

**PENGELOMPOKKAN FAKTOR-FAKTOR YANG MENJADI PENYEBAB
KEBAKARAN HUTAN MENGGUNAKAN ALGORITMA K-MEAN CLUSTERING**



**UNIVERSITAS
TEKNOLOGI
SUMBAWA**

Disusun oleh:

- 1. FAUZAN RUSLAN (20.01.013.055)**
- 2. MUHAMMAD FAHROJI (20.01.013.010)**
- 3. MAGHFIRA MEILANI PUTRI (20.01.013.008)**
- 4. VIRNA FEBRI ANDINI (20.01.013.017)**
- 5. WULANDARI (20.01.013.019)**

**FAKULTAS REKAYASA SISTEM
UNIVERSITAS TEKNOLOGI SUMBAWA
SUMBAWA**

2022

ABSTRAK

Titik api di tiap daerah hampir setiap tahunnya ada dan menimbulkan dampak kebakaran hutan. Salah satu penyebabnya adalah maraknya pembukaan lahan baru. Perlu adanya penanggulangan secepat mungkin untuk mengatasi masalah ini karena kebakaran hutan dapat mengakibatkan polusi udara dalam jangka pendek dan pemanasan global dalam jangka panjang. Sehingga masalah-masalah tersebut perlu diwaspadai dengan cepat salah satunya dengan mengetahui terlebih dahulu faktor yang kemungkinan akan muncul berdasarkan data-data yang ada sebelumnya. Pengolahan data mining sangat diperlukan guna menghasilkan data yang relevan untuk mengetahui kemungkinan terjadinya kebakaran hutan. Diharapkan pada hasil clustering yang menggunakan algoritma K-Means dapat membantu pihak terkait dalam memprediksi lebih dini kebakaran hutan yang kemungkinan akan muncul. Menjawab permasalahan diatas maka dibutuhkan sebuah sistem yang dapat membantu mengoptimalkan dalam proses penentuan clustering kebakaran hutan dengan menggunakan algoritma K-means Clustering.

Kata kunci: Clustering, Data Mining, K-Means, Kebakaran Hutan.

KATA PENGANTAR

Alhamdulillah, puji syukur kepada Allah SWT, karena berkat limpahan rahmat dan karunia-Nya penulis mampu menyelesaikan tugas akhir yang berjudul “pengelompokan faktor faktor yang menjadi penyebab kebakaran hutan menggunakan Algoritma k-mean clustering”. Tugas akhir ini disusun sebagai salah satu syarat untuk memenuhi mata kuliah kecerdasan buatan . Selama penyusunan tugas akhir ini, penulis mendapatkan banyak pengetahuan, bimbingan, dukungan, arahan, serta masukan yang menuju kearah kebaikan dari semua pihak sehingga penulisan tugas akhir ini dapat diselesaikan. Untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih, semoga laporan yang kami buat ini bisa bermanfaat kedepannya.

Sumbawa, 11 januari 2022

Penulis

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
BAB I.....	1
1.1 LATAR BELAKANG	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Pembuatan Makalah.....	2
BAB II.....	3
2.1 Kebakaran Lahan Dan Hutan	3
2.2 Pengelompokan (<i>clustering</i>)	3
2.3 Definisi Data Mining	4
2.4 Algoritma K-means.....	5
2.5 CRISP-DM.....	6
2.6 Rapid Minner	8
2.7 Data set.....	9
BAB III	10
3.1 Metode Penelitian	10
3.2 Alat dan Bahan penelitian.....	11
3.3 waktu dan tempat penelitian	11
BAB IV	12
4.1 Implementasi Sistem	12
4.2 Hasil dari clustering data.....	14
BAB V	15
5.1 Kesimpulan	15
5.2 Saran	15
DAFTAR PUSTAKA	16

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Perkembangan teknologi merupakan salah satu yang menjadi trend saat ini. Sehingga mendorong banyaknya data yang dihasilkan dari berbagai bidang, misalnya bidang kedokteran, pendidikan sampai bidang pertanian. Banyaknya data yang dihasilkan menjadikan Data Mining ikut ambil bagian dalam perkembangan zaman. Setiap hitungan detik, data-data dalam berbagai *database* akan senantiasa terbarukan, baik dikarenakan adanya update data maupun karena adanya penambahan data baru. Hal ini memicu permasalahan penting bagaimana caranya mendapatkan pengetahuan yang memberikan informasi bermakna dan berguna dari kumpulan data-data tersebut. Dalam kondisi semacam ini lahirlah sebuah metode Data Mining yang merupakan suatu metode penambangan data yang dapat memberikan suatu cara yang efektif untuk memproses data lebih lanjut.

Data mining dapat dipergunakan untuk membantu memecahkan masalah yang selalu dihadapi di hutan hutan antara lain masalah kebakaran hutan yang sangat marak akhir-akhir ini. Kebakaran hutan yang hampir setiap tahunnya ada, Penyebab kebakaran hutan antara lain pembukaan lahan yang sering terjadi sebagai pemicu dari kebakaran hutan dan faktor alam yang terjadi akibat terik matahari atau terkena sambaran petir disaat musim hujan. Perlu adanya penanggulangan secepat mungkin untuk mengatasi munculnya masalah kebakaran hutan. Maka sangat penting untuk mengetahui faktor faktor yang rawan menyebabkan kebakaran hutan sejak dini. Hasil yang ditimbulkan dari kebakaran hutan yang dapat mengakibatkan polusi udara dalam jangka pendek dan pemanasan global dalam jangka panjang.

Masalah-masalah diatas perlu diwaspadai dengan cepat salah satunya adalah dengan mengetahui terlebih dahulu kebakaran hutan yang kemungkinan akan muncul berdasarkan data-data yang ada sebelumnya. Salah satu upaya pencegahan kebakaran hutan yakni dengan mengetahui faktor terbesar yang menyebabkan kebakaran hutan serta

persebaran pengelompokan titik api yang memiliki potensi tinggi terhadap terjadinya kebakaran hutan. Pengolahan data mining sangat diperlukan guna menghasilkan data yang relevan untuk mengetahui penyebabnya. Dengan hasil visualisasi pemetaan yang menggunakan metode K-means diharapkan dapat membantu pihak terkait dalam menangani lebih dini yang kemungkinan akan munculnya kebakaran hutan.

Berdasarkan permasalahan yang sudah diuraikan maka pada penelitian ini dibutuhkan sebuah sistem yang dapat membantu mengoptimalkan dalam proses pemetaan titik api dan penyebab kebakaran hutan dengan menggunakan algoritma K-means clustering. Salah satu teknik yang dikenal dalam Data Mining yaitu Clustering, membagi data kedalam grup- grup yang mempunyai obyek yang karakteristiknya sama.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas sebelumnya, maka permasalahan yang akan dikaji dalam penelitian ini dapat dirumuskan menjadi suatu masalah yang dapat dijadikan bahan penelitian “Bagaimana mencari Cluster faktor terbesar penyebab kebakaran hutan?, menggunakan Algoritma *K-means* Clustering”

1.3 Batasan Masalah

Agar pembahasan tidak menyimpang dari ruang lingkup pembahasan, maka perlu dibuat suatu batasan masalah sebagai berikut :

1. Penelitian ini khusus membahas tentang faktor penyebab kebakaran hutan dan lahan
2. Penerapan akan menggunakan Algoritma K-Means yang diambil dari data para peneliti data mining yaitu di UCI Repository.

1.4 Tujuan Pembuatan Makalah

Tujuan pembuatan makalah ini adalah :

Tujuan yang ingin dicapai adalah menentukan cluster factor yang menyebabkan kebakaran hutan kedalam beberapa cluster menggunakan algoritma *K-Means Clustering*.

BAB II

LANDASAN TEORI

2.1 Kebakaran Lahan Dan Hutan

Pengertian Lahan dan Hutan adalah hamparan lahan yang tumbuhan yang didominasi oleh pepohon atau tumbuhan berkayu lain, tumbuh secara bersama dan cukup rapat dengan luasan tertentu. Hamparan lahan hutan biasanya lebih luas dari 0,5 Ha dengan tinggi pohon lebih 5 m, dan penutupan tajuk lebih dari 10% (Suhendang, 2013).

Kebakaran hutan merupakan masalah yang serius karena dapat mengakibatkan dampak buruk terhadap lingkungan (sonita veronica, 2014). Kebakaran lazimnya akan menyebabkan kerusakan atau kemusnahan pada binaan dan kecederaan atau kematian kepada manusia. Kebakaran kadang kala turut menyebabkan kebakaran atau kebakaran liar. Kebakaran bisa menyebabkan kecederaan atau kematian yang berpuncak pada terhiru asap ataupun melecur. Kebakaran saat kecil dibilang teman tetapi saat sudah besar menjadi musuh.

Kebakaran Hutan adalah Pembakaran yang penjalaran apinya bebas sertamengkonsumsi bahan bakar alam dari hutan seperti serasah, rumput, ranting/cabang pohon mati yang tetap berdiri, log, tunggak pohon, gulma, semakbelukar, dedaunan, pohon-pohon dan dapat menimbulkan kerugian ekonomis (Suharjo, 2003). Dampak dari kebakaran hutan bisa bernilai positif dan negatif.

2.2 Pengelompokan (*clustering*)

Pengelompokan merupakan mengelompokkan *records*, pengamatan, atau memperhatikan dan membentuk kelas-kelas objek-objek yang memiliki sebuah kemiripan. Kluster adalah kumpulan *records* yang memiliki kemiripan satu dengan yang lainnya tetapi tidak memiliki kemiripan dengan *records* dalam cluster yang lain. Berbeda dengan klasifikasi, kluster tidak memiliki variable target. Pengklasteran tidak melakukan klasifikasi, estimasi, atau peridiksi nilai dari variable target, akan tetapi algoritma kluster tersebut yang akan melakukan pembagian terhadap keseluruhan data menjadi kelompok kelompok yang memiliki kemiripan. Dimana kemiripan *records* dalam satu kelompok

akan bernilai maksimal sedangkan kemiripan dengan *records* yang lain akan bernilai minimum.

2.3 Definisi Data Mining

Data mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran data mining ini bisa dipakai untuk membantu pengambilan keputusan dimasa depan. Pengembangan KDD ini menyebabkan penggunaan *pattern recognition* semakin berkurang karena telah menjadi bagian data mining (Mujiasih, 2011).

Metode ini merupakan gabungan 4 (empat) disiplin ilmu yakni *statistic*, *visualisasi*, *database*, dan *machine learning*. Adapun *machine learning* adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik pemrograman berdasarkan pembelajaran masa lalu dan bersinggungan dengan ilmu statistika dan juga optimasi.

2.3.1 Tahapan Data Mining

Tahapan dalam Data Mining Menurut (Buulolo, 2013) ada empat tahap yang dilalui dalam data mining, antara lain:

1. Tahap pertama: pernyataan yang tepat terhadap permasalahan sebelum mengakses perangkat lunak data mining, jika tidak ada masalah dalam penelitian tersebut maka solusi yang dihasilkan menjadi sia-sia.
2. Tahap kedua: tahap ini merupakan tahap mempersiapkan data yang termasuk dalam data mining dengan melakukan *cleaning*, transformasi data, memilih subset record, data set, langkah awal seleksi. Mendeskripsikan dan memvisualisasikan data adalah kunci dari tahap ini.
3. Tahap ketiga: tahap ini melibatkan pertimbangan terhadap ragam permodelan dan memilih yang terbaik bagi performansi prediktif.
4. Tahap keempat: tahap ini memilih aplikasi yang tepat serta melakukan permodelan untuk membuat (*generate*) prediksi.

2.4 Algoritma K-means

K-means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan supervised learning yang menerima masukan berupa vektor $(-x-1, y1), (-x-2, y2), \dots, (-x-i, yi)$, di mana x_i merupakan data dari suatu data pelatihan dan merupakan label kelas untuk x_i .

Pada algoritma pembelajaran ini, computer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam unsur pervised learning. Masukan yang diterima adalah data atau objek dan buah kelompok (cluster) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam ke buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut.

Data clustering menggunakan metode K-Means ini secara umum dilakukan dengan algoritma dasar sebagai berikut (Yudi Agusta, 2007) :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/ rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/ rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai

2.4.1 Kekurangan dan Kelebihan K-Means

Ada beberapa kelebihan pada algoritma k-means, yaitu:

1. Mudah untuk diimplementasikan dan dijalankan.
2. Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat.
3. Mudah untuk diadaptasi.

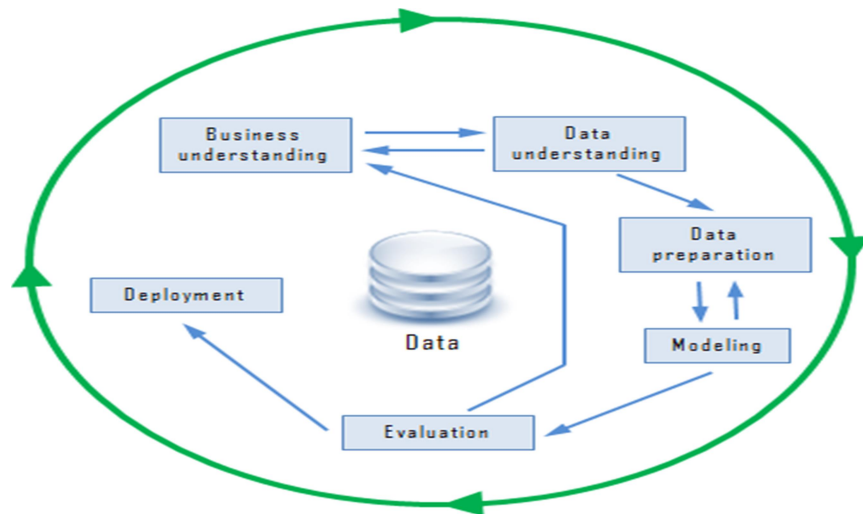
4. Umum digunakan.

Algoritma k-means memiliki beberapa kelebihan, namun ada kekurangannya juga. Kekurangan dari algoritma tersebut yaitu :

1. Sebelum algoritma dijalankan, k buah titik diinisialisasi secara random sehingga pengelompokkan data yang dihasilkan dapat berbeda-beda. Jika nilai random untuk inisialisasi kurang baik, maka pengelompokkan yang dihasilkan pun menjadi kurang optimal.
2. Dapat terjebak dalam masalah yang disebut curse of dimensionality. Hal ini dapat terjadi jika data pelatihan memiliki dimensi yang sangat tinggi (Contoh jika data pelatihan terdiri dari 2 atribut maka dimensinya adalah 2 dimensi. Namun jika ada 20 atribut, maka akan ada 20 dimensi). Salah satu cara kerja algoritma ini adalah mencari jarak terdekat antara k buah titik dengan titik lainnya. Jika mencari jarak antar titik pada 2 dimensi, masih mudah dilakukan. Namun bagaimana mencari jarak antar titik jika terdapat 20 dimensi. Hal ini akan menjadi sulit.
3. Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan k titik yang diinisialisasi secara random. Namun jika terdapat banyak sekali titik data (misalnya satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti kD-Tree atau hashing.

2.5 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan suatu standar yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis dari suatu industri sebagai strategi pemecahan masalah dari bisnis atau unit penelitian[5]. Untuk data yang dapat diproses dengan CRISP-DM ini, tidak ada ketentuan atau karakteristik tertentu, karena data tersebut akan diproses kembali pada fase-fase dalamnya. Terdapat enam fase dalam CRISP-DM ini yakni dijelaskan pada gambar berikut :



Gambar 2.5 fase dalam crisp DM

2.5.1 Fase Pemahaman Bisnis (Business Understanding Phase)

1. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
2. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
3. Menyiapkan strategi awal untuk mencapai tujuan.

2.5.2 Fase Pemahaman Data (Data Understanding Phase)

1. Mengumpulkan data, jika data berasal dari lebih dari satu database maka dilakukan proses integrasi data atau Data Integration.
2. Mengembangkan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
3. Mengevaluasi kualitas data, memeriksa data dan membersihkan data yang tidak valid atau proses Data Cleaning.
4. Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.

2.5.3 Fase Pengolahan Data (Data Preparation Phase)

1. Siapkan data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya atau proses Data Selection.
2. Pilih kasus dan variabel yang akan dianalisis, sesuai dengan analisis yang akan dilakukan.
3. Lakukan perubahan pada variabel jika diperlukan.
4. Siapkan data awal sehingga siap untuk perangkat permodelan atau Data Transformation

2.5.4 Fase Permodelan (Modelling Phase)

1. Pilih dan aplikasikan teknik permodelan yang sesuai.
2. Kalibrasi aturan model untuk mengoptimalkan hasil.
3. Dapat menggunakan beberapa teknik yang sama untuk permasalahan yang sama
4. Dapat kembali ke fase pengolahan data jika diperlukan untuk menjadikan data ke dalam bentuk kebutuhan tertentu

2.5.5. Fase Evaluasi (Evaluation Phase)

1. Mengevaluasi satu atau lebih model yang digunakan dalam fase permodelan atau proses Evaluation Pattern.
2. Menetapkan apakah model tadi sudah sesuai dengan tujuan pada fase awal.
3. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
4. Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.

2.5.6 Fase Penyebaran (Deployment Phase)

Menggunakan model yang dihasilkan dan di persentasikan atau proses knowledge persentation.

2.6 Rapid Minner

RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk

persiapan data, pembelajaran mesin, pembelajaran dalam, penambahan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi

2.7 Data set

dataset adalah sebuah kumpulan data yang berasal dari informasi-informasi pada masa lalu dan siap untuk dikelola menjadi sebuah informasi baru.

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

metode penelitian menggunakan cross-industry standard process for data mining (crisp-DM).

3.1.1 Pemahaman Bisnis (business understanding)

Seperti yang sudah dijelaskan bahwa pengimplementasian *data mining* pada aplikasi ini adalah berhubungan dengan data kebakaran hutan. Faktor-faktor apa saja yang mempengaruhi kebakaran hutan sesuai data yang sudah ada. Memeriksa keterkaitan antara tiap atribut yang mempengaruhi kebakaran hutan yang nantinya dapat dijadikan acuan untuk memprediksi faktor kebakaran hutan selanjutnya sehingga memaksimalkan pemahaman data yang ada.

3.1.2 Pemahaman Data (data understanding)

Pada tahap ini pengumpulan data terlebih dahulu dilakukan, semua data yang diperlukan nanti akan dilakukan pengolahan secara menyeluruh. Pengumpulan data berupa dengan mengambil data yang ada lewat publik data set, UCI Repository.

3.1.3 Pengolahan Data

Pada fase ini akan dilakukan proses pemilihan dan pengolahan data yang nanti akan diperlukan dalam tahap permodelan sehingga nanti permodelan yang dilakukan dapat memberikan hasil maksimal sesuai dengan target yang diinginkan.

3.1.4 Permodelan

Pemilihan Data merupakan dalam fase Pengolahan data dimana menyesuaikan dengan proses mining yang akan dilakukan.

3.2 Alat dan Bahan penelitian

3.2.1 Alat

Tool yang kami gunakan dalam penelitian ini adalah rapid minner

3.2.2 Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini merupakan data kebakaran hutan yang diambil dari Public Data Set para peneliti data mining

3.2.3 Analisa Kebutuhan Data

Tahap ini adalah tahap dimana data yang dibutuhkan dalam penelitian ini dikumpulkan untuk dapat diolah selanjutnya. Data-data yang dikumpulkan adalah data kebakaran lahan. Adapun sumber data, metode pengumpulan data dan jumlah data adalah sebagai berikut:

1. Data yang digunakan merupakan data sekunder yang diambil langsung dari UCI Repository.
2. jumlah data yang digunakan dalam penelitian ini adalah terdiri dari 100 record dengan 13 atribut.

3.3 waktu dan tempat penelitian

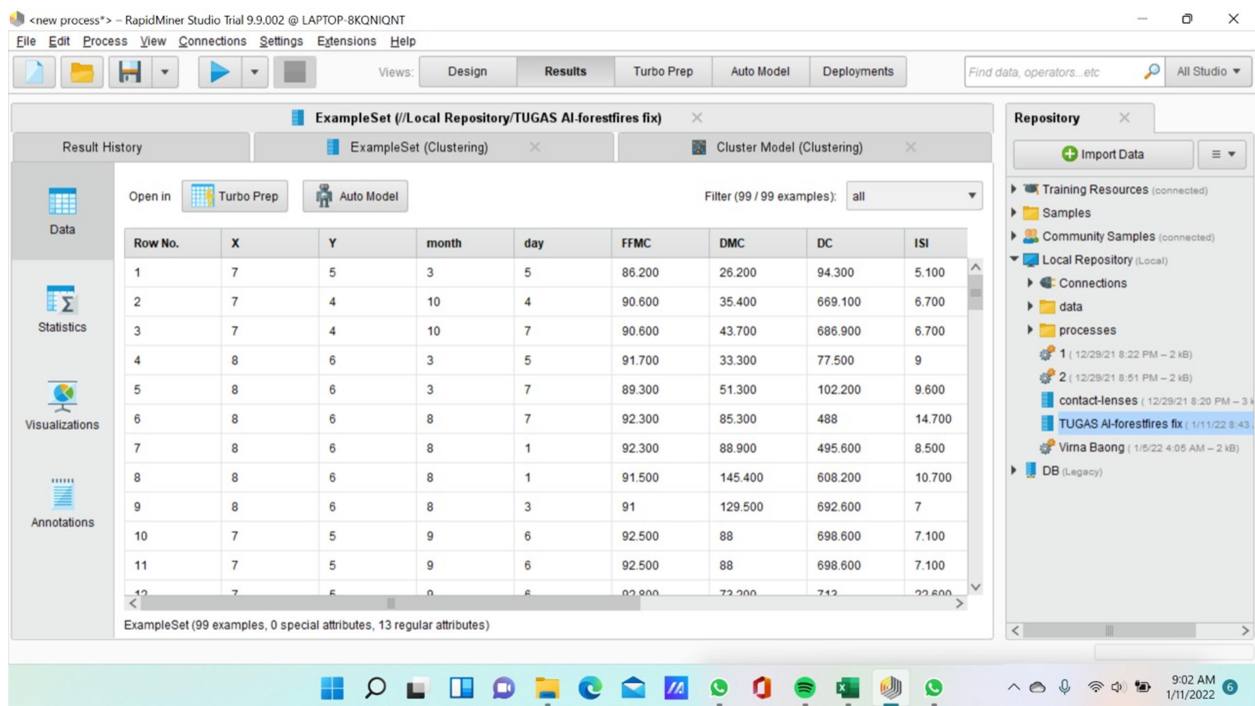
1. waktu penelitian : selasa, 11 Januari 2022
2. tempat penelitian : rumah masing masing

BAB IV HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

Pada proses peminingan, terdapat 3 form utama yakni form penyiapan data, hasil, dan *report mining*. Pada form penyiapan data dan hasil dapat dilihat dari sampel data set yang digunakan. Kemudian untuk hasil, Report yang dihasilkan berupa model data yang telah di cluster menggunakan tool rapid minner dengan spesifikasi faktor terbesar pemicu dari kebakaran hutan.

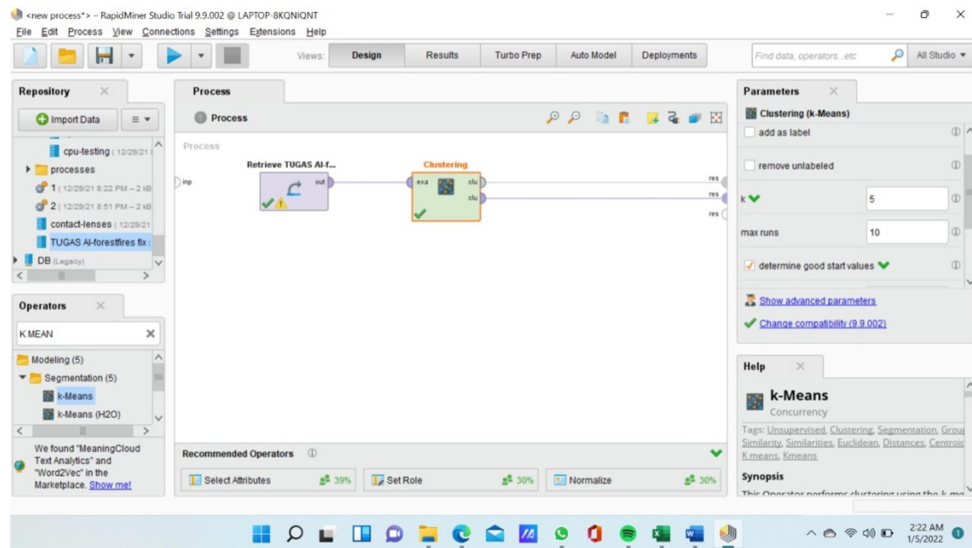
4.1.1 Berikut sampel data set yang digunakan :



Row No.	X	Y	month	day	FPMC	DMC	DC	ISI
1	7	5	3	5	86.200	26.200	94.300	5.100
2	7	4	10	4	90.600	35.400	669.100	6.700
3	7	4	10	7	90.600	43.700	686.900	6.700
4	8	6	3	5	91.700	33.300	77.500	9
5	8	6	3	7	89.300	51.300	102.200	9.600
6	8	6	8	7	92.300	85.300	488	14.700
7	8	6	8	1	92.300	88.900	495.600	8.500
8	8	6	8	1	91.500	145.400	608.200	10.700
9	8	6	8	3	91	129.500	692.600	7
10	7	5	9	6	92.500	88	698.600	7.100
11	7	5	9	6	92.500	88	698.600	7.100
12	7	5	9	6	92.500	88	698.600	7.100

Gambar 4.1 Tabel Sample Data Set

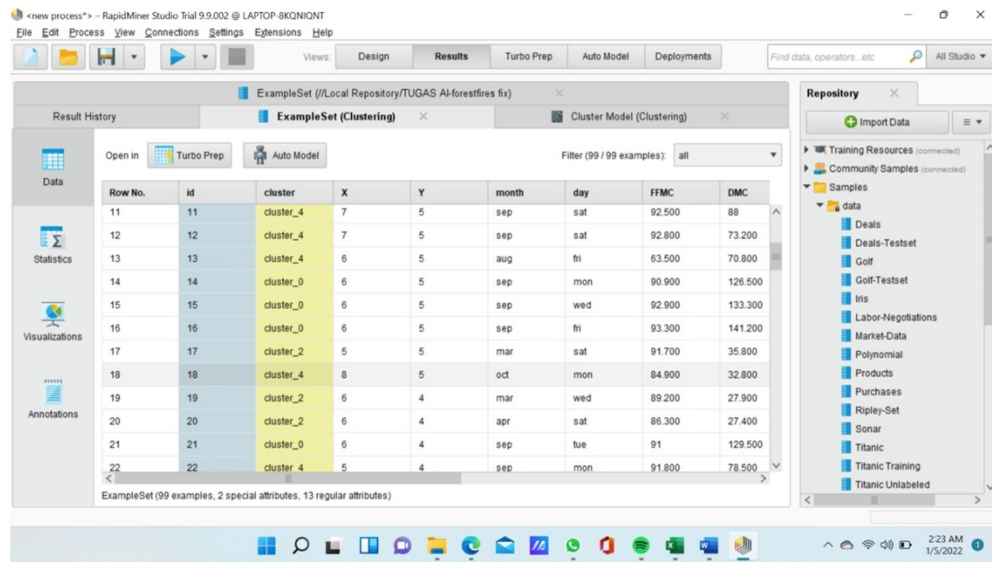
4.1.2 Pengujian Menggunakan RapidMinner



Gambar 4.1.2 Tabel Hasil Uji Data

The screenshot shows the 'Result History' panel in RapidMinner Studio. The 'ExampleSet (Clustering)' tab is selected, displaying a table of 11 rows of data. The table includes columns for Row No., Id, cluster, X, Y, month, day, FPMC, and DMC. The 'Repository' panel on the right shows a list of data sources.

Row No.	Id	cluster	X	Y	month	day	FPMC	DMC
1	1	cluster_2	7	5	mar	fri	86.200	26.200
2	2	cluster_4	7	4	oct	tue	90.600	35.400
3	3	cluster_4	7	4	oct	sat	90.600	43.700
4	4	cluster_2	8	6	mar	fri	91.700	33.300
5	5	cluster_2	8	6	mar	sun	89.300	51.300
6	6	cluster_3	8	6	aug	sun	92.300	85.300
7	7	cluster_3	8	6	aug	mon	92.300	88.900
8	8	cluster_1	8	6	aug	mon	91.500	145.400
9	9	cluster_0	8	6	sep	tue	91	129.500
10	10	cluster_4	7	5	sep	sat	92.500	88
11	11	cluster_4	7	5	sep	sat	92.500	88

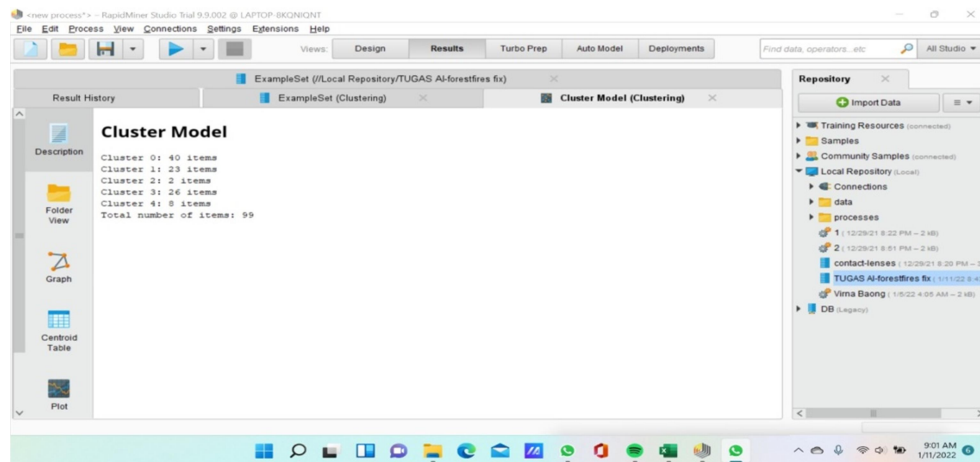


Gambar 4.1.2 data yang siap untuk di cluster

4.2 Hasil dari clustering data

Dari data yang telah di cluster, kami mendapatkan model data sebagai berikut

1. Cluster 0 = 40 item
2. Cluster 1 = 23 item
3. Cluster 2 = 2 item
4. Cluster 3 = 26 item
5. Cluster 4 = 8 item



Gambar 4.2 hasil cluster data

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengolahan data yang ada maka dapat diambil kesimpulan sebagai berikut:

1. Dari pengelompokkan yang dilakukan menggunakan metode *Clustering* dengan algoritma *K-Means* ini, dapat diketahui kelompok faktor yang paling kuat dan besar yang menimbulkan kebakaran hutan.
2. Pada pengujian dengan *RapidMiner*, sangat mudah digunakan dan hasilnya sesuai dengan harapan.
3. Metode *K-Means* dapat digunakan untuk pengelompokkan data kebakaran lahan dan hutan. Pengimplementasian metode *K-Means* pada penelitian ini dimulai dengan mengunduh data pada UCI Repository. Data yang telah diambil akan dimasukkan kedalam proses *clustering* dengan menggunakan algoritma *K-Means*.

5.2 Saran

Pada penelitian ini tentunya tidak terlepas dari kekurangan dan kelemahan, adapun saran yang peneliti berikan adalah melakukan optimasi kombinasi nilai parameter algoritma *K-Means* untuk memperoleh hasil yang optimal, menggabungkan beberapa algoritma agar hasil yang didapat lebih baik.

DAFTAR PUSTAKA

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Freitas, A. A. (2013). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.
- Liautaud, B., & Hammond, M. (2000). *e-Business intelligence: turning information into knowledge into profit*. McGraw-Hill, Inc..
- Maimon, O., & Rokach, L. (Eds.). (2005). *Data mining and knowledge discovery handbook* (Vol. 2). New York: Springer
- Ong, J. O. (2013). *Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing* President University.