

IMPLEMENTASI RAPIDMINER DALAM MENGLASIFIKASIKAN INDEKS DEMOKRASI

Marnis Nasution¹, Ali Akbar Ritonga², Angga Putra Juledi³

Email: marnisnst@gmail.com, aliakbarritonga@gmail.com, anggaj19@gmail.com

^{1,2} Sistem Informasi, Fakultas Sains dan Teknologi Universitas Labuhanbatu

² Teknologi Informasi, Fakultas sains dan Teknologi, Universitas Labuhanbatu

ABSTRACTS

Demokrasi yang menjadi dasar Negara sudah seharusnya terwujud dan terlaksana dengan baik, dan hal tersebut termaksud dalam cakupan daerah yang mengecil, sehingga indeks demokrasi di tiap provinsi mempengaruhi berjalannya demokrasi di Indonesia. Setiap tahunnya Badan Pusat Statistik menerbitkan Indonesia Dalam Statistik yang di dalamnya terdapat banyak data tentang Indonesia yang telah dihimpun dalam satu tahun kebelakang. Menggunakan data yang di dapat dari terbitan BPS yang terdapat dalam bahian Indeks Dekomkrasi Indonesia 2013-2020 provinsi-provinsi yang ada di Indonesia dapat diklasifikasinya sesuai dengan indek demokrasinya menggunakan data mining K-Means yang diimplementasikan menggunakan Rapidminer.sehingga terlihat kelompok daerah yang memiliki indeks demokrasi yang baik.

ARTICLE INFO

Article History:

Received

Revised

Accepted

Available online

Keywords:

Data Mining

K-Means

Rapidminer

Demokrasi

Indeks

© Journal Computer Science and Information Technology(JCoInT)

I. PENDAHULUAN

Negara Indonesia adalah negera berdasarkan demokrasi, dimana dalam periode waktu demokrasi di Indonesia dibedakan menjadi beberapa bagian yaitu Demokrasi Parlemerter periode 1945-2959, Demokrasi Terpimpin periode 1959-1665, Demokrasi Pancasila Periode 1965-1998 dan yang terakhir adalah demokrasi Periode Reformasi yang berjalan hingga saat ini (Evi, 2016)

Melihat perkembangan tersebut dapat dilihat Negara sedang berkembang dalam demokrasi demi dapat mendengar aspirasi rakyatnya sendiri, maka penting bagi sebuah darerah untuh menjadi demokrasi yang mungkin dapat dipindai dari beberapa variabel.

Setiap tahunnya Badan Pusat Statistik meluncurkan buku yang berisi kumpulan angka statistic mengenai Indonesia. Begitupun tahun 2022 terdahulu yang

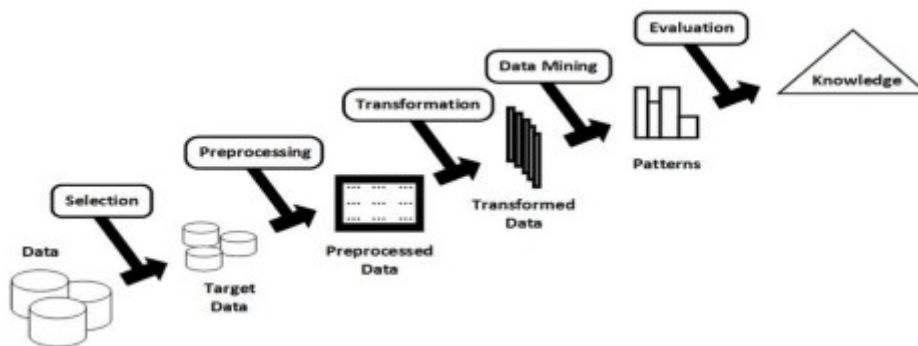
merangkum hal-hal yang terjadi di tahun 2021 dalam bentuk statik, begitupun demokrasi dalam sebuah daerah.

Dengan data yang sebelumnya telah dikumpulkan oleh BPS, untuk dapat melihat daerah daerah yang telah menjalankan demokrasi dengan baik yang dikelompokkan dengan data mining. Salah satu metode *data mining clustering* adalah *K-Means*, yang dapat digunakan untuk melakukan pengelompokkan

2. METODE PENELITIAN

2.1. Knowledge Discovery In Database (KDD)

Knowledge Discovery in Database (KDD) merupakan sebuah terknik yang digunakan untuk membentuk pola atau *rule* dalam informasi. Informasi yang dihasilkan bersumber dari sebuah data yang beukuran besar yang biasa dikenal dengan tambang data yang biasanya disimpan daam database yang dimana awalnya belum diketahui dan kemudian menghasilkan data yang bermanfaat. Berikut lagkah-langkah dalam KDD (Januardi, 2020):



Gambar 2.1 *Knowing Discovery In Database*

1. *Data selection*

Pemilihan atau selessi data dari yang seblumnya adalah sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi di dalam KDD dimulai. Hasil dari selsi kemudian akan digunakan untuk proses data mining serta dipisah dari basis data operasional.

2. *Pre-Processing/Cleaning*

Proses *cleaning* digunakan untuk membuang duplikasi data, memperbaiki inkonsistensi dan kesalahan-kesalahan lain. Dalam proses *cleaning* juga dilakukan proses untuk memperkaya data yang telah ada dengan informasi yang relevan.

3. *Transformationa*

Coding adalah proses transformasi. Dimana merubah data kedalam bentuk yang sesuai dengan proses data *mining*, dan proses ini sangat bergantung terhadap jenis atau pola yang dicari dalam *database*.

4. *Data Mining*

Data mining merupakan prosus untuk menemukan pola ataupun informasi yang menarik daam data yang terpilih dengan teknik ataupun metode tertentu. Teknik ataupun metode dalam data mining sangat bervariasi bergantung pada proses KDD secara keseluruhan

5. *Interpretation/Evaluation*

Dari hasil data mining yang berupa pola informasi perlu ditampilkan dalam bentuk yang mudah dimengerti bagi pihak yang memiliki kepentingan. Tahapan ini merupakan pemeriksaan apakah pola ataupun informasi yang ditemukan bertentangan dengan fakta yang ada. (Yuli, ?)

2.2 *Data Mining*

Data mining merupakan sebuah proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. *Data mining* mencakup pengumpulan data, ekstraksi data, analisis data, statistic data. *Data mining* juga merupakan proses logis untuk menentukan infoamsi yang berguna. Setelah diemukan informasi ataupun pola dapat digunakan untuk alat pendukung dalam pengambilan keputusan dalam mengembangkan bisnis (Muhammad Arhami, dkk. 2020)

Pengelompoka *Data Mining* dapat dibagi menjadi beberapa kelompok:

a. Deskripsi

Deskripsi adalah cara dalam menggambarkan pola serta kecenderungan yang terdapat dalam data yang dimiliki

b. Estimasi

Estimasi hampir sama dengan klasifikasi namun variabel target estimasi lebih kea rah numeric daripada kea rah kategori. Serta model yang dibangun menggunakan record lengkap yang menyediakan nilai variabel target sebagai nilai prediksi

c. Prediksi

Prediksi adalah menerka sebuah nilai yang belum diketahui serta memperkirakan nilainya dimasa mendatang

d. Klasifikasi

Untuk klasifikasi ada target variabel kategori yang dipisahkan kedalam kategori

e. Pengklasteran

Merupakan pengelompokan record, pengamatan ataupun memperhatikan dan membentuk kelas dari objek-objek yang memiliki kemiripan

f. Asosiasi

Asosiasi bertugas menemukan atribut yang munvul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (Novianto, 2020)

2.3 K-Means Clustering

K-Means merupakan salah satu dari pengelompokan yang paling sederhana, dinamakan prosedurnya sederhana dan mudah untuk melakukan klasifikasi data yang sebelumnya diberikan sejumlah *cluster*. Penentuan *centroid* dilakukan dengan cara mengambil data pertama sebagai *centroid* pertama, data kedua sebagai *centroid* kedua, dan seterusnya hingga jumlah *centroid* yang diperlukan. Langkah lanjutannya adalah dengan menghitung jarak dari titik yang akan di *cluster* ke setiap *centroid* yang ada dan dikelompokkan sesuai dengan jarak yang terdekat dari *centroidnya*. Apabila semua titik telah masuk ke dalam pengelompokan maka proses pertama telah selesai. Langkah berikutnya adalah menghitung kembali nilai *centroid* baru sebagai barycenters dari kelompok yang sebelumnya dihasilkan. Setelah memiliki *k-centroid* yang baru, pengelompokan diuji kembali terhadap *centroid* (novianto, 2020)

2.4 Rapidminer

Rapidminer merupakan aplikasi yang berfungsi sebagai alat pembelajaran dalam ilmu *data mining*. Platform ini didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. *rapidminer* memiliki banyak solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi (Vincentius dkk, 2020)

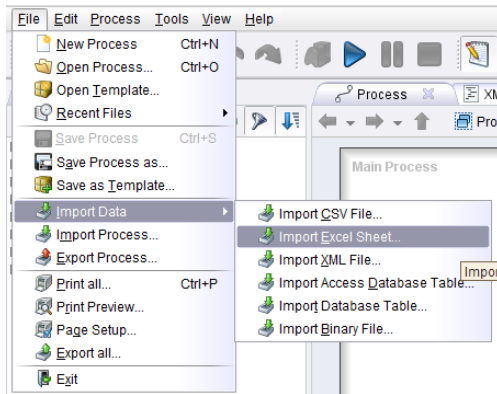
Rapidminer merupakan perangkat lunak independen yang digunakan untuk menganalisa data dan mesin penambangan data, yang dapat diintegrasikan dengan berbagai bahasa pemrograman secara mudah.

3. HASIL DAN PEMBAHASAN

3.1. Pembentukan Model

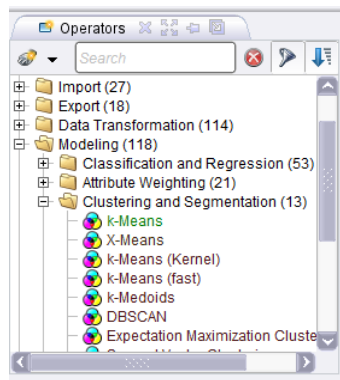
Pemodelan dilakukan dengan memanfaatkan aplikasi *rapidminer*. Proses dilakukan pertama kali dengan membuat data menggunakan Excel. Salah satu bentuk data yang dapat diimport ke dalam *rapidminer* adalah *excel*.

Hal pertama yang dilakukan setelah membuka aplikasi *rapidminer* yaitu mengimport data, dengan cara pilih *toolbox file* kemudian pilih *Import data* selanjutnya pilih bentuk data yang tersedia. Karena sebelumnya data dibuat dalam bentuk *excel* maka yang dipilih adalah *import excel sheet*. Data otomatis akan masuk ke dalam *repository*. Pilihan dapat dilihat pada gambar di bawah ini:



Gambar 2.1 Import Data Excel

Selanjutnya data yang telah masuk ke dalam *repository* dapat ditarik kedalam lembar kerja jika ingin ditambahkan dengan algoritma yang disediakan oleh rapidminer, yang mana saat ini algoritma yang ingin diuji adalah k-means. Algoritma yang ingin diimplementasikan pada data dapat dicari bagian *toolbox* operator. Untuk lebih lengkapnya dapat dilihat pada gambar dibawah ini:



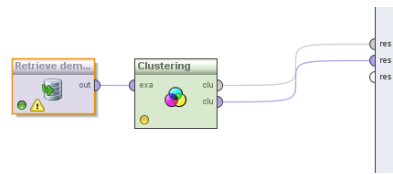
Gambar 2.2 Pilih Model K-Means

Dalam *toolbox* operator dapat ditemukan berbagai macam kebutuhan. Sedangkan untuk algoritnya *k-means* dapat ditemukan didalam menu modeling, kemudian pilih *clustering* dan *segmentation* karena algoritma k-means merupakan bentuk dari *clustering*, setelah baru pilih *k-means*. Algoritma *k-means* dapat ditarik menuju lembar kerja untuk dapat diterapkan kepada data yang sebelumnya telah ada.

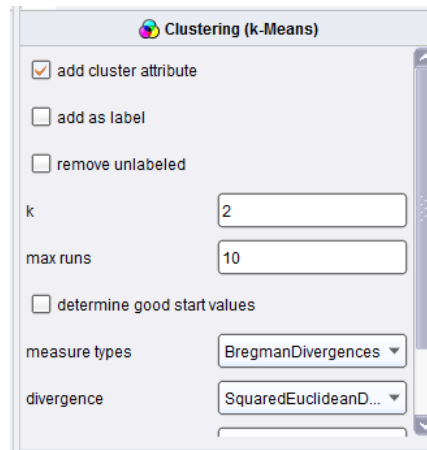


Gambar 2.3 Data dan Model K-means

Data yang telah ada dan algoritma *k-means* yang telah ditarik menuju lembar kerja selanjutnya dihubungkan untuk melihat hasil clusteringnya. Nilai K yang akan diberikan adalah 2 untuk, sehingga 2 pengelompokan, yaitu provinsi-provinsi yang dinilai memiliki demokrasi yang baik sesuai dengan variabel yang ada pada data statistik BPS



Gambar 2.4 Data dan K-Means dihubungkan



Gambar 2.4 Pilih Nilai K

Hasilnya dapat dihasilkan sesaat setelah di-runningkan, akan terbentuk dua *cluster* yakni *cluster 0* dan *cluster 1*. Dimana dari ke 34 provinsi akan dibedakan menjadi 2 cluster sesuai dengan jarak tiap-tiap nilai variabelnya yang berupa indeks demokrasi tiap provinsi setiap tahunnya

Dari hasil pengolahan menggunakan rapidminer didapatkan 11 anggota dari clustering 0 dan 22 anggota dari clustering 1, data yang diperoleh dari buku BPS ada sebanyak 34 provinsi namun dikarenakan provinsi Kalimantan utara yang baru resmi berdiri pada tahun 2012 dan gubernur Kalimantan Utara baru dilantik tahun 2013 sehingga data untuk Kalimantan utara pada taun 2012 dan 2013 kosong.

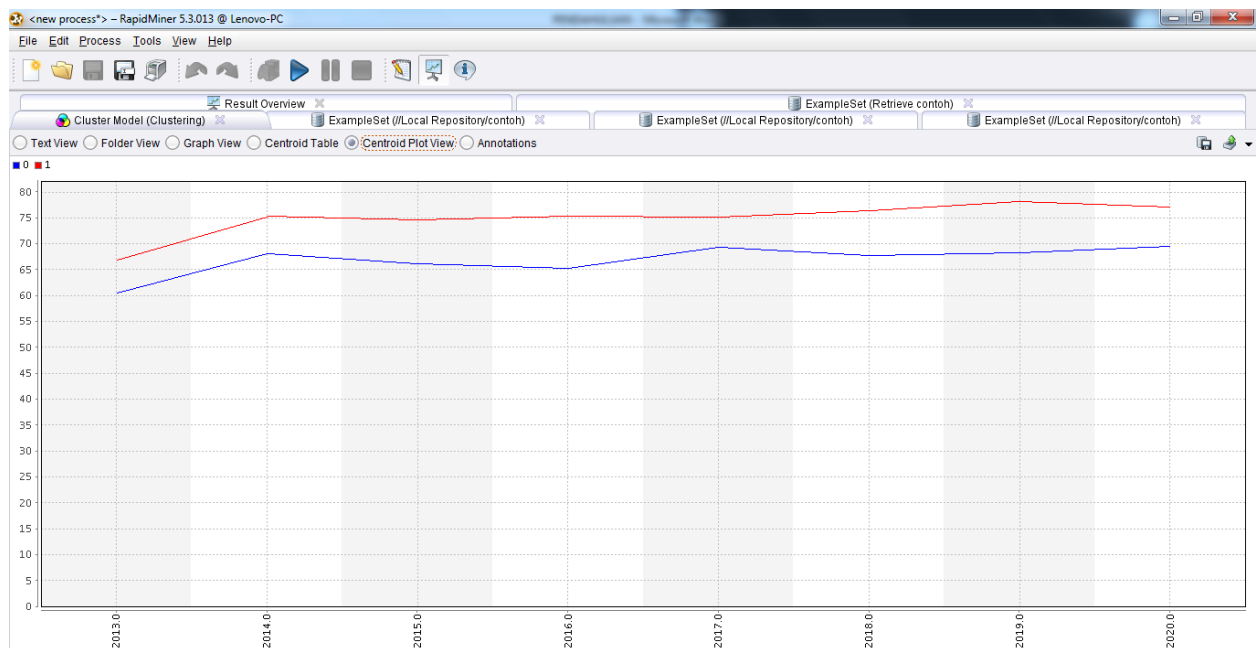
Proses algoritma *K-means data mining* sendiri harus terbebas dari data tidak lengkap untuk dapat diproses sehingga Kalimantan utara tidak dapat diikuti. Sehingga data awal yang berjumlah 34 provinsi menjadi 33 dikarena pembersihan dari data yang tidak lengkap.

```

Cluster Model
Cluster 0: 11 items
Cluster 1: 22 items
Total number of items: 33
  
```

Gambar 2.5 Hasil Cluster dari Rapidminer

Dari proses k-means yang menggunakan rapidminer sendiri juga dapat dilihat plot viewnya, dimana garis berwarna merah merupakan cluster 1, dan garis berwarna biru merupakan cluster 0. Dilihat garis berwarna merah yang merupakan cluster 1 lebih tinggi dari pada cluster berwarna biru yang artinya punya indeks yang lebih baik untuk demokratis



Gambar 2.6 Plot View Clustering

Hasil *clustering* yang didapat dari rapidminer dimana nilai K-nya adalah 2 yang artinya data dikelompokkan menjadi 2

Cluster 0: {Sumatera Utara, Jambi, Lampung, Jawa Barat, Nusa Tenggara Barat, Sulawesi Selatan, Sulawesi tenggara, Maluku Utara, Papua Barat, Papua}

Cluster 1: {Aceh, Riau, Sumatera Selatan, Bengkulu, Kepulauan Riau, DKI Jakarta, Jawa tengah, DI Yogyakarta, Jawa Timir, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Maluku}

4. KESIMPULAN

Pada penelitian ini, penulis dapat merangkum beberapa poin dibawah ini:

1. Algoritma *Data Mining K-Means* dapat berguna mengklasifikasikan sebuah data kedalam beberapa kelompok, dan semakin banyak pengkalsifikasian semakin baik
2. Dengan data indeks demokrasi yang didapat dari Badan Pusat Statistik dapat dikelompokkan profinsi-profinsi yang memiliki tingkat demokrasi yang baik
3. Kelompok provinsi yang memiliki nilai lebih tinggi dapat di klasifikasikan sebagai profinsi yang memiliki indeks demokrasi yang lebih baik

5. DAFTAR PUSTAKA

- Badan Pusat Statistik. (2022). Statistik Indonesia Statistical Yearbook Of Indonesia 2022.
- Dhuhita, Windha Mega Pradnya. (2015). Clustering Menggunakan Metode K-Means Untuk Menentukan Status Balita. *Jurnal Informatika*. Vol.15, No.2, 160-174
- Mardi, yuli. (2016). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*. Vol.2, No.2
- Meilina, Popy. (2015). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi. *Jurnal Teknologi*. Vol.7, no.1
- Nasir, Januardi. (2020). Penerapan data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means. *Jurnal SIMETRIS*. Vol.11, No.2
- Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Jurnal Paradigma*. Vol.22, No.2, 183-189
- Prasetyo, Vincentius Riandaru., Hamzah Lazuardi.m Aldo Adhi Mulyono., Christian Lauw. (2021). Penerapan Aplikasi Rapidminer Untuk Prediksi Nilai Tukar Rupiah terhadap US Dollar Dengan Metode Regresi Linier. *Jurnal Nasional Teknologi dan Sistem Informasi*. Vol.07, No.01, 8-17
- Purnamasari, Tika., Marnis Nasution & Gomali Juni Yanris. (2021). *Jurnal Teknologi dan Informasi*. Vol.8, No.3, 251-258
- Purnawati, Evi. (2020). Perjalanan Demokrasi Di Indonesia. *Jurnal SOLUSI*. Vol. 18, No. 2, 251-264