PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA PEMBESARAN JENIS IKAN UNGGULAN DI PROVINSI JAWA BARAT

Dede Setiadi [1], Bambang Irawan [2], Agus Bahtiar [3]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B, Majasem, Cirebon Indonesia
dedesetiadi321@gmail.com

ABSTRAK

Indonesia sebagai negara kepulauan yang besar, memiliki perairan yang lebih luas dibandingkan dengan daratannya. Karena ukuran perairannya yang besar, Indonesia memiliki potensi yang besar dalam produksi budidaya perikanan, terutama untuk meningkatkan pertumbuhan jenis ikan unggulan. Penelitian ini bertujuan untuk menerapkan metode data *mining* dengan menggunakan algoritma *K-Means clustering*. Data produksi perikanan tahun 2020 akan digunakan sebagai sumber data untuk penelitian ini. Data diambil berasal dari Open Data Jabar (ODJ). Data *mining* diterapkan untuk menganalisis data produksi budidaya pembesaran dari masing-masing jenis ikan unggulan di Provinsi Jawa Barat. Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery and Data Mining (KDD)* yang merupakan salah satu metode yang paling populer dalam data *mining*. Langkah pertama dalam penelitian ini adalah pengumpulan dan persiapan data, data produksi perikanan budidaya pembesaran akan dikumpulkan dari sumber yang relevan. Dengan menggunakan algoritma *K-Means*, pengelompokkan produksi perikanan budidaya pembesaran jenis ikan unggulan dapat diidentifikasi. Misalnya, dibagi menjadi 5 *cluster*, *cluster* 1 (C1), *cluster* 2 (C2), *cluster* 3 (C3), *cluster* 4 (C4), dan *cluster* 5 (C5). Hasil penelitian ini yaitu *cluster* 0: 63 item, *cluster* 1: 9 item, *cluster* 2: 64 item, *cluster* 3: 93 item, *cluster* 4: 1 item. *items*. Klaster dengan pertumbuhan ikan sangat cepat berada di cluster 2 yaitu jenis ikan lele, patin, sepat siam, udang krosok, udang windu, udang danau, rumput laut, kepiting bakau sedangkan pertumbuhan ikan yang sangat lambat ada di cluster 1 yaitu jenis ikan tambakan.

Kata kunci: *Data Mining, Algoritma K-Means, Pengelompokkan, Produksi Perikanan Budidaya, Provinsi*

1. PENDAHULUAN

Industri budidaya perikanan yang semakin berkembang mempunyai peran strategis dalam mendukung ketahanan pangan dan perekonomian daerah. Provinsi Jawa Barat sebagai salah satu pusat kegiatan ekonomi dan pertanian di Indonesia memiliki potensi besar dalam pengembangan produksi perikanan. Dalam upaya meningkatkan efisiensi dan produktivitas sektor perikanan, diperlukan suatu pendekatan yang sistematis dan efektif untuk mengelompokkan jenis ikan unggulan yang dihasilkan. Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means* sebagai metode klasterisasi dalam produksi perikanan budidaya pembesaran berdasarkan jenis unggulan di Provinsi Jawa Barat. Jenis ikan unggulan yang dimaksud disini adalah ikan dengan kriteria pemeliharaannya yang paling cepat. Klasterisasi jenis ikan unggulan diharapkan dapat memberikan informasi yang lebih rinci mengenai ikan mana yang paling cepat pertumbuhannya untuk dikonsumsi masyarakat di Provinsi Jawa Barat. Di tengah perkembangan teknologi informasi yang semakin pesat, data menjadi harta karun yang tak ternilai harganya. Namun, pengelompokkan produksi perikanan budidaya pembesaran secara efektif dalam skala besar menjadi tantangan yang kompleks.

Pada penelitian sebelumnya [1] Klasifikasi hasil tangkapan tahun 2015 hingga 2017 didasarkan pada algoritma *K-Means*, dimana telah teridentifikasi dua *cluster* yaitu *cluster* 1 dengan jumlah ikan lebih sedikit dan *cluster* 2 dengan jumlah ikan lebih banyak. Spesies ikan yang berbeda dapat ditempatkan dalam satu atau lebih kelompok sebagai hasil untuk mendapatkan hasil yang pasti. Ikan madidihihang, selar, gemuk, tembang, teri, kaisar, julung, kakap, swanggi, baronang, kwee, mackerel, nangka potong, tongkol krai, kerapu dan miss blue merupakan 16 jenis ikan yang termasuk dalam *cluster* 1 (C2). Ada dua spesies ikan di dalam *cluster* 2, yaitu fly fish dan cakalang. Kehadiran spesies ikan unggul di *cluster* 2 (C2) karenanya dapat dilihat. Untuk menerapkan algoritma serupa dalam analisis segmentasi yang bertujuan mendeteksi penyakit pada daun jagung, digunakan pendekatan *K-Means*. Tahapan prosesnya melibatkan penetapan jumlah klaster awal (k=3), penggunaan *centroid* acak, perhitungan jarak antara nilai piksel dengan *centroid*, pengelompokkan nilai piksel berdasarkan jarak minimum, perhitungan rata-rata klaster untuk fokus baru, serta proses stokastik pusat gravitasi jika nilai piksel masih bergerak. Iterasi ini terus dilakukan hingga tidak ada perubahan pada nilai piksel yang bergerak. Dalam penelitian ini, digunakan kumpulan data sebanyak 30 tipe A dan B untuk tahap pelatihan, dan 10 tipe data untuk tahap pengujian yang berhasil mencapai akurasi sebesar 90%. [2]. Penelitian sebelumnya [3] *K-Means*, pengelompokan dan penggunaan data dukungan perangkat lunak *RapidMiner* dapat ditingkatkan agar nelayan dapat menggunakan sumber daya perikanan

menurut provinsi lagi. Penelitian menunjukkan bahwa Jawa Tengah merupakan salah satu dari 21 provinsi di Indonesia yang mencapai tingkat klasterisasi tinggi (C1). Sementara itu, Provinsi Kalimantan Tengah, Jawa Timur, Sulawesi Tenggara, Sumatera Barat, Jawa Barat, Gorontalo, Nusa Tenggara Barat, Bangka Belitung, Sulawesi Utara, Bengkulu, Kalimantan Barat, DI Yogyakarta, Sumatera Utara, Sulawesi Tengah, Kepulauan Seribu, Bali, Lampung, Sulawesi Selatan, DKI Jakarta, dan Banten tergolong dalam klaster tingkat rendah (C2).

Tujuan utama dari penelitian ini adalah untuk mengimplementasikan algoritma *K-Means* sebagai metode klasterisasi untuk memahami pola produksi perikanan budidaya pembesaran jenis ikan unggulan di Provinsi Jawa Barat, juga bertujuan untuk mengisi kesenjangan pengetahuan terkait dengan optimasi produksi perikanan menggunakan pendekatan informatika, khususnya dalam konteks klasterisasi data produksi perikanan. Kontribusi utama penelitian ini terletak pada penerapan metode informatika yang inovatif untuk mengoptimalkan praktik budidaya perikanan di wilayah Jawa Barat, yang selanjutnya dapat menjadi panduan bagi masyarakat untuk mengonsumsi ikan yang pemeliharaannya paling cepat guna ketahanan pangan di Provinsi Jawa Barat.

Metode yang digunakan adalah dengan melibatkan implementasi algoritma *K-Means* sebagai teknik klasterisasi untuk mengelompokkan jenis ikan unggulan. Pendekatan ini dianggap tepat untuk memahami karakteristik produksi secara holistik, memungkinkan identifikasi kelompok ikan dengan perilaku serupa. Pendekatan eksperimental juga akan diterapkan untuk memvalidasi hasil klasterisasi dengan menggunakan data produksi aktual. Dalam upaya mengoptimalkan penggunaan teknologi informasi, pendekatan komputasi akan difokuskan pada pengelolaan dan analisis data besar yang dihasilkan oleh sektor perikanan. Integrasi antara algoritma klasterisasi dan analisis data akan menjadi landasan utama dalam penelitian ini, memastikan bahwa hasil yang diperoleh dapat memberikan pandangan yang lebih mendalam terkait dengan pola produksi perikanan budidaya di Provinsi Jawa Barat. Dengan menggunakan algoritma *K-Means*, pengelompokkan produksi perikanan budidaya pembesaran jenis ikan unggulan dapat diidentifikasi. Misalnya, dibagi menjadi 5 *cluster*, *cluster* 1 (C1), *cluster* 2 (C2), *cluster* 3 (C3), *cluster* 4 (C4), dan *cluster* 5 (C5).

Hasil penelitian ini dapat digunakan oleh Dinas Kelautan dan Perikanan untuk mengoptimalkan produksi perikanan budidaya pembesaran dan merencanakan strategi manajemen yang lebih efektif dalam industri perikanan. Penelitian ini akan memberikan kontribusi penting dalam penggunaan teknologi informasi untuk mengoptimalkan sektor perikanan budidaya pembesaran, yang merupakan aspek vital dalam pemenuhan kebutuhan pangan dan pertumbuhan ekonomi di berbagai wilayah. Dengan demikian, penelitian ini menjadi relevan dan penting dalam konteks perkembangan pesat di bidang informatika yang telah mengubah cara kita mengelola dan memahami data untuk meningkatkan kualitas hidup dan efisiensi dalam berbagai sektor kehidupan. Penelitian ini dapat meningkatkan pemahaman kita tentang kemampuan algoritma *K-Means* dalam mengelompokkan data produksi perikanan budidaya pembesaran jenis ikan unggulan. Hal ini akan memberikan pemahaman yang lebih mendalam tentang efektivitas algoritma *K-Means* dalam analisis data sektor perikanan budidaya. Temuan dari penelitian ini dapat digunakan oleh praktisi di sektor perikanan, produsen, serta pemerintah daerah dan pusat untuk membuat keputusan yang lebih tepat dan efisien dalam mengelola produksi perikanan budidaya pembesaran. Dengan demikian, hasil penelitian ini memiliki potensi untuk tidak hanya memperkaya pemahaman kita tentang informatika, tetapi juga memberikan manfaat praktis yang signifikan bagi pengelolaan sumber daya ikan di Provinsi Jawa Barat dan dapat membuka jalan bagi perkembangan teknologi informasi dalam berbagai sektor.

Berdasarkan pemaparan latar belakang diatas maka rumusan masalahnya sebagai berikut; Bagaimana penerapan algoritma *K-Means* dalam pengelompokkan produksi perikanan budidaya pembesaran jenis ikan unggulan di Provinsi Jawa Barat dan Jenis ikan mana yang memiliki tingkat pertumbuhan paling cepat dan jenis ikan mana yang memiliki tingkat pertumbuhan paling lambat.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Paper [4] membahas tentang bagaimana produksi ikan air tawar di Sumatera Selatan dipetakan secara visual, menggunakan metode *K-Means Cluster* untuk mengelompokkan kota/kabupaten berdasarkan produksi perikanan. Tujuan penelitian ini adalah untuk membantu pengaturan kebijakan pemerintah provinsi dalam meningkatkan produksi perikanan air tawar di setiap kota/kabupaten. Hasilnya menunjukkan pembentukan 2 kluster berdasarkan analisis metode *Elbow*, *Silhoutee*, dan *Gaps Statistics*. Kluster 1, terdiri dari 12 kota/kabupaten, memiliki tingkat produksi rendah, sementara kluster 2, hanya terdiri dari 2 kota/kabupaten, memiliki tingkat produksi yang tinggi.

2.2. Data Mining

Data *mining* merupakan proses yang mencakup pengumpulan serta penggunaan data historis guna mengidentifikasi pola, keteraturan, dan hubungan dalam dataset yang besar. Salah satu fokus utama dalam data *mining* adalah pengelompokkan *clustering*, yaitu saat data yang dikelompokkan tidak memiliki contoh kelompok sebelumnya [1].

Data *mining* memiliki peran yang krusial dalam sejumlah aspek kehidupan, termasuk industri, keuangan, cuaca, ilmu, dan teknologi karena

digunakan dalam pengolahan data besar. Di dalamnya, terdapat beragam metode seperti klasifikasi, *clustering*, *regresi*, seleksi variabel, dan analisis keranjang belanja, yang sangat berguna dalam mengeksplorasi informasi dari kumpulan data yang besar [5].

2.3. Algoritma K-Means

Algoritma *K-Means*, bagian dari jenis pembelajaran tanpa supervisi, difungsikan untuk mengelompokkan data menjadi beberapa klaster dengan metode pemisahan tertentu. Pembelajaran tanpa supervisi adalah metode dalam data *mining* yang bertujuan untuk mengidentifikasi pola dari berbagai variabel (atribut), tanpa menentukan variabel mana yang akan menjadi label. [6].

K-Means merupakan salah satu teknik dalam data *mining* yang melakukan proses analisis data tanpa supervisi, di mana pendekatannya terfokus pada pengelompokan berdasarkan kesamaan atau pola yang ada. Data-data yang memiliki sifat yang mirip akan dikelompokkan dalam satu *cluster*, sedangkan data yang memiliki perbedaan karakteristik akan ditempatkan dalam *cluster* yang berbeda [7]**.**

2.4. Clustering

Clustering adalah teknik yang mencari dan mengelompokkan data yang memiliki kesamaan karakteristik. Fokus utama dari *clustering* adalah mengelompokkan data atau objek ke dalam cluster sehingga setiap *cluster* memuat data yang serupa. Dalam proses ini, clustering berupaya menyatukan objek yang serupa ke dalam satu *cluster* sambil menjaga jarak yang jauh antar *cluster* sebisa mungkin. Hasilnya, objek dalam satu *cluster* saling mirip satu sama lain dan berbeda dengan objek dalam *cluster* lainnya [8]

Clustering adalah metode yang digunakan untuk mengelompokkan berbagai entri data ke dalam kategori yang mirip, membentuk kelas objek yang berisi data serupa dan membedakan kelompok satu dari yang lain [9].

2.5. RapidMiner

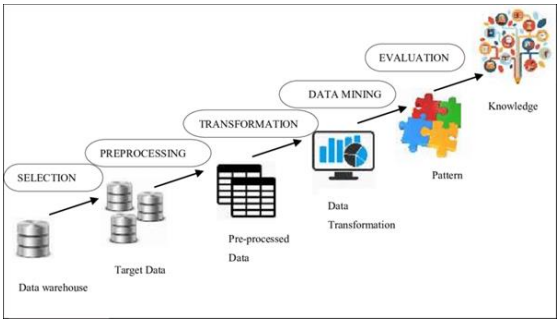
RapidMiner, sebuah perangkat lunak yang dapat diakses oleh siapa saja dan memiliki sifat terbuka (*open source*), telah diadopsi sebagai solusi dalam analisis data. Dalam *platform RapidMiner*, beragam teknik diterapkan, termasuk teknik deskriptif dan prediksi, yang digunakan untuk proses analisis data. Pengoperasian *RapidMiner* menggunakan bahasa pemrograman Java [10].

Sebelum dikenal sebagai *RapidMiner*, *software* ini awalnya dikenal sebagai *YALE (Yet Another Learning Environment).* Pada tahun 2001, *Ralfklinkenberg*, *Ingo Mierswa*, dan *Simon Fischer* mulai mengembangkan versi awalnya di *Artificial Intelligence* Unit dari University of Dortmund. Dengan distribusi menggunakan lisensi *AGPL (GNU Affero General Public License)* versi 3, *RapidMiner* merupakan perangkat lunak sumber terbuka untuk keperluan data mining yang telah mendapat pengakuan global yang solid [11].

3. METODE PENELITIAN

3.1. Metode Penelitian

Metode yang dipakai yaitu *Knowledge Discovery in Database (KDD)*, yang merupakan metode penelitian dalam menemukan pengetahuan dari database. Saat menerapkan teknik data *mining* ini, langkah-langkah analisis data mengikuti tahapan KDD. Tahapan-tahapan untuk menganalisis data bisa ditemukan dalam Gambar 1.



Gambar 1. Metode *Knowledge Data Discovery (KDD)*

Dari Gambar 1 tentang KD dapat dijelaskan sebagai berikut;

1. Data *Selection*
   Tahap awal ini melibatkan pemilihan data yang diperlukan sebelum memulai pencarian informasi dalam proses *Knowledge Data Discovery (KDD).* Data yang telah terpilih akan digunakan pada tahap selanjutnya dan akan disimpan secara terpisah dari basis data yang digunakan secara operasional.
2. *Preprocessing* atau *Cleaning*
   Sebelum memulai data *mining*, langkah awalnya adalah membersihkan data. Tujuannya adalah untuk menghapus duplikat, memeriksa kekonsistenan data, dan memperbaiki kesalahan seperti kesalahan pengetikan atau cetak. Proses *enrichment* juga dilakukan untuk memperkaya data dengan informasi tambahan yang relevan.
3. *Transformation*
   Proses ini terlibat dalam mengubah data yang sudah dipilih sehingga cocok untuk digunakan dalam data *mining*. Langkah ini sangat bergantung pada jenis informasi yang diinginkan dari basis data.
4. Data *Mining*
   Merupakan tahap dimana pola atau informasi yang menarik dicari dari data yang telah terpilih menggunakan teknik atau metode tertentu. Pemilihan metode dan algoritma harus dilakukan dengan tepat sesuai dengan tujuan dari *Knowledge Data Discovery (KDD)* secara keseluruhan.

5. *Interpretation/Evaluation*
   Pola informasi yang diperoleh dari data *mining* harus disampaikan secara jelas kepada pihak yang berkepentingan agar mudah dipahami. Langkah ini merupakan bagian dari proses *Knowledge Data Discovery (KDD)*, di mana pengecekan dilakukan untuk memastikan kesesuaian atau ketidaksesuaian pola informasi dengan fakta dan hipotesis yang telah ada sebelumnya [12].

3.2. Sumber Data

Dalam penelitian ini, sumber data yang digunakan berasal dari Open Data Jabar (ODJ). ODJ adalah lembaga resmi yang bertanggung jawab untuk mengumpulkan, mengolah, dan menyediakan data statistik terpercaya yang berkaitan dengan berbagai aspek kehidupan di Provinsi Jawa Barat. Kepercayaan pada data dari ODJ menjadi kunci penting dalam penelitian ini, karena data ini merupakan referensi utama untuk analisis *clustering* yang akan dilakukan. Dengan memanfaatkan data ODJ, penelitian ini diharapkan dapat memberikan wawasan yang berharga tentang karakteristik produksi perikanan budidaya pembesaran jenis ikan unggulan di Provinsi Jawa Barat, yang dapat menjadi dasar untuk pengambilan keputusan yang lebih baik dalam pengembangan sektor perikanan budidaya. Data yang diambil adalah data produksi perikanan budidaya pembesaran pada tahun 2020. Selain itu ada tambahan data tentang lama pemeliharaan ikan.

3.3. Populasi dan Sampel

Populasi adalah sekumpulan objek atau subjek yang memiliki karakteristik spesifik yang didefinisikan oleh peneliti, yang menjadi dasar untuk membuat kesimpulan. Dalam konteks tahun 2020, kelompok utama dalam industri perikanan merupakan populasi yang dibahas dalam penelitian ini. Sampling merujuk pada proses pengambilan sebagian kecil dari populasi tersebut untuk mewakili nilai atau karakteristik keseluruhan populasi. Bagian yang diambil dari populasi ini disebut sebagai sampel, yang merupakan representasi dari keseluruhan populasi. Dalam konteks penelitian ini, populasi merujuk pada keseluruhan produksi perikanan budidaya pembesaran jenis ikan unggulan di Provinsi Jawa Barat.

Sampel dalam penelitian ini merupakan subset atau bagian yang diambil dari populasi tersebut. Karena memeriksa seluruh data produksi perikanan budidaya pembesaran dari seluruh kabupaten di Provinsi Jawa Barat akan menjadi tugas yang sangat besar dan kompleks. Sampel ini akan digunakan untuk melakukan analisis clustering dengan algoritma *K-Means*. Pemilihan sampel yang representatif akan memungkinkan penelitian ini untuk memberikan wawasan yang relevan tentang pola dan tren dalam produksi perikanan budidaya di Provinsi Jawa Barat tanpa harus memeriksa seluruh populasi yang sangat luas.

3.4. Teknik Pengumpulan Data

Metode penelitian kepustakaan yang kegiatannya dilakukan adalah mempelajari, mencari dan mengumpulkan data yang berkaitan dengan penelitian tersebut, telah digunakan untuk kegiatan pengumpulan data dalam penelitian ini. Data yang digunakan dalam pengelompokan produksi perikanan budidaya menurut jenis ikan diperoleh dari data Open Data Jabar (ODJ) berdasarkan data tahun 2020 tempat penelitian ini dilakukan. Kemudian untuk mengklasifikasikan produksi perikanan budidaya menurut jenis ikan unggulan, data yang diperoleh akan diolah berdasarkan metode *K-Means* yang mengambil beberapa nilai atribut yang terdapat pada data tersebut.

3.5. Teknik Analisis Data

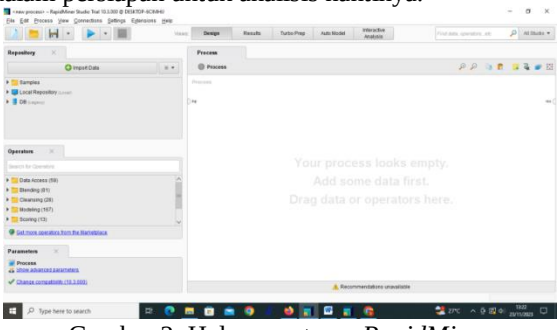
Analisis data melibatkan langkah-langkah sistematis untuk mengumpulkan dan mengatur informasi dari berbagai sumber, mengelompokkan data ke dalam kategori, menggambarkannya dalam unit-unit yang relevan, menyusunnya secara keseluruhan, mencari pola, menyoroti informasi penting, menentukan fokus belajar, dan merangkumnya untuk mendapatkan pemahaman lebih dalam tentang diri sendiri dan orang lain.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

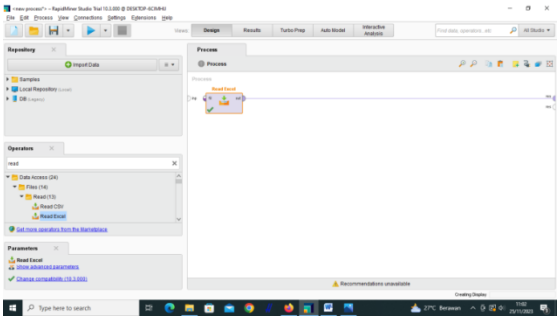
4.1.1. Data Selection

Data *selection* merupakan proses memilih data mana yang akan dimasukkan dan mana yang tidak di dalam aplikasi *RapidMiner*, ini adalah langkah awal dalam persiapan untuk analisis nantinya.



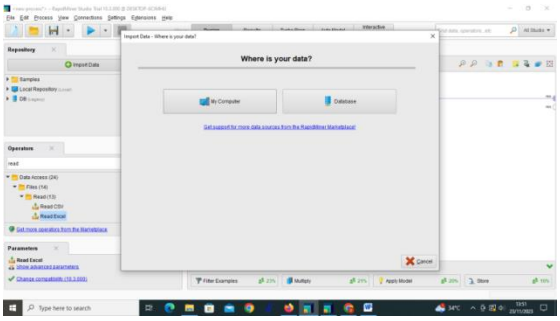
Gambar 2. Halaman utama *RapidMiner*

Setelah Rapidminer dibuka seperti pada Gambar 2, langkah berikutnya adalah drag dan drop operator yang akan dipakai dalam penelitian.



Gambar 3. *Read excel*

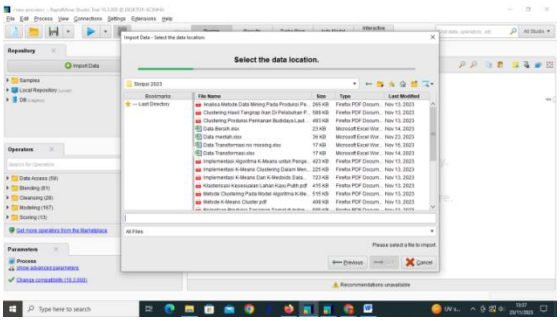
Tahap Gambar 3. Read Excel dipakai untuk membaca informasi dari file excel yang akan diproses, dan untuk mengimpor data, pilihlah wizard konfigurasi impor yang terletak di bagian parameter di sisi kanan layar monitor. Lanjut pilihlah data yang akan diolah.



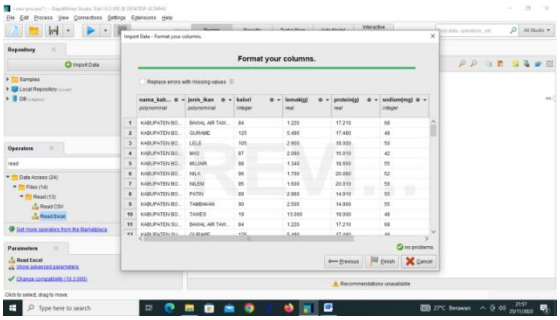
Gambar 4. *Import* data

Pada tahap Gambar 3, langkah yang diperlukan adalah mengklik "read excel" dan kemudian " import configuration" untuk menampilkan hasil di Gambar 4. Setelah itu, klik "my computer", yang akan menampilkan tampilan serupa dengan Gambar 5. Selanjutnya, pilih data yang ingin diproses.

Saat data diimpor, sistem akan menampilkan *preview* data sebelum data tersebut dimasukkan. Hal ini dilakukan untuk memverifikasi kebenaran informasi yang dimasukkan. Setelahnya, Anda dapat mengklik "*finish*" untuk menyelesaikan proses pemilihan atribut.



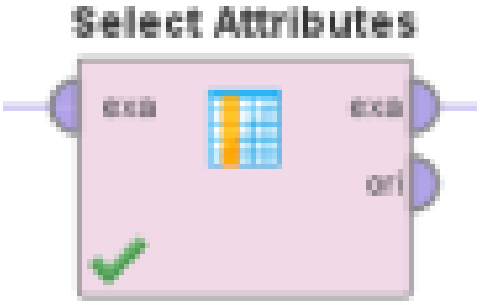
Gambar 5. *Selection* data



Gambar 6. Proses *import* data

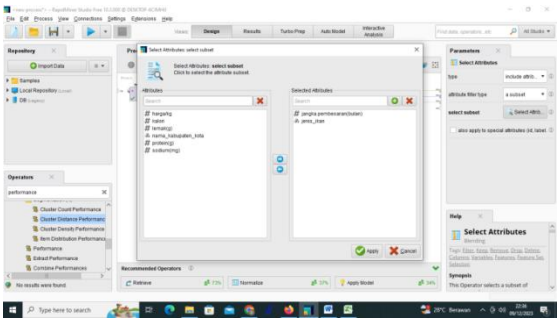
4.1.2. Preprocessing/Data *Cleaning*

Membersihkan data dari entri yang tidak valid seperti entri kosong, duplikat, dan entri yang hilang dilakukan untuk mencegah masalah dan memastikan data dapat diproses dengan lancar.



Gambar 7. Operator *select attributes*

Pada Gambar 7, operator "*Select Attributes*" dalam aplikasi *RapidMiner* digunakan untuk memilih dan mengelola atribut-atribut yang relevan dari dataset. Ini merupakan langkah penting dalam tahap pra-pemrosesan data sebelum melakukan analisis lebih lanjut.



Gambar 8. Proses *select attributes*

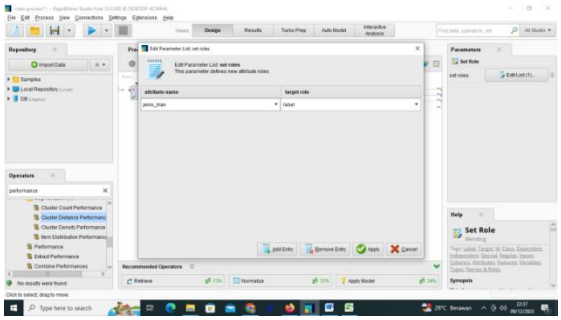
Pada Gambar 8. merupakan proses mengubah atribut jenis ikan menjadi label, operator ini memungkinkan pengguna untuk memilih atribut-atribut tertentu dari dataset yang akan diproses lebih lanjut. Pemilihan atribut bisa dilakukan berdasarkan berbagai kriteria, seperti informasi dari *feature selection*, korelasi, atau metode lain yang relevan.

4.1.3. Transformasi

Setelah proses *Cleaning* selanjutnya dilakukan proses transformasi data sesuai dengan jenis data tersebut. Pada proses ini memerlukan operator *Set Role*. Operator "*Set Role*" dalam aplikasi *RapidMiner* digunakan untuk menetapkan peran atau perbedaan dalam dataset. Ini memungkinkan pengguna untuk mendefinisikan jenis dari atribut (kolom) dalam dataset, seperti menentukan apakah suatu atribut adalah variabel input, output, atau bahkan sebagai metadata.

Gambar 9. Operator *set role*

Digunakan untuk mengubah peran satu atau lebih dari satu atribut yaitu memberikan label atau id kepada variabel yang di inginkan. Pada penelitian ini variabel "jenis_ikan" diubah menjadi id. Untuk mengubah variable "jenis_ikan" menjadi id, pilih *set role* seperti pada Gambar 9.

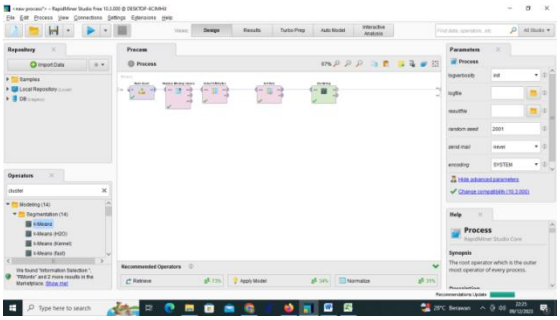


Gambar 10. Proses *set role*

Kemudian akan tampil parameter set role seperti pada Gambar 10. setelah itu select attribute name "jenis_ikan" diubah menjadi id.

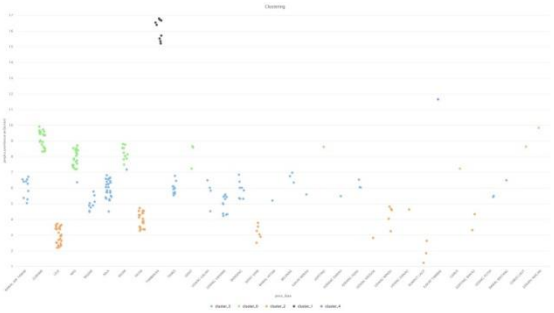
4.1.4. Data Mining

Langkah selanjutnya dalam urutan ini adalah melakukan data *mining*, yang melibatkan penggunaan beragam teknik guna menemukan pola dan informasi khas di dalam data yang dipilih. Pengujian dilakukan dengan Rapidminer.



Gambar 11. *Clustering*

Pada Gambar 11, proses *Clustering* akan menjalani identifikasi bobot nilai *cluster*, yang nantinya akan dianalisis lebih lanjut guna mencapai hasil yang diinginkan.

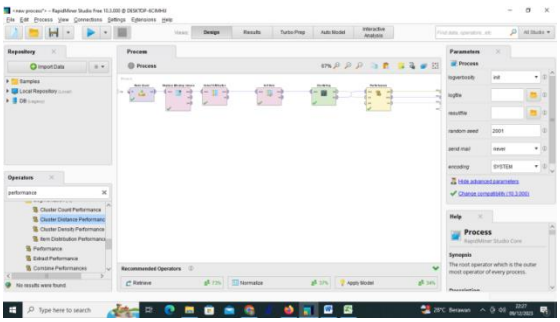


Gambar 12. Hasil *clustering*

Pada Gambar 12, merupakan *Visualisasi* hasil *Clustering* menggunakan aplikasi *RapidMiner* dengan menggunakan algoritma *K-Means*. Disana menggambarkan 5 klaster yaitu, klaster 0 (sedang), klaster 2 (sangat cepat), klaster 3 (cepat), klaster 4 (lambat), dan klaster 1 (sangat lambat).

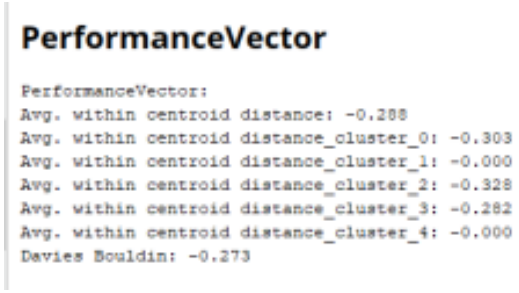
4.1.5. Evaluasi

Langkah berikutnya adalah mengecek performa dalam *RapidMiner*.



Gambar 13. *Performance*

Pada Gambar 13 dimasukkan operator *Performance* untuk pemrosesan data dalam menangani dataset.



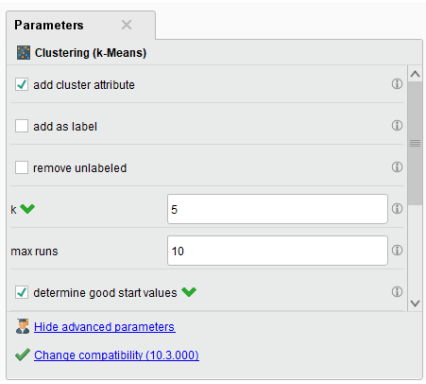
Gambar 14. Hasil *performance*

Dapat dilihat pada Gambar 14 artinya rata-rata jarak dalam *centroid (within centroid distance)* untuk beberapa *cluster* memiliki nilai negatif, mungkin hasil dari normalisasi atau skema evaluasi tertentu yang digunakan. Nilai negatif tidak umum terjadi pada metrik ini, biasanya nilai positif yang lebih diharapkan. Sebagian besar cluster memiliki nilai rata-rata jarak yang negatif, kecuali *cluster* 1 dan *cluster* 4 yang memiliki nilai 0, yang mungkin menunjukkan karakteristik khusus dari *cluster* tersebut dalam hal jarak titik-titiknya dengan *centroid*. *Davies Bouldin* memiliki nilai negatif juga, yang, lagi, mungkin dari normalisasi atau skema evaluasi tertentu.

4.2. Pembahasan

4.2.1. Mengelompokkan Hasil Produksi Perikanan Budidaya Pembesaran Jenis Ikan Unggulan Di Provinsi Jawa Barat Menggunakan Algoritma K-Means Dengan Aplikasi Rapidminer

Setelah memasukkan dan memeriksa data, peneliti mulai memodelkan dataset. Dengan menggunakan algoritma *k-means* melalui perangkat lunak *RapidMiner*. Langkah awal adalah memasukkan data ke dalam *RapidMiner*, diikuti dengan pemrosesan data menggunakan metode *k-means*. Dengan mengatur parameter menjadi k = 5 atau 5 klaster.



Gambar 15. *Parameter* pada *RapidMiner*

Selanjutnya data di proses dengan *RapidMiner* dang menghasilkan 5 model klaster yang berisi *cluster* 0: 63 *items*, *cluster* 1: 9 *items*, *cluster* 2: 64 *items*, *cluster* 3: 93 *items*, *cluster* 4: 1 *items*. *Cluster* dimulai dari 0 karena pada bahasa pemrograman angka 0 adalah angka pertama dari urutan penomoran.

Klaster 0 (sedang) atau dengan lama pemeliharaannya sedang adalah jenis ikan gurame, mas, nilem, sidat, kepiting, gabus, gabus laut, kerapu macan, klaster 2 (sangat cepat) atau dengan lama pemeliharaannya sangat cepat adalah jenis ikan lele, patin, sepat siam, udang krosok, udang windu, udang danau, rumput laut, kepiting bakau, klaster 3 (cepat) atau dengan lama pemeliharaannya cepat adalah jenis ikan bawal air tawar, mujair, nila, tawes, udang galah, udang vaname, bandeng, bawal hitam, belanak, kakap merah, kerang darah, kerang hijau, bawal bintang, klaster 4 (lambat) atau dengan lama pemeliharaannya lambat adalah jenis ikan kakap tambak, dan klaster 1 (sangat lambat) atau dengan lama pemeliharaannya sangat lambat adalah jenis ikan tambakan.

4.2.2. Melihat Hasil Klaster Jenis Ikan Mana Yang Memiliki Tingkat Pertumbuhan Yang Paling Cepat Dan Yang Paling Lambat

Dalam klasifikasi berdasarkan pertumbuhan, beberapa klaster menunjukkan tingkat pertumbuhan yang berbeda untuk jenis ikan yang berbeda pula. Klaster 2, yang termasuk ikan lele, patin, sepat siam, udang krosok, udang windu, udang danau, rumput laut, dan kepiting bakau, menonjol dengan pertumbuhan yang sangat cepat. Sebaliknya, klaster 1, yang hanya berisi ikan tambakan, dikenali sebagai klaster dengan pertumbuhan yang paling lambat di antara kelompok yang teridentifikasi. Perbedaan ini memberikan gambaran jelas tentang variasi pertumbuhan yang dapat diamati pada jenis ikan yang berbeda dalam konteks klasterisasi ini.

Ikan yang berada di klaster 2 memiliki nilai gizi yang memadai, tersedia dengan harga yang ekonomis, dan memiliki lama pemeliharaan yang relatif singkat dalam perbandingan dengan spesies lainnya. Sedangkan ikan yang berada di klaster 1 memiliki gizi yang tinggi, harganya mahal dan juga lama pemeliharaannya paling lambat. Gizi yang mencukupi, harga yang ekonomis, dan waktu pemeliharaan yang relatif singkat menjadikan ikan sebagai sumber protein yang penting dan dapat diandalkan dalam konteks ketahanan pangan. Aspek kesehatan yang terkandung dalam gizi ikan, seperti asam lemak omega-3, protein berkualitas tinggi, serta mineral dan vitamin, menjadikannya kontributor utama dalam memenuhi kebutuhan nutrisi masyarakat di Provinsi Jawa Barat. Ketersediaan ikan dengan harga yang terjangkau juga memainkan peran vital dalam memastikan akses terhadap sumber protein yang sehat bagi populasi yang kurang mampu secara ekonomi.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa: Penerapan algoritma *K-Means* dalam pengelompokkan jenis ikan unggulan berdasarkan kriteria pertumbuhan, mulai dari yang sangat cepat hingga yang sangat lambat, telah terbukti efektif. Hasil perhitungan dari *K-Means clustering* dapat digunakan untuk mengorganisir kelompok ikan unggulan dengan menggunakan pendekatan jarak terdekat terhadap *centroid cluster* lain. Setiap klaster menunjukkan distribusi data yang berbeda, yaitu *cluster* 0: 63 item, *cluster* 1: 9 item, *cluster* 2: 64 item, *cluster* 3: 93 item, *cluster* 4: 1 item. *items*. Klaster dengan pertumbuhan ikan sangat cepat berada di *cluster* 2 yaitu jenis ikan lele, patin, sepat siam, udang krosok, udang windu, udang danau, rumput laut, kepiting bakau sedangkan pertumbuhan ikan yang sangat lambat ada di *cluster* 1 yaitu jenis ikan tambakan. Masyarakat di Provinsi Jawa Barat mengetahui jenis ikan yang memiliki gizi memadai, harganya relatif murah dan lama pemeliharaan yang paling cepat ataupun lambat.

Diperlukan kelanjutan penelitian yang bertujuan untuk memperbaiki serta memperluas penerapan algoritma ini dalam lingkup yang lebih luas. Pentingnya menjalankan penelitian yang lebih komprehensif dengan mempertimbangkan penggunaan metode lain sebagai pembanding terhadap data yang dihasilkan dari setiap metode penelitian. Dengan demikian, dapat diperoleh pemahaman yang lebih mendalam tentang keefektifan algoritma ini dalam memahami dan mengelompokkan jenis ikan unggulan berdasarkan kriteria pertumbuhan, serta meningkatkan keakuratan hasil penelitian untuk mendukung pengembangan praktik perikanan yang lebih efisien dan berkelanjutan.

DAFTAR PUSTAKA

[1] R. Hablum, A. Khairan, and R. Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.

[2] A. Firlansyah, A. A. N. Risal, F. Adiba, and A. B. Kaswar, "Clustering Produksi Perikanan Budidaya Laut Berdasarkan Provinsi Menggunakan Algoritma K-Means," *J. Embed. Syst. Secur. Intell. Syst.*, vol. 2, no. 1, p. 55, 2021, doi: 10.26858/jessi.v2i1.20378.

[3] H. U. Sari, A. P. Windarto, and D. Hartama, "Analisa Metode Data Mining Pada Produksi Perikanan Laut Yang Dijual Di Tempat Perikanan Ikan (Tpi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 630–636, 2019, doi: 10.30865/komik.v3i1.1671.

[4] Y. Febriani, Y. P. Sari, and D. Octaria, "Metode K-Means Cluster Untuk Mengelompokkan Kota/Kabupaten di Sumatera Selatan Berdasarkan Produksi Ikan Air Tawar," *Sainmatika J. Ilm. Mat. dan Ilmu Pengetah. Alam*, vol. 18, no. 2, p. 175, 2021, doi: 10.31851/sainmatika.v18i2.6722.

[5] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," *Semin. Nas. Inov. Teknol.*, no. 2549–7952, pp. 285–291, 2020.

[6] T. M. M. Tyas and A. I. Purnamasari, "Penerapan Algoritma K-means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 277–283, 2023, doi: 10.56211/blendsains.v1i4.231.

[7] L. Darwan Ali, A. Maulana, and K. Nur Akbar, "Penerapan Clustering Menggunakan Algoritma K-Means Sebagai Analisis Produksi Komoditas Perikanan Provinsi di Indonesia," *EJECTS E-Journal Comput. Technol. Informations Syst.*, vol. 01, no. 01, pp. 1–6, 2021.

[8] M. Anjelita, A. P. Windarto, A. Wanto, and I. Sudahri, "Pengembangan Datamining Klastering Pada Kasus Pencemaran Lingkungan Hidup," *Semin. Nas. Teknol. Komput. Sains* , pp. 309–313, 2020.

[9] I. Ramadhani and M. Megawati, "Implementasi Algoritma K-Means Untuk Klustering Data Produktivitas Kelapa Sawit: Implementation Of K-Means Algorithm For Palm Oil Productivity Data Clustering," *Indones. J. Inform. …*, vol. 3, no. 1, pp. 56–64, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/ijirse/article/view/488%0Ahttps://journal.irpi.or.id/index.php/ijirse/article/download/488/261

[10] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.

[11] D. Bahtiar *et al.*, "Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering," *Sci. Sacra J. Sains*, vol. 3, no. 2, pp. 29–39, 2023, [Online]. Available: http://pijarpemikiran.com/index.php/Scientia

[12] P. R. Wulandhari, N. Rahaningsih, I. Ali, and C. L. Rohmat, "IMPLEMENTASI DATA MINING DALAM PERENCANAAN PERSEDIAAN OBAT," vol. 7, no. 1, 2023.