BAB II
LANDASAN TEORI

2.1. Kepuasan Masyarakat

Kepuasan masyarakat merupakan bentuk evaluasi subjektif yang menggambarkan sejauh mana pelayanan publik yang diterima mampu memenuhi atau bahkan melampaui harapan masyarakat [4]. Dalam konteks pelayanan publik, kepuasan ini tidak hanya menjadi refleksi dari kualitas pelayanan yang diberikan, tetapi juga menjadi indikator penting dalam menilai keberhasilan lembaga publik dalam menjalankan fungsinya [5]. Pelayanan yang berkualitas dapat dilihat dari beberapa aspek seperti kecepatan, kemudahan prosedur, sikap petugas, kebersihan lingkungan, hingga ketersediaan fasilitas. Oleh karena itu, pelayanan publik yang efektif harus mampu menjawab kebutuhan masyarakat secara optimal agar tercipta kepuasan yang merata.

Untuk mencapai pelayanan yang sesuai dengan standar atau bahkan melebihi ekspektasi masyarakat, diperlukan sistem evaluasi yang objektif dan terstruktur. Dalam hal ini, keterlibatan masyarakat dalam memberikan penilaian terhadap pelayanan menjadi sangat penting sebagai dasar ukur kinerja instansi publik. Penelitian ini dilakukan untuk mengetahui tingkat kepuasan masyarakat terhadap pelayanan di Kantor Camat Silangkitang, Kabupaten Labuhanbatu Selatan, dengan menggunakan teknik pengumpulan data melalui angket atau kuesioner. Data yang diperoleh kemudian dianalisis menggunakan metode klasifikasi algoritma C4.5 melalui bantuan aplikasi Orange. Pendekatan ini diharapkan dapat mengidentifikasi faktor-faktor yang paling berpengaruh terhadap tingkat kepuasan masyarakat serta

memberikan masukan konstruktif bagi peningkatan kualitas pelayanan ke depannya.

2.2. Indikator Kepuasan Masyarakat

Indikator kepuasan masyarakat di Kantor Camat merupakan elemen krusial dalam menilai sejauh mana kualitas pelayanan publik yang diberikan telah memenuhi harapan warga. Dalam penelitian ini, digunakan lima indikator utama yang berperan dalam mengukur kepuasan masyarakat, yakni Kecepatan Pelayanan, Kemudahan Prosedur, Sikap Petugas, Kebersihan, Ketersediaan Fasilitas, serta relevansi layanan terhadap Keperluan masyarakat. Hasil temuan menunjukkan bahwa meskipun sebagian besar indikator telah terpenuhi dengan baik, masih terdapat celah yang perlu mendapat perhatian lebih, khususnya dalam aspek fisik pelayanan dan kehandalan layanan. Salah satu kelemahan yang teridentifikasi adalah minimnya fasilitas untuk menampung pengaduan, seperti absennya kotak saran atau sistem pelaporan yang mudah diakses masyarakat. Kondisi ini mengindikasikan bahwa selain pelayanan langsung, aksesibilitas untuk menyampaikan keluhan atau masukan juga memiliki peran penting dalam membentuk persepsi kepuasan masyarakat secara menyeluruh.

Lebih lanjut, keberadaan Indeks Kepuasan Masyarakat (IKM) menjadi alat ukur yang relevan untuk menilai keberhasilan penyelenggaraan pelayanan publik. Penelitian ini menegaskan bahwa kualitas pelayanan yang optimal dan citra positif dari Kantor Camat Silangkitang mampu memberikan kontribusi signifikan terhadap tingkat kepuasan masyarakat. Citra yang baik dari institusi publik, terutama di tingkat kecamatan, mampu memperkuat kepercayaan masyarakat dan mendorong

partisipasi aktif dalam pembangunan wilayah. Oleh karena itu, inovasi dalam pelayanan menjadi hal yang mutlak diperlukan, seperti penerapan sistem digital dalam pengurusan dokumen atau penyediaan aplikasi pengaduan online yang mudah diakses oleh warga. Dengan menggabungkan kualitas layanan yang unggul, keterbukaan terhadap masukan, serta inovasi yang adaptif terhadap perkembangan teknologi, diharapkan Kantor Camat mampu menciptakan pelayanan publik yang lebih responsif dan berorientasi pada kepuasan masyarakat secara berkelanjutan.

2.3. Data Mining

Data mining merupakan suatu himpunan teknik dan proses sistematis yang digunakan untuk menemukan pola-pola tersembunyi, valid, baru, berguna, dan dapat dimengerti dari kumpulan data dalam jumlah besar [6]. Tujuan utama dari proses ini adalah untuk mengekstraksi informasi yang bermakna dan bernilai dari basis data guna mendukung pengambilan keputusan yang lebih tepat. Tahapan dalam data mining meliputi pemilihan data (data selection), pembersihan dan praproses data (data preprocessing), transformasi data, penerapan teknik data mining itu sendiri, serta interpretasi dan evaluasi hasil temuan pola [7]. Teknik yang umum digunakan dalam data mining antara lain klasifikasi, clustering, asosiasi, dan regresi. Penerapan data mining sudah banyak digunakan dalam berbagai bidang, seperti penentuan penerima program bantuan sosial, peramalan penjualan, analisis perilaku konsumen, hingga pengelompokan Puas mahasiswa dalam dunia pendidikan.

Dalam pelaksanaannya, data mining memanfaatkan sejumlah algoritma populer, seperti algoritma C4.5 untuk klasifikasi berbasis pohon keputusan,

algoritma Apriori untuk analisis asosiasi, algoritma K-Means untuk pengelompokan data, serta algoritma Naive Bayes untuk klasifikasi probabilistik. Pemilihan metode atau algoritma yang digunakan sangat bergantung pada jenis data dan tujuan analisis yang ingin dicapai [8]. Data mining tidak hanya berperan sebagai alat bantu teknis, melainkan telah menjadi bagian penting dalam strategi pengambilan keputusan di berbagai sektor, termasuk bisnis, pendidikan, pemerintahan, dan kesehatan. Melalui pendekatan yang tepat, data mining mampu menghasilkan wawasan yang strategis, meningkatkan efisiensi operasional, serta memberikan solusi konkret terhadap permasalahan yang kompleks.

2.3.1. Ciri-Ciri Data Mining

Data mining merupakan proses sistematis untuk memperoleh informasi yang berguna dari kumpulan data berskala besar [9]. Teknik ini berakar dari disiplin ilmu Kecerdasan Buatan (Artificial Intelligence/AI) dan Rekayasa Pengetahuan (Knowledge Engineering/KE), serta memanfaatkan pendekatan statistik dan machine learning dalam pengolahan data. Tujuan utama dari data mining adalah melakukan klasifikasi, prediksi, perkiraan, serta mengekstraksi informasi tersembunyi yang bermanfaat dari data [10]. Proses ini dapat menggunakan berbagai algoritma seperti C4.5 yang menghasilkan pohon keputusan yang mudah dipahami, algoritma Apriori untuk menemukan asosiasi antar item, serta algoritma Support Vector Machine (SVM) dalam analisis sentimen [11]. Data mining telah diaplikasikan secara luas dalam berbagai bidang seperti kesehatan, keuangan, lingkungan, dan pemerintahan. Seiring kemajuan perangkat keras, pengolahan data besar semakin efisien dan mendukung implementasi data mining. Secara umum,

data mining merupakan salah satu tahapan penting dalam proses Knowledge Discovery in Database (KDD) dan melibatkan teknik-teknik utama seperti klasifikasi, regresi, clustering, dan asosiasi [12] [13] [14].

Beberapa karakteristik utama data mining antara lain:

1. Proses Ekstraksi Informasi, yaitu mengambil data penting dari sumber tak terstruktur (teks, gambar, video) ke dalam bentuk yang terstruktur dan siap digunakan. Tahapan ekstraksi meliputi preprocessing, identifikasi entitas, ekstraksi relasi, normalisasi, dan evaluasi hasil.
2. Berbasis Algoritma, di mana algoritma menjadi dasar dalam memecahkan masalah dan membangun sistem otomatis berbasis data.
3. Skala Besar, karena data mining digunakan untuk mengolah data dengan volume besar yang cakupannya luas dan kompleks.
4. Multi-Disiplin, melibatkan kombinasi ilmu komputer, statistik, matematika, manajemen data, dan domain aplikasi lainnya untuk mengkaji permasalahan secara komprehensif.
5. Identifikasi Pola dan Tren, di mana proses ini membantu mengungkap kecenderungan, perubahan, serta struktur tersembunyi dalam data, yang pada gilirannya dapat dijadikan dasar dalam pengambilan keputusan strategis.

2.4. Knowledge

Pengetahuan (knowledge) merupakan hasil dari proses memahami, mengorganisasi, dan menginterpretasikan informasi yang diperoleh melalui pengalaman, pembelajaran, maupun penelitian [15]. Pengetahuan meliputi berbagai

jenis informasi, mulai dari fakta, konsep, prinsip, teori, hingga keterampilan, yang semuanya berfungsi sebagai dasar untuk memecahkan masalah, mengambil keputusan, atau menciptakan sesuatu yang baru [16]. Pengetahuan tidak hanya berkaitan dengan apa yang diketahui, tetapi juga menyangkut bagaimana informasi tersebut diolah dan diaplikasikan secara tepat dalam berbagai konteks. Dalam praktiknya, pengetahuan dapat dibedakan menjadi beberapa jenis, yaitu:

1. Pengetahuan Deklaratif, yang mencakup fakta dan informasi dasar, seperti pengetahuan tentang produk halal atau fungsi paru-paru.
2. Pengetahuan Prosedural, yang berkaitan dengan keterampilan melakukan sesuatu, seperti cara membuat blog atau menyampaikan informasi.
3. Pengetahuan Kontekstual, yaitu pengetahuan yang relevan dengan kondisi atau situasi tertentu, seperti pemahaman tentang pariwisata lokal atau persepsi masyarakat terhadap inovasi tertentu.

Ciri-ciri dari pengetahuan antara lain mencakup kemampuannya untuk dipelajari melalui pelatihan atau pengalaman, dapat diukur melalui evaluasi tertentu, serta dapat meningkatkan pengambilan keputusan dan performa individu [17]. Pengetahuan juga dapat dibagi berdasarkan dimensi utamanya, seperti pengetahuan faktual (tentang informasi dasar), pengetahuan konseptual (berisi teori dan konsep), pengetahuan prosedural (langkah dan teknik pelaksanaan), dan pengetahuan metakognitif (kesadaran terhadap proses berpikir sendiri) [18]. Selain itu, literasi sains yang baik juga mencakup dimensi pengetahuan (memahami konsep ilmiah), konteks (mengaitkan pengetahuan dalam kehidupan), kompetensi (kemampuan menyelidiki dan menganalisis), serta sikap (berpikir kritis dan ilmiah).

Dengan demikian, pengetahuan memiliki cakupan luas yang tidak hanya bersifat kognitif, tetapi juga aplikatif dan kontekstual sesuai dengan tantangan zaman.

2.5. Algoritma C4.5

Algoritma C4.5 merupakan salah satu metode klasifikasi yang banyak digunakan dalam data mining dan machine learning [19]. Dikembangkan oleh Ross Quinlan sebagai penyempurnaan dari algoritma ID3, C4.5 memiliki kemampuan untuk menghasilkan pohon keputusan yang efisien, akurat, serta mampu menangani berbagai tipe atribut, baik kontinu maupun diskrit [20]. Salah satu fitur unggulan dari algoritma ini adalah proses *pruning*, yaitu pemangkasan cabang-cabang pohon yang tidak memberikan kontribusi signifikan terhadap peningkatan akurasi [21]. Pruning ini sangat penting dalam mengurangi *overfitting*, yaitu kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan dan kehilangan kemampuan untuk melakukan generalisasi terhadap data baru. Selain itu, algoritma C4.5 juga mampu menangani data yang memiliki *missing value* dengan pendekatan probabilistik, menjadikannya cukup fleksibel dalam pengolahan data riil yang tidak selalu sempurna [22]. C4.5 bekerja berdasarkan penghitungan *gain ratio*, yang mengukur sejauh mana sebuah atribut mampu mengurangi entropi atau ketidakpastian dalam data, lalu membagi dataset ke dalam subset yang lebih homogen secara rekursif hingga tercapai kriteria penghentian.

Dalam praktiknya, algoritma C4.5 telah diterapkan secara luas dalam berbagai studi dan sektor, seperti klasifikasi pasien berdasarkan jenis penyakit, prediksi masa studi mahasiswa, hingga pengelompokan kepribadian berdasarkan teori psikologis tertentu. Beberapa penelitian menunjukkan bahwa penerapan C4.5

dapat meningkatkan akurasi klasifikasi dalam berbagai konteks, seperti pada sistem rekomendasi organisasi pelatihan, klasifikasi jurusan siswa di SMK, hingga penentuan penerima bantuan sosial [23]. Keunggulan utama lainnya adalah kemampuannya untuk menghasilkan aturan keputusan (*decision rules*) yang berasal dari struktur pohon, sehingga memudahkan interpretasi hasil oleh pengguna akhir. Dengan demikian, C4.5 tidak hanya berfungsi sebagai alat analisis data, tetapi juga mendukung pengambilan keputusan berbasis data (data-driven decision making). Integrasinya dengan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) juga menambah nilai praktisnya, karena menjadikannya cocok untuk proses analitik data yang sistematis dan iteratif di berbagai bidang, termasuk bisnis, pendidikan, dan pemerintahan.

Namun, meskipun C4.5 memiliki banyak keunggulan, sejumlah kelemahan tetap harus diperhatikan. Beberapa studi mencatat bahwa algoritma ini kurang efektif saat menangani data yang tidak seimbang (imbalanced data) atau yang memiliki *overlapping* antar kelas, sehingga berisiko menghasilkan model yang bias [24]. C4.5 juga sensitif terhadap keberadaan *noise* dalam data, yang dapat mengakibatkan struktur pohon yang rumit dan kurang optimal. Selain itu, ketika dihadapkan pada dataset yang sangat besar dan mengandung banyak atribut serta kelas, algoritma ini cenderung menghasilkan kompleksitas tinggi dan waktu komputasi yang lama, sehingga efisiensinya menurun dalam skenario real-time. Untuk mengatasi hal tersebut, beberapa peneliti telah mengusulkan pendekatan optimasi seperti penggunaan teknik *bagging* atau integrasi dengan *Particle Swarm Optimization (PSO)* untuk meningkatkan performa algoritma. Oleh karena itu,

meskipun C4.5 tetap menjadi algoritma yang kuat dan relevan dalam dunia klasifikasi data, pemahaman terhadap kelebihan dan keterbatasannya sangat penting untuk memastikan implementasi yang efektif dan optimal dalam berbagai aplikasi.

2.5.1. Proses Algoritma C4.5

Proses algoritma C4.5 diawali dengan tahap pembentukan pohon keputusan, di mana data dilatih menggunakan atribut dan label yang telah ditentukan untuk menghasilkan struktur pohon yang mampu merepresentasikan pola klasifikasi. Pemilihan atribut terbaik pada setiap node pohon dilakukan dengan menggunakan perhitungan Gain Ratio, yaitu rasio antara Information Gain dan entropi dari atribut tersebut. Pendekatan ini dipilih karena Gain Ratio mampu mengatasi kelemahan dari Information Gain murni, terutama dalam menghadapi atribut dengan banyak nilai. Setelah pohon keputusan terbentuk, dilakukan tahap pemangkasan (pruning) guna mengurangi kompleksitas model dan mencegah terjadinya overfitting, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data pelatihan sehingga kurang akurat saat digunakan pada data baru. Tahap akhir dari proses algoritma C4.5 adalah klasifikasi, di mana pohon keputusan yang telah terbentuk digunakan untuk memprediksi atau mengklasifikasikan data baru berdasarkan jalur keputusan yang relevan dari akar hingga daun pohon. Terdapat tahapan yang perlu dilakukan pada Algoritma C4.5 yaitu sebagai berikut.

1. Perhitungan Entropy dan Gain
    1) Menghitung nilai entropy dari seluruh dataset untuk mengetahui tingkat ketidakpastian (disorder) dalam data.

2) Kemudian menghitung information gain untuk masing-masing atribut.
3) Setelah itu, menghitung Gain Ratio, yaitu rasio antara information gain dan split information, untuk menentukan atribut terbaik sebagai node keputusan.

2. Pembuatan Pohon Keputusan
   1) Atribut dengan Gain Ratio tertinggi dipilih sebagai node (akar) utama.
   2) Dataset kemudian dibagi berdasarkan nilai dari atribut tersebut, dan proses perhitungan berlanjut secara rekursif untuk setiap cabang hingga seluruh data terklasifikasi atau atribut habis.
3. Pemangkasan (Pruning)
   1) Setelah pohon selesai dibangun, dilakukan proses pruning untuk menyederhanakan struktur pohon.
   2) Tujuannya adalah mengurangi overfitting, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data latih dan tidak general terhadap data baru.
4. Klasifikasi
   1) Pohon yang telah terbentuk digunakan untuk mengklasifikasikan data baru.
   2) Proses klasifikasi dilakukan dengan menelusuri pohon berdasarkan nilai atribut pada data uji hingga mencapai simpul daun yang menyatakan kelas prediksi.

2.5.2. Kelebihan Algoritma C4.5

Algoritma C4.5 memiliki sejumlah kelebihan yang menjadikannya salah satu metode klasifikasi yang banyak digunakan dalam data mining. Salah satu keunggulannya adalah kemampuannya untuk menangani data baik yang bersifat kategorikal maupun numerik, sehingga cocok diterapkan pada berbagai jenis atribut dalam dataset. Selain itu, C4.5 juga mampu mengatasi permasalahan data yang tidak lengkap (missing values) tanpa harus menghapus data tersebut, dengan memperkirakan distribusi nilai yang hilang secara proporsional. Dalam pemilihan atribut, algoritma ini menggunakan Gain Ratio yang merupakan pengembangan dari Information Gain pada algoritma ID3, sehingga mampu mengurangi bias terhadap atribut dengan banyak nilai. C4.5 juga dilengkapi dengan proses pemangkasan (pruning) untuk menyederhanakan struktur pohon keputusan dan menghindari overfitting, yang berarti model tetap dapat bekerja dengan baik terhadap data baru. Keunggulan lainnya terletak pada hasil model yang berbentuk pohon keputusan yang mudah dipahami dan dijelaskan dalam bentuk aturan logis, sehingga memudahkan dalam interpretasi dan pengambilan keputusan berdasarkan hasil klasifikasi.

Selain itu, kelebihan lain dari algoritma C4.5 terletak pada fleksibilitasnya dalam menangani dataset dengan banyak kelas target. Tidak seperti beberapa algoritma lain yang hanya bekerja optimal untuk klasifikasi biner, C4.5 mampu membagi data ke dalam lebih dari dua kelas tanpa memerlukan modifikasi khusus. Algoritma ini juga dapat menangani noise atau ketidakteraturan dalam data dengan cukup baik karena prinsip pemangkasan (pruning) yang digunakan mampu

mengurangi kompleksitas pohon dan mempertahankan hanya cabang-cabang yang signifikan. Keunggulan ini menjadikan C4.5 sangat berguna dalam penerapan dunia nyata, termasuk dalam sektor pelayanan publik seperti penelitian kepuasan masyarakat, karena algoritma ini tidak hanya memberikan hasil akurat, tetapi juga menyajikannya dalam bentuk yang mudah dimengerti oleh pihak non-teknis. Dengan demikian, C4.5 tidak hanya kuat secara analisis, tetapi juga unggul dalam hal keterbacaan dan kepraktisan hasil klasifikasinya.

2.5.3. Kekurangan Algoritma C4.5

Meskipun algoritma C4.5 memiliki banyak kelebihan, terdapat beberapa kekurangan yang perlu diperhatikan. Salah satu kekurangan utama adalah kompleksitas komputasinya yang cukup tinggi, terutama ketika digunakan pada dataset yang sangat besar atau memiliki banyak atribut. Proses perhitungan gain ratio untuk setiap atribut dapat memakan waktu dan sumber daya yang cukup besar, yang bisa menjadi kendala dalam sistem dengan keterbatasan performa. Selain itu, algoritma ini sensitif terhadap data yang tidak seimbang (imbalanced data), di mana kelas mayoritas bisa mendominasi hasil klasifikasi dan menyebabkan penurunan akurasi terhadap kelas minoritas.

Kelemahan lainnya terletak pada cara algoritma C4.5 menangani data numerik, karena perlu dilakukan proses pembagian nilai secara berulang hingga ditemukan titik pemisah (threshold) terbaik, yang bisa meningkatkan risiko overfitting jika tidak disertai dengan proses pruning yang tepat. Selain itu, meskipun hasil klasifikasinya disajikan dalam bentuk pohon keputusan yang visual, struktur pohon dapat menjadi sangat besar dan rumit jika dataset memiliki banyak

atribut, sehingga menyulitkan interpretasi secara menyeluruh. Oleh karena itu, meskipun C4.5 merupakan algoritma yang andal, penggunaannya harus disesuaikan dengan karakteristik data dan tujuan analisis agar tetap memberikan hasil yang optimal.

2.5.4. Integrasi Analisis Kepuasan dengan Algoritma C4.5

Integrasi analisis kepuasan masyarakat dengan algoritma C4.5 dilakukan dengan memanfaatkan keunggulan algoritma ini dalam mengolah data kategori dan menghasilkan model klasifikasi yang mudah dipahami dalam bentuk pohon keputusan. Dalam konteks pelayanan publik, seperti di Kantor Camat Silangkitang, atribut-atribut seperti kecepatan pelayanan, kemudahan prosedur, sikap petugas, kebersihan, dan ketersediaan fasilitas dijadikan variabel penentu tingkat kepuasan masyarakat. Algoritma C4.5 kemudian menganalisis setiap atribut tersebut untuk menentukan pola dan hubungan antara variabel yang memengaruhi keputusan masyarakat merasa puas atau tidak puas.

Proses ini diawali dengan tahap pelatihan model menggunakan data yang telah diklasifikasikan berdasarkan tingkat kepuasan responden. Algoritma C4.5 menghitung nilai gain dan gain ratio dari setiap atribut untuk menentukan atribut mana yang paling signifikan dalam membedakan kelas (puas atau tidak puas). Selanjutnya, model pohon keputusan dibentuk berdasarkan urutan atribut yang memiliki pengaruh terbesar, sehingga setiap cabang dalam pohon menunjukkan jalur keputusan yang logis berdasarkan data yang tersedia. Dengan cara ini, analisis tidak hanya menghasilkan prediksi, tetapi juga memberikan gambaran faktor-faktor dominan yang perlu diperbaiki oleh pihak kantor camat.

Hasil dari integrasi ini memungkinkan instansi pemerintah untuk mengambil keputusan yang lebih tepat sasaran berdasarkan data lapangan. Melalui visualisasi pohon keputusan, pihak pengelola pelayanan dapat lebih mudah memahami faktor mana yang paling berpengaruh terhadap kepuasan masyarakat, dan tindakan apa yang perlu diambil untuk meningkatkan kualitas layanan. Selain itu, model klasifikasi yang dihasilkan juga dapat digunakan untuk memantau dan mengevaluasi kepuasan secara berkala, serta sebagai dasar pengambilan kebijakan berbasis data dalam upaya mewujudkan pelayanan publik yang lebih responsif dan efisien di masa depan.

2.5.5. Langkah-Langkah Algoritma C4.5

Algoritma C4.5 merupakan salah satu metode klasifikasi berbasis pohon keputusan yang populer dalam data mining karena kemampuannya menangani data kategorikal maupun numerik, serta menghasilkan model yang mudah diinterpretasikan. C4.5 bekerja dengan membagi dataset ke dalam subset berdasarkan atribut yang memberikan informasi paling besar (information gain) dan memperhitungkan gain ratio untuk menghindari bias terhadap atribut dengan banyak nilai unik. Dengan membangun struktur pohon secara rekursif, algoritma ini mampu mengidentifikasi pola tersembunyi dalam data dan menyederhanakan proses pengambilan keputusan berdasarkan logika yang sistematis.

Berikut ini adalah langkah-langkah dalam algoritma C4.5:

1. Menentukan Atribut Target, Tentukan atribut mana yang akan dijadikan sebagai label kelas (target) dalam proses klasifikasi.

2. Menghitung Entropy Semua Atribut, Hitung entropy dari setiap atribut dalam dataset untuk mengukur tingkat ketidakpastian atau keragaman nilai pada masing-masing atribut.
3. Menghitung Gain dan Gain Ratio, Hitung nilai information gain dari masing-masing atribut, lalu hitung gain ratio untuk menyesuaikan pengaruh atribut yang memiliki banyak nilai unik.
4. Menentukan Atribut Terbaik Sebagai Node, Pilih atribut dengan gain ratio tertinggi sebagai node (akar atau cabang) dari pohon keputusan.
5. Membagi Data Menjadi Subset, Bagi data berdasarkan nilai dari atribut terbaik yang telah dipilih, dan buat cabang pohon keputusan berdasarkan nilai-nilai tersebut.
6. Rekursi untuk Setiap Subset, Ulangi proses dari langkah pertama untuk setiap subset hingga seluruh data pada cabang tersebut memiliki kelas yang sama atau tidak ada lagi atribut yang bisa dipilih.
7. Pemangkasan (Pruning), Setelah pohon selesai dibentuk, lakukan pemangkasan untuk menghilangkan cabang yang tidak signifikan atau menyebabkan overfitting, agar pohon lebih sederhana dan generalisasi model menjadi lebih baik.

Langkah-langkah ini menjadikan algoritma C4.5 sebagai alat yang efektif untuk mengklasifikasikan data serta memberikan pemahaman yang jelas terhadap pola-pola yang tersembunyi dalam kumpulan data besar.

2.6. Model Klasifikasi

Model klasifikasi merupakan salah satu pendekatan dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang ditemukan dari data historis. Model ini bekerja dengan mempelajari hubungan antar atribut dalam data training dan kemudian membentuk aturan atau pohon keputusan yang dapat digunakan untuk memprediksi kelas dari data baru. Tujuan utama dari klasifikasi adalah untuk membangun model prediktif yang mampu mengklasifikasikan data secara akurat ke dalam kelas-kelas yang telah ditentukan. Berbagai algoritma klasifikasi seperti Decision Tree (C4.5), Naive Bayes, dan K-Nearest Neighbor (KNN) umum digunakan tergantung pada jenis data dan tujuan analisis.

Dalam penelitian ini, model klasifikasi diterapkan untuk mengelompokkan tingkat kepuasan masyarakat terhadap pelayanan di Kantor Camat Silangkitang. Dengan menggunakan algoritma C4.5, data yang telah dikumpulkan dari masyarakat dianalisis untuk menemukan pola yang berkontribusi terhadap kepuasan atau ketidakpuasan mereka. Hasil klasifikasi ini tidak hanya memberikan gambaran objektif mengenai kualitas pelayanan publik, tetapi juga membantu pihak kecamatan dalam menentukan area yang perlu diperbaiki. Dengan demikian, model klasifikasi berperan penting dalam mendukung pengambilan keputusan berbasis data dan meningkatkan kualitas layanan secara berkelanjutan.

2.7. Alat Bantu Program Aplikasi Orange

Aplikasi Orange merupakan salah satu alat bantu analisis data berbasis grafis yang sangat populer dalam dunia data mining dan machine learning. Program ini bersifat open-source dan menyediakan antarmuka visual (visual programming)

yang memudahkan pengguna, terutama yang tidak memiliki latar belakang pemrograman, untuk membangun dan menguji berbagai model analisis data. Orange dilengkapi dengan berbagai widget yang dapat digunakan untuk melakukan proses seperti preprocessing data, visualisasi, klasifikasi, regresi, clustering, dan evaluasi model. Keunggulan utama Orange adalah kemudahan drag-and-drop dalam menyusun alur kerja analisis data, serta tersedianya berbagai algoritma populer seperti Decision Tree, Naive Bayes, KNN, dan lainnya yang dapat digunakan secara instan.

Dalam penelitian mengenai analisis tingkat kepuasan masyarakat terhadap pelayanan di Kantor Camat Silangkitang, aplikasi Orange digunakan sebagai alat bantu utama dalam membangun dan mengevaluasi model klasifikasi berbasis algoritma C4.5 (Decision Tree). Penggunaan Orange memungkinkan peneliti untuk menyusun alur kerja analisis secara sistematis, mulai dari input data, proses preprocessing, pembuatan model klasifikasi, hingga evaluasi hasil dengan lebih efisien dan akurat. Dengan tampilan visual yang intuitif, Orange membantu peneliti untuk memahami pola-pola dalam data masyarakat yang memengaruhi tingkat kepuasan, serta mempermudah dalam penyajian hasil analisis dalam bentuk pohon keputusan yang mudah dipahami oleh pihak terkait.

2.8. Evaluasi Model Algoritma C4.5

Evaluasi model pada algoritma C4.5 merupakan proses penting untuk menilai sejauh mana model yang dibangun mampu melakukan klasifikasi data secara akurat dan andal. Evaluasi ini biasanya dilakukan dengan menggunakan metrik seperti akurasi, presisi, dan recall yang dihasilkan dari confusion matrix. Akurasi

menunjukkan seberapa besar persentase prediksi yang benar dari keseluruhan data, presisi menggambarkan seberapa tepat model dalam memprediksi kelas positif, sedangkan recall mengukur kemampuan model dalam menemukan seluruh data yang termasuk dalam kelas positif. Selain itu, evaluasi dapat menggunakan metode validasi silang (cross-validation) atau membagi data ke dalam set pelatihan dan pengujian, untuk memastikan bahwa model tidak hanya bekerja baik pada data pelatihan, tetapi juga dapat digeneralisasi pada data baru.

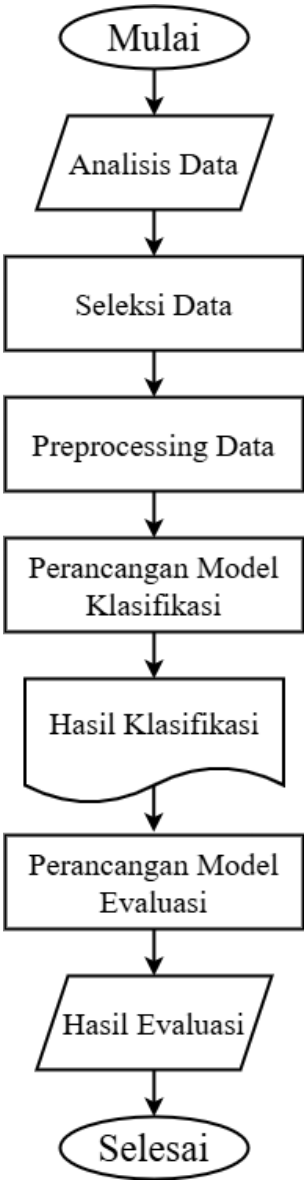
Dalam penelitian mengenai tingkat kepuasan masyarakat terhadap pelayanan di Kantor Camat Silangkitang, evaluasi model C4.5 dilakukan terhadap data yang telah diklasifikasikan berdasarkan atribut-atribut seperti kecepatan pelayanan, kemudahan prosedur, sikap petugas, kebersihan, dan ketersediaan fasilitas. Hasil evaluasi menunjukkan bahwa model memiliki akurasi yang tinggi, menunjukkan bahwa pohon keputusan yang dibentuk mampu mengidentifikasi pola dari data dengan baik. Hal ini membuktikan bahwa C4.5 efektif digunakan dalam menganalisis data survei masyarakat, serta memberikan kontribusi yang signifikan dalam mendukung proses pengambilan keputusan guna meningkatkan kualitas pelayanan publik di wilayah tersebut.

2.9. Kelebihan Penelitian

Penelitian ini memiliki sejumlah kelebihan yang signifikan, terutama dalam penerapan algoritma C4.5 yang mampu menghasilkan model klasifikasi berbasis pohon keputusan yang mudah dipahami dan diinterpretasikan. Dengan pendekatan visual melalui aplikasi Orange, proses analisis data menjadi lebih interaktif dan transparan, sehingga memudahkan dalam mengidentifikasi faktor-faktor utama

yang memengaruhi tingkat kepuasan masyarakat terhadap pelayanan publik di Kantor Camat Silangkitang. Selain itu, penggunaan data nyata dari responden masyarakat membuat hasil analisis bersifat aplikatif dan representatif terhadap kondisi di lapangan, sehingga rekomendasi yang dihasilkan lebih relevan dan dapat diimplementasikan secara praktis.

2.10. Kerangka Kerja Penelitian



Gambar 2. 1. Kerangka Kerja Penelitian

Berikut adalah penjelasan dari masing masing poin yang ada pada kerangka kerja penelitian yaitu sebagai berikut.

1. Analisis Data, Analisis data dilakukan untuk memahami struktur dan karakteristik data yang dikumpulkan dari masyarakat pengguna layanan di Kantor Camat Silangkitang. Tahapan ini penting untuk menentukan atribut-atribut yang relevan dalam mengukur tingkat kepuasan masyarakat.
2. Seleksi Data, Seleksi data merupakan proses pemilihan data yang sesuai dan layak digunakan untuk penelitian. Data yang tidak relevan atau tidak lengkap akan diabaikan agar tidak mengganggu akurasi hasil analisis.
3. Preprocessing Data, Preprocessing data bertujuan untuk membersihkan dan mempersiapkan data sebelum dianalisis, seperti menghapus nilai kosong atau data duplikat. Dengan preprocessing yang baik, kualitas data akan lebih terjamin sehingga hasil analisis lebih valid.
4. Perancangan Model Klasifikasi, Tahapan ini dilakukan untuk membangun model klasifikasi menggunakan algoritma C4.5 dengan bantuan aplikasi Orange. Model ini digunakan untuk mengidentifikasi pola dari atribut-atribut yang memengaruhi kepuasan masyarakat.
5. Hasil Klasifikasi, Hasil klasifikasi menunjukkan prediksi tingkat kepuasan masyarakat berdasarkan model yang telah dibangun. Data akan dikelompokkan ke dalam kategori puas dan tidak puas sesuai dengan pola yang ditemukan oleh algoritma.
6. Perancangan Model Evaluasi, Perancangan model evaluasi dilakukan untuk menilai sejauh mana model klasifikasi mampu bekerja secara optimal.

Evaluasi ini dilakukan dengan menghitung metrik seperti akurasi, presisi, dan recall.

7. Hasil Evaluasi, Hasil evaluasi memberikan gambaran objektif mengenai performa model klasifikasi dalam memprediksi tingkat kepuasan masyarakat. Nilai akurasi dan metrik lainnya menjadi dasar untuk menilai keefektifan metode yang digunakan.