

BAB IV

HASIL DAN PEMBAHASAN

4.1. Seleksi Data

Tahapan seleksi data adalah proses penting dalam mengumpulkan data yang akan digunakan untuk melakukan clustering. Setelah data diperoleh, langkah selanjutnya adalah menyeleksi data yang layak untuk digunakan dalam penelitian ini. Pada penelitian ini, data sampel yang digunakan sebanyak 60, karena dari total data yang tersedia, hanya 60 data yang memenuhi kriteria yang sesuai dengan tujuan penelitian. Data yang diperoleh juga terbatas pada jumlah tersebut, sehingga hanya 60 data sampel yang dapat dianalisis lebih lanjut.

Tabel 4. 1. Data Sampel Penelitian

Nama Lengkap	Status Tempat Tinggal	Pendapatan	Pekerjaan
Agus Supriadi	10	2000000	5
Ahmad Fikri Pratama	10	2000000	10
Ahmad Syafii	10	4000000	5
Aisyah Maulani	10	2000000	10
Aldi Santoso	10	2000000	10
Andre Yudha	10	2000000	10
Ani Widiанти Purnama	10	6000000	5
Arfan Pradana	10	2000000	10
Arif Hidayat	10	6000000	5
Ayu Permata Sari	10	4000000	10
Bayu Saputra	10	2000000	5
Bella Nuryanti	10	2000000	5
Budi Santoso	10	2000000	10
Budi Santoso Utomo	10	2000000	10
Citra Anggraeni	10	2000000	10
Dedi Firmansyah	10	6000000	5
Desi Andriani	10	2000000	10

Dian Septian Nugroho	10	2000000	5
Doni Prasetyo	10	4000000	10
Dony Ardiansyah	10	4000000	5
Eka Putri Dewi	10	2000000	10
Eko Saputra Perdana	5	4000000	10
Eko Susilo	10	4000000	10
Endang Susilawati	10	4000000	10
Fahmi Pratama	5	2000000	10
Faisal Ramadhan	5	4000000	10
Fajar Pratama	10	4000000	10
Farida Hanum	10	4000000	10
Fitri Handayani	10	4000000	10
Hendra Saputra	10	6000000	5
Heri Gunawan	10	2000000	10
Indah Lestari	5	4000000	10
Intan Permatasari	5	2000000	10
Iqbal Maulana	10	4000000	5
Irfan Maulana Hakim	10	4000000	10
Irwan Maulana	10	2000000	10
Laras Setyowati	10	2000000	5
Lilis Kartika	5	2000000	10
Maya Indah Cahyani	5	2000000	10
Maya Puspita	10	2000000	10
Nanda Rizky	5	2000000	10
Nurul Hidayati	10	6000000	5
Rani Apriani	10	2000000	5
Rian Putra	5	4000000	10
Rika Mariska	10	4000000	5
Riko Ardian	10	2000000	10
Rina Anggraini	10	2000000	10
Rina Fitriani Siregar	10	2000000	10
Rini Amalia	5	4000000	10
Rizki Putra Ramadhan	10	4000000	5
Rudi Kurniawan	5	2000000	10
Sari Andini	10	4000000	5
Siti Aisyah Wulandari	10	4000000	10
Siti Nurhaliza	10	2000000	10

Taufik Rahman	10	4000000	10
Tika Anjani	10	4000000	10
Wahyu Triyono	5	4000000	10
Widya Lestari	10	4000000	10
Wulan Rahmawati	10	2000000	10
Yuni Kurniawati	10	6000000	5

4.2. Perancangan Model *Clustering*



File

Gambar 4. 1. Widget File

Widget File yang ditunjukkan dalam gambar di atas adalah komponen antarmuka yang memungkinkan pengguna untuk mengunggah atau memilih file dari perangkat mereka ke dalam suatu aplikasi atau sistem. Biasanya, widget ini berbentuk ikon atau tombol dengan gambar file, yang jika diklik akan membuka dialog pemilihan file. Setelah pengguna memilih file yang diinginkan, widget ini akan memproses atau mengunggah file tersebut untuk keperluan lebih lanjut, seperti penyimpanan, analisis, atau pembagian di dalam aplikasi. Widget seperti ini sangat umum digunakan dalam sistem pengolahan data dan aplikasi web untuk memudahkan interaksi antara pengguna dan sistem dalam hal berbagi berkas.



Gambar 4. 2. Widget Data Tabel untuk Data

Widget Data Table yang ditampilkan pada gambar di atas digunakan untuk menampilkan data dalam bentuk tabel, memungkinkan pengguna untuk melihat, mengelola, dan menganalisis informasi dengan cara yang lebih terstruktur. Jika file sudah diunggah melalui widget File, widget Data Table dapat digunakan untuk membuka dan menampilkan isi file tersebut, seperti data dalam format CSV, Excel, atau lainnya, ke dalam tampilan tabel. Dengan begitu, pengguna bisa lebih mudah memahami pola atau struktur data, melakukan pengurutan, penyaringan, serta manipulasi data untuk keperluan lebih lanjut dalam aplikasi atau analisis.



Gambar 4. 3. Widget Metode *K-Means*

Widget *K-Means* digunakan sebagai metode clustering untuk mengelompokkan data kesejahteraan masyarakat ke dalam beberapa kelompok

berdasarkan karakteristik tertentu, seperti pendapatan, pendidikan, dan akses layanan. Dalam konteks analisis kesejahteraan masyarakat, widget ini memproses data dan mempartisiny ke dalam cluster, di mana setiap cluster mewakili tingkat kesejahteraan yang serupa. Widget *K-Means* memanfaatkan algoritma *K-Means* untuk menghitung pusat cluster dan mengelompokkan data berdasarkan jarak terdekat ke pusat-pusat tersebut. Dengan menggunakan widget ini, pengguna dapat memahami pola kesejahteraan masyarakat, mengidentifikasi kelompok-kelompok yang memerlukan intervensi, dan membuat keputusan yang lebih tepat berdasarkan hasil clustering yang dihasilkan.



Gambar 4. 4. Widget *Predictions*

Widget *Predictions* digunakan untuk memberikan hasil prediksi berdasarkan model clustering yang telah dibuat, seperti model *K-Means*, dalam menganalisis tingkat kesejahteraan masyarakat. Setelah proses clustering selesai dan model terlatih, widget ini dapat memprediksi kelompok kesejahteraan masyarakat baru berdasarkan data input yang dimasukkan. Misalnya, jika terdapat data baru tentang pendapatan, pendidikan, atau akses layanan, widget *Predictions* akan menggunakan model yang telah dilatih untuk menentukan ke dalam cluster mana data tersebut masuk. Dengan demikian, widget ini membantu memproyeksikan

tingkat kesejahteraan masyarakat dan dapat digunakan untuk pengambilan keputusan kebijakan yang lebih efektif berdasarkan prediksi yang dihasilkan.



Scatter Plot

Gambar 4. 5. Widget *Scatter Plot*

Widget *Scatter Plot* digunakan untuk memvisualisasikan hasil clustering tingkat kesejahteraan masyarakat dengan menampilkan data dalam bentuk titik-titik di diagram dua dimensi. Setiap titik mewakili individu atau kelompok data, dan posisi titik ditentukan oleh nilai-nilai dari dua variabel yang dipilih, seperti pendapatan dan pendidikan. Warna atau bentuk titik-titik tersebut biasanya digunakan untuk menunjukkan kelompok atau cluster yang terbentuk dari proses clustering, seperti menggunakan metode *K-Means*. Dengan *Scatter Plot*, pengguna dapat dengan mudah melihat pola distribusi data, mengevaluasi keseragaman dalam cluster, dan mengidentifikasi anomali atau tren dalam tingkat kesejahteraan masyarakat.



Data Table

Gambar 4. 6. Widget *Data Tabel* untuk Hasil

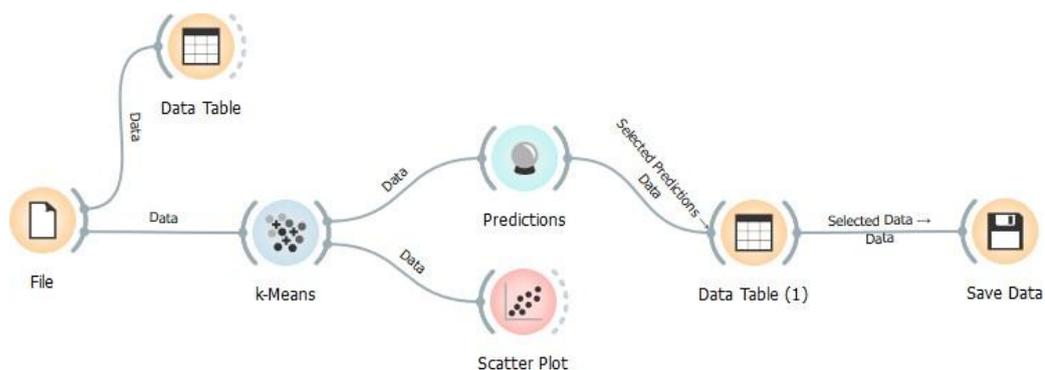
Widget *Data Table* digunakan untuk menampilkan hasil clustering dalam format yang mudah dibaca dan dianalisis. Setelah proses clustering selesai, hasilnya dapat disimpan dalam bentuk file, seperti CSV atau Excel, yang kemudian dapat dibuka menggunakan widget ini. Widget *Data Table* akan menampilkan hasil clustering, seperti kelompok-kelompok yang terbentuk berdasarkan atribut yang diolah dalam proses clustering. Pengguna dapat dengan mudah memeriksa hasil pengelompokan setiap data, mengurutkan data, serta melakukan analisis lebih lanjut, seperti validasi hasil atau mencari pola yang relevan di dalam dataset.



Gambar 4. 7. Widget *Save File*

Widget *Save File* pada aplikasi Orange digunakan untuk menyimpan hasil clustering yang telah diperoleh ke dalam file dengan format yang diinginkan, seperti CSV, Excel, atau format lain. Setelah proses clustering selesai, widget ini memungkinkan pengguna untuk menyimpan data hasil, termasuk label cluster dan atribut lainnya, ke dalam file yang dapat diakses dan digunakan untuk analisis lebih lanjut di luar aplikasi Orange. Fitur ini sangat berguna untuk mendokumentasikan dan membagikan hasil penelitian, atau untuk keperluan laporan dan pengolahan data lanjutan.

Pada tahap perancangan model cluster, penulis akan membuat sebuah model menggunakan aplikasi Orange untuk melakukan clustering data. Model ini dirancang dengan tujuan untuk mengelompokkan data ke dalam cluster yang relevan berdasarkan kesamaan karakteristik tertentu. Dalam prosesnya, penulis akan mempertimbangkan berbagai parameter dan metode clustering yang tersedia di Orange untuk memastikan model yang dihasilkan memiliki akurasi dan efisiensi yang optimal. Model ini nantinya akan menjadi alat utama dalam analisis data, memungkinkan penulis untuk mengidentifikasi pola-pola tertentu dalam dataset yang besar dan kompleks.



Gambar 4. 8. Perancangan Model Clustering

Pada gambar diatas merupakan model yang akan digunakan untuk melakukan clustering data pada *Data Mining*. Dengan adanya model diatas, proses clustering data dapat dilakukan dengan baik. Gambar yang ditunjukkan adalah alur proses analisis clustering menggunakan metode *K-Means*. Setiap widget dalam diagram ini memiliki peran penting dalam keseluruhan proses analisis data. Berikut penjelasan untuk masing-masing widget: File: Widget ini digunakan untuk memuat dataset yang akan dianalisis. Dataset ini dapat berasal dari berbagai format file seperti CSV, Excel, atau format data lainnya. Data yang

dimuat akan digunakan sebagai input untuk proses lebih lanjut. Data Table: Widget ini berfungsi untuk menampilkan data yang telah dimuat dari file, sehingga Anda dapat memeriksa dan memvalidasi data tersebut sebelum melanjutkan ke tahap berikutnya. Data yang ditampilkan di sini juga akan menjadi input untuk algoritma *K-Means*. *K-Means*: Ini adalah widget utama yang melakukan proses clustering. Algoritma *K-Means* akan mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan karakteristik antar data. Hasil dari proses clustering ini kemudian akan diteruskan ke widget berikutnya untuk dianalisis lebih lanjut. Predictions: Setelah proses clustering, widget ini digunakan untuk melakukan prediksi berdasarkan model yang telah dibuat oleh algoritma *K-Means*. Prediksi ini akan memberikan informasi mengenai cluster mana setiap data masuk. Scatter Plot: Widget ini memungkinkan visualisasi data berdasarkan hasil clustering. Scatter Plot akan menampilkan data dalam bentuk plot dua dimensi, di mana Anda dapat melihat bagaimana data dikelompokkan ke dalam cluster yang berbeda. Ini membantu dalam memverifikasi apakah clustering yang dilakukan sudah sesuai atau perlu disesuaikan. Data Table (1): Setelah prediksi dilakukan, widget ini menampilkan hasil prediksi dalam bentuk tabel. Anda dapat melihat data asli beserta informasi mengenai cluster yang diprediksi untuk setiap data. Save Data: Widget ini digunakan untuk menyimpan hasil akhir dari analisis, termasuk data yang telah dikelompokkan dan informasi prediksi. Data yang disimpan ini dapat digunakan untuk analisis lebih lanjut atau sebagai referensi di masa mendatang. Secara keseluruhan, alur ini menggambarkan proses lengkap

dari memuat data, melakukan clustering dengan *K-Means*, visualisasi hasil, hingga menyimpan output akhir.

4.3. Hasil Cluster

Pada hasil cluster yang dilakukan oleh penulis, hasil cluster akan disajikan dalam bentuk table. Untuk cluster nya juga terdapat C1, C2, dan C3. Untuk hasil nya dapat dilihat pada table dibawah ini.

Tabel 4. 2. Hasil Clustering

Nama Lengkap	Status Tempat Tinggal	Pendapatan	Pekerjaan	Cluster
Agus Supriadi	10	2000000	5	C2
Ahmad Fikri Pratama	10	2000000	10	C3
Ahmad Syafii	10	4000000	5	C2
Aisyah Maulani	10	2000000	10	C3
Aldi Santoso	10	2000000	10	C3
Andre Yudha	10	2000000	10	C3
Ani Widianti Purnama	10	6000000	5	C2
Arfan Pradana	10	2000000	10	C3
Arif Hidayat	10	6000000	5	C2
Ayu Permata Sari	10	4000000	10	C3
Bayu Saputra	10	2000000	5	C2
Bella Nuryanti	10	2000000	5	C2
Budi Santoso	10	2000000	10	C3
Budi Santoso Utomo	10	2000000	10	C3
Citra Anggraeni	10	2000000	10	C3
Dedi Firmansyah	10	6000000	5	C2
Desi Andriani	10	2000000	10	C3
Dian Septian Nugroho	10	2000000	5	C2
Doni Prasetyo	10	4000000	10	C3
Dony Ardiansyah	10	4000000	5	C2
Eka Putri Dewi	10	2000000	10	C3
Eko Saputra Perdana	5	4000000	10	C1
Eko Susilo	10	4000000	10	C3
Endang Susilawati	10	4000000	10	C3

Fahmi Pratama	5	2000000	10	C1
Faisal Ramadhan	5	4000000	10	C1
Fajar Pratama	10	4000000	10	C3
Farida Hanum	10	4000000	10	C3
Fitri Handayani	10	4000000	10	C3
Hendra Saputra	10	6000000	5	C2
Heri Gunawan	10	2000000	10	C3
Indah Lestari	5	4000000	10	C1
Intan Permatasari	5	2000000	10	C1
Iqbal Maulana	10	4000000	5	C2
Irfan Maulana Hakim	10	4000000	10	C3
Irwan Maulana	10	2000000	10	C3
Laras Setyowati	10	2000000	5	C2
Lilis Kartika	5	2000000	10	C1
Maya Indah Cahyani	5	2000000	10	C1
Maya Puspita	10	2000000	10	C3
Nanda Rizky	5	2000000	10	C1
Nurul Hidayati	10	6000000	5	C2
Rani Apriani	10	2000000	5	C2
Rian Putra	5	4000000	10	C1
Rika Mariska	10	4000000	5	C2
Riko Ardian	10	2000000	10	C3
Rina Anggraini	10	2000000	10	C3
Rina Fitriani Siregar	10	2000000	10	C3
Rini Amalia	5	4000000	10	C1
Rizki Putra Ramadhan	10	4000000	5	C2
Rudi Kurniawan	5	2000000	10	C1
Sari Andini	10	4000000	5	C2
Siti Aisyah Wulandari	10	4000000	10	C3
Siti Nurhaliza	10	2000000	10	C3
Taufik Rahman	10	4000000	10	C3
Tika Anjani	10	4000000	10	C3
Wahyu Triyono	5	4000000	10	C1
Widya Lestari	10	4000000	10	C3
Wulan Rahmawati	10	2000000	10	C3
Yuni Kurniawati	10	6000000	5	C2

Untuk tabel diatas merupakan hasil cluster yang sudah dilakukan. Hasil clustering dari 60 data sampel menunjukkan bahwa data tersebut terbagi ke dalam tiga cluster yang berbeda. Sebanyak 12 sampel tergabung dalam Cluster C1, yang mungkin mewakili kelompok dengan karakteristik atau pola tertentu yang cukup berbeda dari yang lain. Cluster C2 mencakup 18 sampel, menunjukkan bahwa kelompok ini memiliki karakteristik yang sedikit lebih umum dibanding C1. Sedangkan Cluster C3, yang terdiri dari 30 sampel, merupakan kelompok terbesar, menandakan bahwa karakteristik yang diidentifikasi dalam cluster ini mungkin lebih umum atau dominan di antara data yang dianalisis. Pembagian ini memberikan gambaran tentang distribusi karakteristik dalam data yang dapat membantu dalam analisis lebih lanjut atau dalam pengambilan keputusan berbasis data.