jl. Kartini, Proklamasi, Kec. Siantar Bar., Kota Pematang Siantar, Sumatera Utara, Indonesia

Email: ¹*Purnama19990827@gmail.com, ²jayatata@amiktunasbangsa.ac.id, ³yuegilionpranayama@amiktunasbangsa.ac.id

Email Penulis Korespondensi: Purnama19990827@gmail.com

Abstrak—Kementerian perindustrian mencatat jumlah industri kecil menengah di Indonesia mencapai 44 juta. Namun sayangnya 99% dari seluruh industri kecil memiliki masalah dan kesulitan mengembangkan usaha. Masalah yang sering dihadapi di Indonesia adalah keterbatasan kemampuan digital marketing. Produksi yang belum stabil dan lainnya. Guna membantu/mendukung usaha kecil dan mikro maka dilakukan pengelompokan dikecamatan di pematangsiantar yang memiliki usaha kecil/mikro menurut jenis usaha menggunakan metode k-means. Data penelitian yang digunakan di peroleh dari BPS dan dikelompokkan menjadi 2 kelompok kecil dan mikro. Dengan mengimplementasikan metode k-means pada aplikasi rapidminer dapat menghasilkan suatu pohon keputusan berdasarkan data set learning usaha industri kecil dan micro menengah sebelumnya dengan menggunakan kriteria yang dimaksudkan agar menghasilkan hasil klasifikasi usaha industri kecil, mikro dan menengah yang bisa relevan dan akurat.

Kata Kunci: Data Mining; Algoritma K-Means; Pengelompokan; Industri Kecil dan Micro

Abstract—The Ministry of Industry noted that the number of small and medium industries in Indonesia reached 44 million. But unfortunately 99% of all small industries have problems and difficulties in developing their businesses. The problem that is often faced in Indonesia is the limited ability of digital marketing. Unstable production and others. In order to help/support small and micro businesses, a grouping is carried out in the Pematangsiantar sub-district that has small/micro businesses according to the type of business using the K-Means method. The research data used were obtained from BPS and grouped into 2 high and low groups. By implementing the k-means method in the rapidminer application, it is possible to produce a decision tree based on the previous learning data set of small and micro-medium industries using criteria intended to produce relevant and accurate results of small and medium-sized industrial business classification.

Keywords: Data Mining; K-Means Algorithm; Grouping; Small and Micro Industries

1. PENDAHULUAN

Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer untuk menganalisis dan mengekstrak pengetahuan secara otomatis atau serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [1], [2]. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan database [3]. Data mining juga merupakan salah satu bagian proses Knowledge Discovery from Databases (KDD) yang bertugas untuk mengekstarak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik [4]. Pengembangan UMKM di Kota Pematangsiantar tidak lepas dari peran serta pemerintah dan para pelaku UMKM. Pemerintah sebaiknya dapat melakukan pendampingan kepada para pelaku UMKM, serta sosialisasi dan pelatihan-pelatihan, dan para pelaku UMKM sebaiknya lebih kreatif dan inovatif dalam memproduksi barang. Bahwa pengusaha UMKM mendapatkan kendala dalam pengembangan usaha yang dijalaninya. Semakin banyak keterampilan yang diperoleh dan dikuasai industri kreatif kota pematang siantar, semakin banyak pula peluang untuk menambah perekonomian daerah. Usaha kecil dan menengah (UKM) merupakan salah satu bagian penting dari perekonomian suatu negara ataupun daerah. Berbagai upaya pengembangan UKM telah dilakukan, salah satunya dengan membangkitkan dan memperbanyak orang atau pengusaha baru di bidang UKM, sehingga masyarakat desapun diberi keterampilan dengan harapan keterampilan tersebut menjadi sebuah usaha kreatif yang memberi manfaat bagi perekonomian keluarga dan masyarakat desa. Demikian pula usaha kreatif tersebut juga dapat membuka kesempatan dan lapangan kerja baru bagi masyarakat.

Maka perlu dibuat sebuah sistem menggunakan data mining metode K-Means untuk mengklasifikasi desa/kelurahan menurut jenis industri kecil dan mikro (UMKM), sistem ini dapat membantu untuk mempermudah dalam mengklasifikasi desa/kelurahan yang tergolong dalam industri sesuai perkembangannya pola pikir yang kreatif dan inovatif. Dimana hal ini dapat berguna bagi masyarakat dalam peningkatan perkembangan kebutuhan ekonomi yang meningkat. Penelitian ini dilandasi oleh peneliti terdahulu ada 2 jurnal yang menjadi referensi dalam penulisan skripsi ini yaitu pertama, penelitian yang dilakukan [5]. Dalam jurnal tersebut, peneliti melakukan Strategi Pengembangan Usaha Mikro, Kecil dan Menengah (UMKM) di Kota Pematangsiantar. Kedua, yaitu penelitian yang dilakukan oleh (Katrina, 2019) membahas tentang model aturan klasifikasi minat mahasiswa berwirausaha dengan algoritma naïve bayes. dengan atribut yang digunakan adalah Mikro, kecil, dan menengah.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

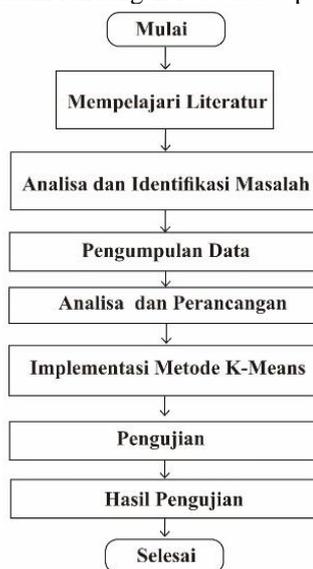
Metode penelitian terdiri atas dua metode, yaitu metode penelitian kualitatif dan metode penelitian kuantitatif. Metode penelitian kualitatif merupakan metode yang berlandaskan pada filsafat post positifisme, serta sebagai metode artistic karena proses penelitian lebih bersifat seni (kurang terpola). Sedangkan Metode penelitian kuantitatif merupakan salah satu jenis penelitian yang spesifikasinya adalah sistematis, terencana, dan terstruktur dengan jelas sejak awal hingga pembuatan desain penelitiannya[6]. Setelah tahu perbedaan pendekatan penelitian kuantitatif dan kualitatif, Penulis dapat mengambil kesimpulan bahwa penulis menggunakan pendekatan yang dapat diselesaikan dengan komputasi numerik yaitu dengan penelitian kuantitatif yang menuntut lebih banyak terhadap penggunaan angka-angka. Dimana, komputasi numerik merupakan suatu pendekatan penyelesaian masalah matematika dengan menggunakan beberapa metode numerik.

2.1.1 Lokasi dan Waktu Penelitian

Dalam penyusunan penelitian ini, penulis melakukan penelitian di Desa/Kelurahan yang ada di Pematangsiantar. Waktu pengumpulan data dan pengolahan data dilakukan selama satu minggu dengan waktu dimulai tanggal 31 Agustus s/d 05 September 2020.

2.1.2 Rancangan Penelitian

Rancangan atau model penelitian disajikan dalam rancangan *Flowchart* pada gambar 1[7], [8]:



Gambar 1. Rancangan Penelitian

Gambar 1 menjelaskan rancangan penelitian yang dilakukan untuk mencari daerah dengan hasil industri kecil dan micro secara optimal dengan Algoritma K-Means yang terdiri dari[9], [10] :

a) Analisis Masalah

Masalah yang terkait dengan minimnya hasil industri kecil dan micro yang di panen di wilayah Sumatera Utara maka penulis berusaha mencari penyebab dengan memetakan daerah yang belum optimal. Dengan mengetahui kedudukan masalah dalam konteks keilmuan yang ada, peneliti dapat menelusuri dan mendalami permasalahan itu dan menempatkannya dalam pokok bahasan atau sub pokok bahasan bidang ilmu tersebut. Dengan cara ini peneliti dengan mantap memiliki pangkal tolak dan sudut pandang keilmuan yang ada.

b) Mempelajari Literatur

Penelitian ini harus didasari rujukan yang digunakan untuk mendapatkan rujukan yang digunakan untuk mendapatkan informasi dalam penelitian. Tinjauan kepustakaan biasa disebut pula studi literatur. Pada kerangka ini pembaca mencari, membaca, dan mereview bacaan yang relevan dengan topik penelitian. Tak jarang literatur yang relevan banyak jumlahnya. Pada kasus ini, pembaca harus menyeleksi mana yang lebih relevan, mana yang kurang relevan.

c) Mengumpulkan Data

Metode pengumpulan data yang diterapkan harus mampu menjawab rumusan masalah. Jadi, tugas peneliti disini untuk memilih dan memilah mana metode pengumpulan data yang relevan. Relevansi metode pengumpulan data sangat tergantung pada rumusan masalah yang dirumuskan.

d) Analisa

Proses yang dilakukan untuk mencari daerah dengan hasil industri kecil dan micro secara optimal dengan pemetaan. Analisa data bisa dilakukan secara manual atau dengan bantuan komputer. Cara melakukannya juga berbeda-beda

tergantung metode riset apa yang digunakan. Riset kuantitatif sering kali menerapkan metode analisis yang berbeda dengan riset kualitatif. Perbedaan analisis data bisa dilihat dari perbedaan software yang digunakan jika analisis dilakukan dengan bantuan komputer.

e) Implementasi

Implementasi dilakukan dengan menggunakan aplikasi *RapidMiner versi 8.1* sebagai sistem yang mencari keputusan dalam analisa daerah dengan hasil industri kecil dan mircro secara optimal dengan pemetaan.

f) Keputusan

Hasil yang diberikan oleh sistem akan ditindak lanjuti oleh pihak Dinas Perkebunan Provinsi Sumatera Utara. Sehingga perlu mendapat perhatian dan penanganan efektif karena berkaitan dengan pengambilan kebijakan penyaluran bantuan yang dilakukan.

2.2 Analisis Data

Tahap praproses data merupakan tahap seleksi data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Melakukan seleksi dengan memperhatikan konsistensi data, *missing value* dan *redundant* data. Kemudian data diolah menggunakan *RapidMiner* menggunakan *Performance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data[11]. Dalam penelitian, data yang digunakan akan diolah dari BPS Pematangsiantar tahun 2018 - 2020 dengan jumlah 8 Kecamatan. Dengan tiga type usaha yang sedang berjalan seperti **Swadaya** berarti kekuatan atau tenaga sendiri, **Swakarsa** berarti keinginan atau kemauan sendiri yang muncul tanpa adanya paksaan atau dorongan dari orang lain dan **Swasembada** berarti suatu usaha untuk mencukupi kebutuhan diri[12].

Baik itu **Swakarsa**, **Swadaya** dan **Swasembada** adalah prinsip kegiatan dari usaha yang merupakan ciri khas lembaga ekonomi tersebut. Prinsip swadaya usaha berarti usaha sendiri, swakarsa berarti buatan sendiri dan swasembada berarti dengan kemampuan sendiri. Berikut tabel Data Usaha Kecil dan Mikri dair BPS Pematangsiantar yang digunakan:

Tabel 1. Data Usaha Kecil dan Mikro Pematangsiantar 2018

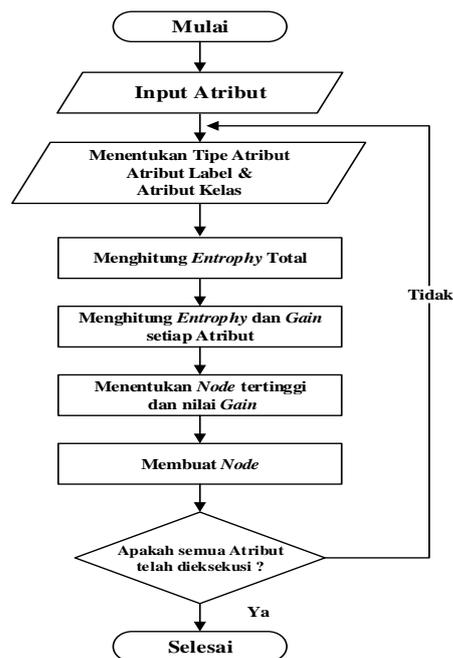
No	Nama Usaha	Modal Usaha	Laba (Keuntungan) / Bulan	Jumlah Karyawan (orang)	Bangunan	Alamat	Skala Usaha
1	ADD GRUP KONVEKSI	Rp25.000.000	Rp5.500.000	4	Miliki Sendiri	Siantar Barat	Mikro
2	AJ CATERING	Rp5.000.000	Rp900.000	5	Miliki Sendiri	Siantar Utara	Kecil
3	AJS ENTERPRISE	Rp11.000.000	Rp2.500.000	5	Sewa	Siantar Timur	Mikro
4	AKSESORIS	Rp25.000.000	Rp5.500.000	3	Sewa	Siantar Marihat	Kecil
5	AYAM GORENG	Rp30.000.000	Rp4.500.000	4	Miliki Sendiri	Siantar Martoba	Menengah
.
.
48	UD ALDI	Rp15.000.000	Rp5.500.000	15	Miliki Sendiri	Siantar Sitalasari	Menengah
49	UD ALEX'S	Rp2.500.000	Rp1.000.000	4	Miliki Sendiri	Siantar Marimbun	Mikro
50	UD REZEKI PUTRO SEKAWAN	Rp250.000.000	Rp9.500.000	8	Miliki Sendiri	Siantar Barat	Mikro
51	UD SEJAHTRA	Rp50.000.000	Rp12.000.000	4	Miliki Sendiri	Siantar Selatan	Menengah
52	WARUNG PECEL DINDA	Rp150.000.000	Rp15.000.000	5	Miliki Sendiri	Siantar Martoba	Kecil

Tabel 2. Rekap Data Usaha Kecil dan Mikro Pematangsiantar

No	Kecamatan / District	Tahun 2020												2020
		Januar i	Februar i	Mare t	Apri l	Me i	Jun i	Jul i	Agustu s	Septembe r	Oktobe r	Novembe r	Desembe r	
1	Siantar Marihat	1	0	2	1	0	1	0	1	0	0	1	0	7
2	Siantar Marimbun	1	0	0	1	0	0	2	1	1	0	0	0	6
3	Siantar Selatan	0	1	0	0	1	0	0	0	1	1	1	1	6
4	Siantar Barat	1	1	1	1	0	0	0	1	0	1	1	1	8
5	Siantar Utara	1	1	2	2	0	1	0	0	0	0	0	0	7
6	Siantar Timur	0	1	0	1	1	0	1	0	1	0	1	1	7
7	Siantar Martoba	1	0	1	0	1	0	2	0	0	2	0	0	7
8	Siantar Sitalasari	1	1	1	0	0	2	0	0	0	0	0	0	5
	Jumlah / Total	6	5	7	6	3	4	5	3	3	4	4	3	53

2.3 Pemodelan Metode

Dalam penelitian ini penulis menggunakan teknik klasifikasi Pada Metode *K-Means* dan melakukan pengujian di *RapidMiner*. Berikut ini pemodelan yang penulis gunakan dalam penelitian ini[13]. Berikut ini pemodelan Metode *K-Means* yang digambarkan dalam *flowchart* pada Gambar 2[14].

Gambar 2. Pemodelan Metode *K-Means*

Berdasarkan pemodelan gambar 2. dapat diketahui alur Pemodelan *K-Means* yang digunakan. Pada persiapan awal ditentukan penginputan atribut yang digunakan kemudian melakukan penentuan tipe atribut dengan mencari atribut label dan atribut kelas berdasarkan perhitungan *entropy* total dari masing-masing atribut. Apabila ditemukan gain disetiap atribut maka gain tersebut akan menjadi node tertinggi dengan nilai gain. Selanjutnya dilakukan penentuan cabang dengan cara yang sama dengan melihat gain tertinggi dari tiap hasil partisi[15].

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Pada bagian ini menjelaskan tentang hasil yang diperoleh dalam penelitian yang sudah dilakukan oleh penulis. Dalam proses penelitian penulis menggunakan data yang diolah dari BPS Pematangsiantar tahun 2020 dengan jumlah 8 Kecamatan sebagai bahan untuk menyelesaikan penelitian. Untuk mendukung penelitian penulis menggunakan bantuan *software Rapidminer 8.1* dalam pengimplementasian data agar mendapatkan hasil yang sesuai dari perhitungan manual dengan aplikasi. Data UMKM yang diolah dari BPS Pematangsiantar tahun 2020 dengan jumlah 8 Kecamatan akan di ubah ke dalam format data excel 2010 kemudian data dimasukkan kedalam rumus algoritma *k-means* untuk mendapatkan *cluster* tertinggi, Sedang dan terendah.

3.1.1 Perhitungan Menggunakan Algoritma *K-Means*

a) Menentukan Data Yang Akan Diolah

Data mentah yang digunakan adalah Data UMKM yang diolah dari BPS Pematangsiantar tahun 2020. Data diperoleh dari dari BPS Pematangsiantar. Berikut data yang akan digunakan untuk pengolahan metode *K-Means* :

Tabel 1. Data Hasil Kuesioner

No	Nama Usaha	Modal Usaha	Laba (Keuntungan) / Bulan	Jumlah Karyawan (orang)	Bangunan	Skala Usaha	Alamat
1	Add Grup Konveksi	Rp25.000.000	Rp5.500.000	4	2	1	Siantar Barat
2	Aj Catering	Rp5.000.000	Rp900.000	5	2	2	Siantar Utara
3	Ajs Enterprise	Rp11.000.000	Rp2.500.000	5	1	1	Siantar Timur
4	Aksesoris	Rp25.000.000	Rp5.500.000	3	1	2	Siantar Marihat
5	Ayam Goreng	Rp30.000.000	Rp4.500.000	4	2	3	Siantar Martoba
.
.
48	Ud Aldi	Rp15.000.000	Rp5.500.000	15	2	3	Siantar Sitalsari
49	Ud Alex's	Rp2.500.000	Rp1.000.000	4	2	1	Siantar Marimbun

No	Nama Usaha	Modal Usaha	Laba (Keuntungan) / Bulan	Jumlah Karyawan (orang)	Bangunan	Skala Usaha	Alamat
50	Ud Rezeki Putro Sekawan	Rp250.000.000	Rp9.500.000	8	2	1	Siantar Barat
51	Ud Sejahtra	Rp50.000.000	Rp12.000.000	4	2	3	Siantar Selatan
52	Warung Pecel Dinda	Rp150.000.000	Rp15.000.000	5	2	2	Siantar Martoba

b) Menentukan Jumlah *Cluster*

Jumlah *Cluster* yang digunakan pada data UMKM Sebanyak 3 *Cluster*. *Cluster* tersebut diantaranya Ditolak dari Target (*C1*), Diterima dengan pertimbangan (*C2*) dan Deterima (*C3*) berdasarkan data dari BPS Kota Pematangsiantar.

c) Tentukan *Centroid*

Menentukan pusat awal *Cluster* (*Centroid*) ditentukan secara manual atau *acak* yang diambil dari data yang ada dalam *range*. Nilai *Cluster* 0 diambil dari data dari data rata-rata atau nilai tengah, Nilai *Cluster* 1 diambil dari data rendah dan Nilai *Cluster* 2 diambil dari data yang paling tertinggi.

Tabel 2. *Centroid*

Centroid	Modal Usaha	Laba (Keuntungan) / Bulan	Jumlah Karyawan (orang)	Bangunan	Skala Usaha
Di ambil data ke-28 sebagai pusat cluster ke-1	300000000	4500000	5	2	1
Di ambil data ke-19 sebagai pusat cluster ke-2	45000000	21000000	6	2	1
Di ambil data ke-32 sebagai pusat cluster ke-3	1500000	500000	3	2	2

d) Menghitung Jarak dari *Centroid*

Untuk menghitung jarak antara titik *Centroid* dengan titik tiap objek menggunakan *Euclidian Distance*. Rumus untuk menghitung jarak dari *Centroid* adalah :

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

Keterangan :

D = Column kluster;

I = Menunjukkan pada data kolumn pertama;

f = Menunjukkan sebagai cluste;r

X_i = Data pada kolum data pertama;

X_j = Menenjukan pada Cluster berapa yang akan di hitung.

Maka penulis menampilkan 5 contoh perhitungan untuk jarak dari *Centroid* ke-1 adalah sebagai berikut :

$$D_{x1,c1} = \sqrt{(25.000.000 - 300000000)^2 + (5.500.000 - 4500000)^2 + (4 - 5)^2 + (2 - 2)^2 + (1 - 1)^2} = 275001818,1758$$

$$D_{x2,c1} = \sqrt{(5000000 - 300000000)^2 + (900000 - 4500000)^2 + (5 - 5)^2 + (2 - 2)^2 + (2 - 1)^2} = 295021966,2839$$

$$D_{x3,c1} = \sqrt{(11000000 - 300000000)^2 + (2500000 - 4500000)^2 + (5 - 5)^2 + (1 - 2)^2 + (1 - 1)^2} = 289006920,3323$$

$$D_{x4,c1} = \sqrt{(25000000 - 300000000)^2 + (5500000 - 4500000)^2 + (3 - 5)^2 + (1 - 2)^2 + (2 - 1)^2} = 275001819,1758$$

$$D_{x5,c1} = \sqrt{(30000000 - 300000000)^2 + (4500000 - 4500000)^2 + (4 - 5)^2 + (2 - 2)^2 + (3 - 1)^2} = 270000004$$

Dan seterusnya sampai dengan $D_{x12,c1}$.

Hasil akhirnya jika centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut :

$$D_{x1,c2} = \sqrt{(25.000.000 - 45000000)^2 + (5.500.000 - 21000000)^2 + (4 - 6)^2 + (2 - 2)^2 + (1 - 1)^2} = 25303161,8577$$

$$D_{x2,c2} = \sqrt{(5000000 - 45000000)^2 + (900000 - 21000000)^2 + (5 - 6)^2 + (2 - 2)^2 + (2 - 1)^2} = 44766171,2628$$

$$D_{x3,c2} = \sqrt{(11000000 - 45000000)^2 + (2500000 - 21000000)^2 + (5 - 6)^2 + (1 - 2)^2 + (1 - 1)^2} = 38707234,4659$$

$$D_{x4,c2} = \sqrt{(25000000 - 45000000)^2 + (5500000 - 21000000)^2 + (3 - 6)^2 + (1 - 2)^2 + (2 - 1)^2} = 25303162,8577$$

$$D_{x5,c2} = \sqrt{(30000000 - 45000000)^2 + (4500000 - 21000000)^2 + (4 - 6)^2 + (2 - 2)^2 + (3 - 1)^2} = 22299107,1209$$

Dan seterusnya sampai dengan $D_{x12,c2}$.

Hasil akhirnya jika centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-3 adalah sebagai berikut :

$$D_{x1,c3} = \sqrt{(25.000.000 - 1500000)^2 + (5.500.000 - 500000)^2 + (4 - 3)^2 + (2 - 2)^2 + (1 - 2)^2} = 24026028,5534$$

$$D_{x2,c3} = \sqrt{(5000000 - 1500000)^2 + (900000 - 500000)^2 + (5 - 3)^2 + (2 - 2)^2 + (2 - 2)^2} = 3522782,9907$$

$$D_{x3,c3} = \sqrt{(11000000 - 1500000)^2 + (2500000 - 500000)^2 + (5 - 3)^2 + (1 - 2)^2 + (1 - 2)^2} = 9708244,9194$$

$$D_{x4,c3} = \sqrt{(25000000 - 1500000)^2 + (5500000 - 500000)^2 + (3 - 3)^2 + (1 - 2)^2 + (2 - 2)^2} = 24026027,5534$$

$$D_{x5,c3} = \sqrt{(30000000 - 1500000)^2 + (4500000 - 500000)^2 + (4 - 3)^2 + (2 - 2)^2 + (3 - 2)^2} = 28779333,8623$$

Dan seterusnya sampai dengan $D_{x22,c3}$.

Hasil akhirnya Untuk pengulangan berikutnya (pengulangan ke-1 sampai selesai), centroid baru dihitung dengan **menghitung nilai rata-rata data pada setiap cluster**. Jika centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai.

Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari ketiga *Centroid*. Tabel Jarak dari *Centroid* adalah sebagai berikut :

Tabel 3. Jarak *Centroid* Iterasi ke-1

	C1	C2	C3	Jarak Centroid
275001818,1758	25303161,8578	24026028,5535	C3	
295021966,2839	44766171,2628	3522782,9908	C3	
289006920,3324	38707234,4659	9708244,9195	C3	
275001819,1758	25303162,8578	24026027,5535	C3	
270000004,0000	22299107,1210	28779333,8623	C2	
285001758,3806	33767592,0098	14396181,0489	C3	
.	.	.	.	
.	.	.	.	
297520587,5229	46970735,5701	1118034,9888	C3	
50249378,1056	205322307,6044	248662925,4580	C1	
250112478,6989	10295634,1410	49844760,0023	C2	
150367051,9121	105171289,8577	149206233,1138	C2	

e) Menentukan *Cluster* atau Pengelompokan

Dalam menentukan *Cluster* dengan mencari nilai *Cluster* berdasarkan nilai minimal dari nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai minimal pada Iterasi 1. Berikut tabel *Cluster* pada Iterasi 1 sebagai berikut :

Tabel 4. *Cluster* Iterasi ke-1

No.	C1	C2	C3
1			1
2			1
3			1
4			1
5		1	
.	.	.	.
.	.	.	.
.	.	.	.
47		1	
48			1
49			1
50	1		
51		1	
52		1	
53			1

No.	C1	C2	C3
Jumlah	8	17	28

Selanjutnya dalam metode *K-Means*, perhitungan berhenti apabila *Cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *Cluster* pada iterasi selanjutnya sampai nilai Iterasinya sama. Untuk mencari nilai *Centroid* selanjutnya dengan menggunakan *Centroid* baru pada Iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada *Cluster* di tabel diatas. Adapun *Centroid* baru untuk mencari *Cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *Cluster* tersebut kemudian memmbagikannya sebanyak jumlah nilai sebagai berikut :

$$C_{1,Modal Usaha} = (250000000 + 250000000 + 250000000 + 250000000 + 300000000 + 250000000 + 250000000 + 250000000)/8 = 256250000$$

$$C_{2,Modal Usaha} = (30000000 + 45000000 + 50000000 + 150000000 + 150000000 + 50000000 + 45000000 + 160000000 + 50000000 + 150000000 + 50000000 + 30000000 + 45000000 + 50000000 + 150000000 + 50000000 + 150000000)/17 = 82647058,8235$$

$$C_{3,Modal Usaha} = (25000000 + 5000000 + 11000000 + 25000000 + 1500000 + 2500000 + 5000000 + 4500000 + 8000000 + 2500000 + 500000 + 2500000 + 5000000 + 11000000 + 25000000 + 4500000 + 15000000 + 1500000 + 2500000 + 5000000 + 25000000 + 25000000 + 5000000 + 11000000 + 2500000 + 15000000 + 2500000 + 25000000)/28 = 11375000$$

Dst hingga C5.

Berikut, data *Centroid* baru Iterasi ke-1 sebagai berikut :

Tabel 5. *Centroid* Baru Iterasi Ke-1

Centroid	Modal Usaha	Laba (Keuntungan) / Bulan	Jumlah Karyawan (orang)	Bangunan	Skala Usaha
C1	256250000,0000	11125000,0000	7,1250	1,8750	1,7500
C2	82647058,8235	8941176,4706	6,0000	1,8824	1,8824
C3	11375000,0000	3289285,7143	5,9643	1,8214	1,9286

Hingga *Cluster* Iterasi ke-3 dan tabel *Cluster* Iterasi ke-4 memiliki nilai *Cluster* yang sama atau tidak berubah pada *Cluster* Iterasi ke-4 maka perhitungan dihentikan dan hasil yang diperoleh yaitu :

- 1) *Cluster* 2 (C1) memiliki 8 data yang diartikan bahwa kelompok pertama adalah kategori yang UMKM dengan Grade tertinggi dengan masuk kelas Menengah pada tahun 2021.
- 2) *Cluster* 1 (C2) memiliki 6 data yang diartikan bahwa kelompok kedua adalah kategori yang UMKM dengan Grade sedang dengan masuk kelas Kecil pada tahun 2021.
- 3) *Cluster* 0 (C3) memiliki 39 data yang diartikan bahwa kelompok ketiga adalah kategori yang UMKM dengan Grade rendah dengan masuk kelas mikro pada tahun 2021.

Untuk proses mengelompokkan data penulis menggunakan rumus pada metode *k-means*. Kemudian data pegawai penerima insentif tersebut akan diproses menggunakan aplikasi *Rapidminer* 8.1 untuk mendapatkan hasil yang akurat. Data UMKM dari BPS Pematangsiantar tahun 2020 dengan jumlah 8 Kecamatan.

3.1.2 Perhitungan Menggunakan Rapidminer

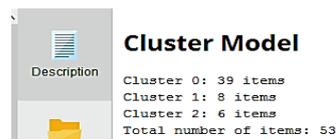
Tahapan pertama untuk menjalankan proses pembentukan permodelan yaitu melakukan *importing* data dari data yang sebelumnya ditransformasikan ke *microsoft excel* dengan format *.xlsx* agar dapat di masukkan ke aplikasi *Rapidminer*. Berikut adalah gambaran dari *importing* data :

Langkah berikutnya *drag* dan *drop* data yang usdah diimport sebelumnya, lalu hubungkan pada *operator K-means*. Langkah berikutnya *drag* dan *drop* operator *k-means* lalu hubungkan pada *output*. Untuk memastikan datanya sudah benar sesuai dengan type datanya, jika type data tidak sesuai maka operator tidak akan melakukan proses run.



Gambar 3. Drag dan Drop Operator *K-Means*

Jika pada operator *k-means* dan data read excel tidak terjadi error akibat kesalahan data ataupun type data maka akan muncul tampilan seperti gambar 3. Sambungan proses antara data read excel dengan operator *k-means* akan berbentuk warna hijau mudah seperti tertera pada gambar 3. Pada saat *tool run* diklik akan muncul pohon keputusan seperti gambar dibawah. Setelah itu menentukan jumlah *Cluster* yang digunakan pada data akta kelahiran sebanyak 3 *Cluster*. *Cluster* tersebut diantaranya dengan jenis UMKM probilitas tinggi dengan kategori Menengah (C1), dengan jenis UMKM probilitas sedang dengan kategori kecil (C2) dan dengan jenis UMKM probilitas rendah dengan kategori kecil (C3) berdasarkan data UMKM dari BPS Pematangsiantar tahun 2020 dengan jumlah 8 Kecamatan.



Gambar 4. Data Hasil *Clustering*

Sehingga berdasarkan data hasil *clustering* pada gambar 4. didapatkan cluster hasil dari rapidminer 8.1 berikut ini.

- a) *Cluster 0 (C3)* memiliki 39 data yang diartikan bahwa kelompok pertama adalah kategori yang UMKM dengan Grade rendah dengan masuk kelas mikro pada tahun 2021.
- b) *Cluster 1 (C2)* memiliki 6 data yang diartikan bahwa kelompok kedua adalah kategori yang UMKM dengan Grade sedang dengan masuk kelas Kecil pada tahun 2021.
- c) *Cluster 2 (C1)* memiliki 8 data yang diartikan bahwa kelompok ketiga adalah kategori yang UMKM dengan Grade tinggi dengan masuk kelas menengah pada tahun 2021.

3.2 Pembahasan

Setelah dilakukan proses perhitungan *cluster* dengan bantuan aplikasi *rapidminer*, diperoleh hasil data pegawai yang termasuk dalam *cluster* tertinggi (*cluster 1*) terdapat 8 kelompok UMKM. dan *cluster* sedang (*cluster 2*) terdapat 6 kelompok UMKM. Sedangkan untuk *cluster* rendah (*cluster 0*) terdapat 39 kelompok UMKM.. Hasil yang didapat dari perhitungan manual dengan implementasi di aplikasi *rapidminer* sesuai atau sama (*balance*).

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, penulis menarik beberapa kesimpulan diantaranya (*C1*) *Cluster* Tertinggi dengan jumlah 8 kelompok UMKM yaitu, memiliki 8 data yang diartikan bahwa kelompok ketiga adalah kategori yang UMKM dengan Grade tinggi dengan masuk kelas menengah pada tahun 2021. (*C2*) *Cluster* Sedang dengan jumlah 6 kelompok UMKM, memiliki 6 data yang diartikan bahwa kelompok kedua adalah kategori yang UMKM dengan Grade sedang dengan masuk kelas Kecil pada tahun 2021. (*C3*) *Cluster* Terendah dengan jumlah 39 kelompok UMKM yaitu, memiliki 39 data yang diartikan bahwa kelompok pertama adalah kategori yang UMKM dengan Grade rendah dengan masuk kelas mikro pada tahun 2021.

REFERENCES

- [1] M. R. Ridlo, S. Defiyanti, A. Primajaya, M. Rosyid Ridlo, S. Defiyanti, and A. Primajaya, "Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang," *Citee 2017*, pp. 426-433, 2017.
- [2] J. T. Informatika-s and F. I. Komputer, "Clusterisasi Data Pasien Menggunakan K-Means Clustering Tri Hertanto," no. 5.
- [3] F. Nurzaman, "Penerapan Algoritma K-Means Dalam Pengelompokan Lokasi Rumah Sakit Provider Pada Asuransi Kesehatan," pp. 61-67, 2018.
- [4] L. Felicia, "Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi Di Kota Semarang," pp. 1-5, 2014.
- [5] D. Mining, "Belajar Mudah Algoritma Data Mining Clustering : k-means," pp. 2-6.
- [6] S. Harlina, "DATA MINING PADA PENENTUAN KELAYAKAN KREDIT MENGGUNAKAN ALGORITMA K-NN BERBASIS FORWARD SELECTION DATA MINING ON CREDIT FEASIBILITY DETERMINATION USING K-NN ALGORITHM BASED ON FORWARD SELECTION," vol. 11, no. 2, pp. 236-244, 2018.
- [7] P. Rapid and M. Studio, "Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," vol. IV, no. 2, pp. 401-411, 2018.
- [8] P. Alkhairi, I. S. Damanik, and A. P. Windarto, "Penerapan Jaringan Saraf Tiruan untuk Mengukur Korelasi Beban Kerja Dosen Terhadap Peningkatan Jumlah Publikasi," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 581, 2019, doi: 10.30645/senaris.v1i0.65.
- [9] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *J. KomTekInfo Fak. Ilmu Komput.*, vol. 2, no. 2, pp. 33-38, 2015.
- [10] T. I. S and F. I. Komputer, "Analisis Data Rawat Inap Rumah Sakit Kota Semarang Untuk Mengetahui Daerah Endemi Penyakit Menggunakan Algoritma K-Means," 2018.
- [11] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, p. 108, 2018, doi: 10.31154/cogito.v2i2.19.108-119.
- [12] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, p. 22, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.
- [13] W. A. Triyanto, "ALGORITMA K-MEDOIDS UNTUK PENENTUAN STRATEGI PEMASARAN PRODUK," *J. SIMETRIS*, vol. 6, no. 1, pp. 183-188, 2015.
- [14] A. ACiputra, D. R. I. M. Setiadi, E. H. Rachmawanto, and A. Susanto, "Klasifikasi Tingkat Kematangan Buah Apel Manalagi Dengan Algoritma Naive Bayes Dan Ekstraksi Fitur Citra Digital," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 465-472, 2018, doi: 10.24176/simet.v9i1.2000.
- [15] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," *Pseudocode*, vol. 4, no. 2, pp. 120-128, 2017, doi: 10.33369/pseudocode.4.2.120-128.