BAB II

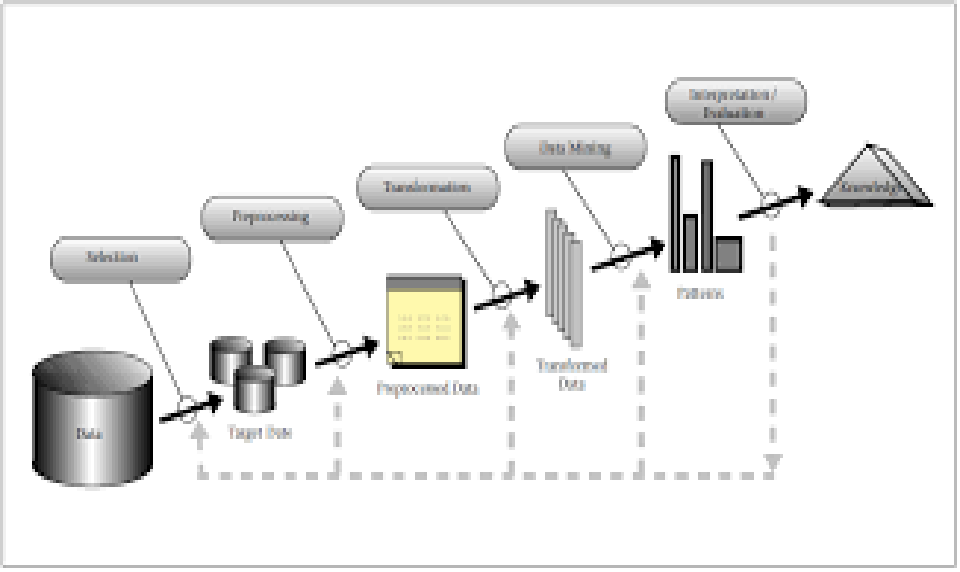
LANDASAN TEORI

2.1 know Discovery in Database (KDD)

Know discovery in Database (KDD) Merupakan salah satu tahapan Penting didalam proses knowledge Discovery in Database (KDD).Terminologi dari kdd dan data mining adalah berbeda. KDD adalah keseluruhan proses di dalam menemukan pengetahuan yang berguna dari suatu kumpulan data sedangkan data mining adlah salah satu tahapan pada KDD dan fokus pada upaya untuk menemukan pengetahuan yang berguna yang berguna dengan menggunakan algoritma. Prosedur penelitian menggunakan tahapan-tahapan knowledge discovery in Database (KDD). Tahapannya diantaranya adalah : selection (Menyeleksi data yang relevan), Prepocessing (menghilangkan noise dan inkosisten data; menggunakan data yang bersumber dari banyak sumber), transformation (Mentraspormasi data dalam bentuk yang sesuai untuk proses data mining), data mining(Memilih algoritma data mining yang sesuai dengan pattern data; Ekstransi pola dari data), Interpretation/evalution (menginterpretasi pola menjadi pengetahuan yang Menghilangkan pola yang redundant dan tidak relewan) Proses KDD yang ada dapat lihat pada gambar 2.1 (Hartama dedi dan hartono, 2016)

Pada data proses kdd Merupakan Teori Rough set dikembangkan oleh zdzislaw Pada awal tahun delapan puluhan. Teori ini muncul karena adanya Rough pada sesuatu himpunan, dimana dalam terori Rough set , data dapat

direpresentasikan dalam dua system yaitu system informasi dan system keputusan yang ada pada(Taufik Hidayat 2024).



Gambar 2.1 Proses dalam knowledge data discovery

Berdasarkan Gambar diatas didefinisikan merupakan salah satu bagian proses knowledge discovery in Database (KDD) yang Bertugas untuk mengekstrak pola atau model dari data dengan menngunakan suatu algoritma yang sepesifik. Adapun proses KDD sebagai Berikut (Maukar et al., 2022):

Data selection: pemilihan data dari sekumpulan data opresional perlu dilakukan sebelum tahapan penggalian informasi dalam KDD dimulai.

Preprocetion: sebelum Proses data mining dapat dilakukan, Perlu dilakukan proses cleaning dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkosisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan Proses enrichment, yaitu proses " memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan di perlukan

untuk KDD, seperti data atau informasi.

Transformasi : yaitu proses coding pada data yang lebih di pilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses keratif dan sangat tergantung pada jenis untuk pola informasi yang akan dicari dalam database

Data mining : proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

Interpretation /evalution : pola informasi yang dihasilkan dari proses data mining perlu di tampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahapan ini merupakan bagian dari proses KDD yang disebut dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentengan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

Sehingga dapat diketahui apa itu KDD sendiri dari kutipan penelitian yang di ambil dari berdasarkan jurnal yang dikutip seperti yang di ketahui yaituknowledge discovery in database (KDD) adalah proses Menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat

2.1.1 Teori data mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data serupa pengetahuan yang selama ini tidak diketahui

secara manual dari suatu kumpulan data. Defenisi lain data miningadalah sebagai Proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan.istilah data mining kadang disebut juga knowledge discovery. Istilah data miningdan Knowledge discovery in Database (KDD) sering kali di gunakan secara bergantian untuk menjelaskan proses penggalihan informasi tersembunyi dalam suatu basias data yang bessar (Maukar et al., 2022).

Data transformasi bertujuan untuk merubah suatu bentuk data mentah yang telah lengkap dalam sesuatu database kedalam suatu bentuk data yangh singakat dan mudah untuk di pahami. Dari teknik ini nantinya akan dihasilkan suatu knowledge yang dapat di inginkan dalam pengambilan keputusan. Dan juga akan didapat hasil pengolahan data yang tepat dan cepat tercepat dapat dicapai. Sedangkan data mining adalah proses pencarian knowledge/ pattern dengan menggunakan salah satu teknik Artificial intelligent yang ada,contohnya seperti Rough Set(Rizky Amalia 2020)

Data miningadalah isu yang sangat panas sekarang ini, karena semakin banyakinformasi yang disimpan secara digital,kemampuan untuk mengumpulkan data jauh melebihi kemampuan seseorang untuk menanalisis itu kedepan data dan penemuan pengetahuan adalah bagian yang sangat penting bisnis saat ini.

Tujuan data ini adalah untuk Menemukan model deskripif dan prediktif dan pola dari data (Lidiya simanjuntak 2024)

Pola yang ditemukan harus penuh arti dan pola tersebut memberikan

keuntungan. Karakteristik data miningsebagai berikut :

Data mining berhubungan dengan penemuan sesuatu yang tertentu yang tidak diketahui sebelumnya.

Data mining biasa menngunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

Assoction rule mining adalah teknik mining untuk menemukan aturan assosiatif antara suatu kombinasi item. Contoh dari aturan assosiatif dari analisa pembelian di suatu pasar swalayan adalah bisa diketahui beberapa besar kemungkinan seseorang pelanggan pembeli roti bersamaan dengan susu.

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk mendapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.

Decision tree adalah satu metode classification yang paling populer karena mudah untuk diinterprestasi oleh manusia. Setiap percabangan menyatakan kondisi yang dipenuhi tiap ujuang pohon menyatakan kelas data.

Clustring, clustring melakukan pengelompokan data tanpa berdasarkan kelas datatersebut. Bahkan clustring dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu clustring sering sering digolongkan sebagai metode usupervised.

Neural network, jaringan syaraf buatan yang terlatih dapat dianggap sebagai pakar dalam kategori informasi yang akan dianalisis. Pakar ini

Dapat digunakan untuk memproyeksi situasi baru dari ketertarikan informasi. (jatisi 2023)

Berdasarkan Pembahasan terhadap penelitian ini maka dapat membuat suatu hasil landaskan teori yang dipelajari dalam mengatasi pembaca

Untuk mengatasi permasalahan diatas, tugas akhir ini akan membuat sistem cerdas dengan mengimlementasikan algoritma roughset. Roughset merupakan perluasan dari teori set untuk studi sistem cerdas yang ditandai dengan informasi eksak, pasti, atau samar-samar. Pendekatan roughset menjadi pendekatan yang penting dalam artificial intelligent (AI) dan ilmu kognitif , terutama pada area machine learing, akuisis pengetahuan, analisis keputusan, Pencarian pengetahuan dari database, sistem pakar, penalaran indikutif, dan pengenalan pola.Roghset telah banyak diterapkan dalam banyak permasalahan nyata pada kodektoran, dan lain-lain. Keuntungan utama menggunakan roughset adalah bahwa roughset tidak membutuhkan data awal atau data tambahan seperti probabilitas, pada teori fuzzy set. Dalam roughset, Kumpulan objek disebut sebagai informasi system (IS). Dari IS tersebut objek-objek dikelarifikasi kedalam area-area tertentu yang disebut dengan lower approximaction, boundray rogion, dan outside region. Dari pengelompokan area tersebut, dapat dilakukan perhitungan dependensi antar atribut, reduksi atribut, rule generation sehinggga dapat diperoleh rule dari data set yang di gunakan(Wanto et al., 2020).

2.1.2 Tahapan Proses Data Mining

Data miningadalah proses yang menggunakan teknikstatistik matematika, kecerdasan buatan, dan machinelearning untuk mengekstraksi dan mengidentifikasiinformasi yang bermanfaat dan pengetahuan yang terkaitdari

berbagai data base besar danData mining merupakan bidang yang berkembang pesat seiring dangan perkembangan teknologi informasiyang melibatkan pemakaian database berskala besarmapun kecil. Informasi yang tersimpan dalam databasemenjadi tidak berguna seiring berjalannya waktu. (kartika sari 2023).

Data Mining merupakan salah satu langkah dari serangkaian proses nerative KDD.DataMiningadalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data Miningadalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting(juisi 2020).

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, data miningdapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional databaseyang besar. Data mininguntuk mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk:

Prediksi trenddan sifat-sifat bisnis, dimana data miningmengotomatisasi proses pencarian informasi pemprediksi di dalam basis data yang besar.

2. Penemuan pola-pola yang tidak diketahui sebelumnya, dimana data mining"menyapu" basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan. (revandi guanawan 2022).

penerapan data miningdapat membantu menggali informasi yang tersimpan dalam databasedengan memanfaatkan Algoritma Rough Set, metode Rough Setmerupakan salah satu metode alat matematika untuk menangani ketidak jelasan dan ketidakpastian yang diperkenalkan untuk memproses ketidak pastian dan informasi yang tidak tepat dengan metode tersebut dapat menghasilkan informasi baru berupa pola aturan (rule) yang dapat digunakan dalam acuan penyeleksian pemohon bantuan, sehingga sangat membantu bagi tim seleksi dalam mengambil keputusan yang tepat sasaran (Wanto et al., 2020).

Data miningmerupakan bidang yang berkembang pesat seiring dengan perkembangan teknologi informasi yang melibatkan pemakaian databaseberskala besar maupun kecil.Informasi yang tersimpan dalam databasemenjadi tidak berguna seiring berjalannya waktu. Data miningdapat meningkatkan nilai tambah suatu database. Kita dapat menggali informasi yang tersimpan dalam databaseyang terakumulasi dalam jangka waktu lama untuk mendapatkan informasi tambahan Banyak algoritma mengimplementasikan data mining Salah satu algoritma yang cukup sederhana dan cukup mudah untuk diimplementasikan adalah algoritma Rough Set. Rough Setmerepresentasikan set data sebagai sebuah tabel, di mana baris dalam tabel merepresentasikan objek dan kolom-kolom merepresentasikan atribut dari objek-objek tersebut(Rohani 2023).

2.2 Representasi Identifikasi Metode Rough Set

Rough Set menawarkan dua bentuk representasi data yaitu Information System(IS) dan Decision System (DS). Definisi Decision System yaitu sebuah

pasangan Information System, di mana "U" adalah anggota bilangan "n" dan merupakan sekumpulan example dan attribute kondisi secara berurutan Rough Set Theorypertama kali diperkenalkanoleh Pawlak di awal 1980-an. Menurut PolkowskiRough Set Theorydapat digunakan untukpemilihan fitur, ekstrasi fitur, reduksi data,generasi aturan keputusan, ekstrasi pola, dan lainlain. Teori ini telah datang sebagai alternatifmetode yang banyak digunakan pada machinelearning dan analisis data statistika. Data yangdipresentasikan dalam kerangka Rough Setberbentuk tabel sistem informasi.Sistem informasiadalah suatu tabel di mana setiap baris data tabelyang mewakili suatu objek dan setiap kolommerupakan suatu atribut yang dapat diukur darisetiap objek.Rough Set Theoryberasal dari teorihimpunan biasa sehingga asumsi teknik penelitianklasik kuantitatif tidak berlaku(risky amalia 2020.)

Teori rough setadalah sebuah alat matematika untuk menangani ketidak jelasan dan ketidak pastian yang diperkenalkan untuk memproses ketidak pastian dan informasi yang tidak tepat.Rough Settelah banyak diterapkan dalam banyak permasalahan nyata pada kedokteran, farmakologi, teknik, perbankan, keuangan, analisis pasar, pengelolaan lingkungan dan lain-lain. Tahapan di dalam penggunaan algoritma Rough Setini sebagai beriku(H. Maulidiya and A. Jananto,2020):

1. Data Selection (Pemilihan data yang akan digunakan).
2. Pembentukan Decision System yang berisikan atribut kondisi dan atribut keputusan.

3. Pembentukan Equivalence Class, yaitu dengan menghilangkan data yang berulang.

4. Pembentukan Discernibility Matrix Modulo D, yaitu matriks yang berisikan perbandingan antar data yang berbeda atribut kondisi dan atribut keputusan.

5. Menghasilkan reductdengan menggunakan aljabar boolean.

6. Menghasilkan rule (pengetahuan).

Salah satu teknik yang digunakan untuk sistem pendukung keputusan ini adalah dengan menggunakan teknik Artificial Intelligent (Al) Rough Set yang mana teknik ini adalah merupakan teknik yang efisien untuk Knowledge Discovery in Database (KDD) proses dan data mining. Dengan teknik Artificial Intelligent (AI) Rough Set ini, nantinya akan didapat suatu hasil knowledge/pattern yang dapat digunakan dalam mengambil suatu keputusan, yaitu dengan melakukan tahapan-tahapan dalam KDD yang terdiri dari data selection, data cleaning, data transformation, data mining, dan evaluation.

Data selection bertujuan untuk menyeleksi data-data yang akan diproses, yang mana data-data yang akan diambil untuk diproses tidak keseluruhan dari data-data yang ada di dalam database. Data cleaning berguna untuk menangani sejumlah data yang besar dalam suatu database. Di mana dengan teknik ini data-data dalam suatu database yang bersifat uncertainty (tidak pasti) seperti incomplete (tidak lengkap) dan inconsistent (tidak konsisten) dapat dijadikan complete (lengkap) dan consistent (konsisten). Sehingga proses ini dapat memberikan suatu keuntungan yang besar untuk menghasilkan suatu data yang konsisten dan data yang lengkap pada suatu database. Data transformasian

bertujuan untuk merubah suatu bentuk data mentah yang telah lengkap dalam suatu database kedalam suatu bentuk data yang singkat dan mudah untuk dipahami. Dari teknik ini nantinya akan dihasilkan suatu knowledge yang dapat di inginkan dalam pengambilan keputusan. Dan juga akan didapat hasil pengolahan data yang mempunyai tingkat keakuratan yang tinggi, sehingga keputusan yang tepat dan cepat tersebut dapat dicapai. Sedangkan data mining adalah proses pencarian knowledge/pattern dengan menggunakan salah satu teknik Artificial Intelligent yang ada, contohnya seperti Rough Set (A. Damuri,).

Rough set adalah sebuah teknik matematik yang dikembangkan oleh Pawlack pada tahun 1980. Rough Set salah satu teknik data mining yang digunakan untuk menangani masalah Uncertainty, Imprecision dan Vagueness dalam aplikasi Artificial Intelligence (AI).Rough set merupakan teknik yang efisien untuk Knowledge Discovery in Database (KDD) dalam tahapan proses dan Data Mining. (Hafisa,2022).

Dalam Rough set, sebuah set data dipersentasikan sebagai sebuah tabel, dimana baris dalam tabel mempersentasikan objec dan kolom mempersentasikan dari atribut keobjek tersebut sehingga untuk penyelesaian dengan menggunakan metode rough set dilakukan dengan beberapa langkah-langkah dalam proses data mining yaitu (Harli Trimulya Suand,2021).

2.2.1 Decision System

Dimana "U" adalah set terhingga yang tidak kosong dari objek yang disebut dengan universe dan "A"set terhingga tidak kosong dari atribut di mana

untuk tiapSet disebutvalue set dari "A"

Tabel 2.1 information system

Objek	Tipe	Lokasi	Fasilitas (Lantai , Kamar, Dapur, WC)	Kelebihan Tanah	Harga Standar sesuai tipe (Diluar Kelebihan) Tanah
Wianda	36	Tengah	Kurang	Ada	Ya
Sinta	36	Tengah	Cukup	Ada	Ya
Aldo khoir	36	Hook	Baik	Ada	Ya
Eka tanjung	36	Hook	Baik	Ada	Ya
Fauzan	36	Tengah	Baik	Ada	Ya

Tabel 1 memperlihatkan sebuah informasi system yang sederhana. Dalam information System, tiap-tiap baris merepresentasi objek sedangkan column merepresentasikan attribute.

Dalam Penggumpulan information system, terdaftar autcome dari Klarifikasi yang telah diketahui yang di buat dengan atribut keputusan, information system tersebut disebut dengan dicision sytem decision sytem dapat digambarkan sebagai :

Dimana : IS=(U,{A,C}

U ={x1,x2,...,xm} yangmerupakan sekumpulan example. A = a{a,a2,...,an} yang merupakan sekumpulan attribut kondisi secara berurutan atau atribut

C = Decision attribut (keputusan).

Decision systems (DS) yang sederhana diperahatikan pada tabel 2

Tabel 2.2 decsilon stystem (DS)

Objec	Tipe	Lokasi	Fasilitas (Lantai , Kamar, Dapur, WC)	Kelebihan Tanah	Harga Standar sesuai tipe (Diluar Kelebihan) Tanah	idaman	
EC1	36	Tengah	Kurang	Ada	Ya	Tidak	
EC 2	36	Tengah	Cukup	Ada	Ya	Tidak	
EC 3	45	Hook	Baik	Ada	Ya	Ya	
EC4	45	Hook	Baik	Ada	Ya	Ya	
EC5	90	Tengah	Baik	Ada	Ya	Ya	

Dalam tabel 2, n-1 attributeTipe, lokasi Fasilitas (lantai Kamar , dapur , dan WC), Kelebihan Tanah, Harga standar Sesuai Tipe (Diluar kelbihan Tanah) adalah attribute dengan Decision sytem dapat diGambarkan sebagai

Di mana: U = {x1,x2,...xm} yang merupakan sekumpulan attribute kondisi secara berurutan atau atribute , Seperti tipe , lokasi, fsilitas (lantai , Kamar , dapur ,dan WC),

Kelebihan Tanah, harga standar sesuai Tipe (diluar kelebihan tanah) dan idaman.

C = Decisiopn attributes (Keputusan). Decision system (DS) yang sederhana diperhatikan pada tabel 2 kondisi sedangkan Idaman adalah decsision

attribute(riski amalia,2020).

2.2.2 Eguivalence Class

Mengelompokkan Objek-Objek yang sama untuk attribute E(U, A). Diberikan Decision systems pada tabel 2, kita dapat memperoleh equvalence class (EC1-EC6 sepertri pada tabel 3

Tabel 2.3 Proses Equivalence class

2.2.3 Discernibility Matrix

Defenisi discerniblity Matrix: diberikan sebuah IS A=(U,A) and B dan A, discernibility matrix dari A adalah MB, di mana tiap-tiap entry Mb(I,J) terdiri dari sekumpulan attribute yang berbeda antara objek Xi dan Xj. Tabel 3 memperlihatkan discernibility tabel 4 Tabel acuan Discernibility Matrix atauDiscernibillity matrix Modul D

Untuk mendapatkan nilai Discernibility Matrix-nya yaitu dengan

Objec	EC2	EC3	EC4	EC5	EC6
EC1	X	C	BC	AC	AC
EC 2	C	X	BC	AC	A
EC 3	BC	BC	X	AB	ABC
EC4	ABC	ABC	A	AB	BC
EC5	AC	AC	AB	X	AC
EC6	AC	A	ABC	AC	X

mengklarifikasikan artibutyang berbeda antara objek ke – i dan objek ke-j (yangdilihat hanya atribut kondisi saja) . Berdasarkan data di atas maka berikut iniadalah discernibility matrixnya:

Tabel 2.4 hasil Discernibility Matrix

Objec	EC2	EC3	EC4	EC5	EC6
EC1	X	C	BC	AC	AC
EC 2	C	X	BC	AC	A
EC 3	BC	BC	X	AB	ABC
EC4	ABC	ABC	A	AB	BC
EC5	AC	AC	AB	X	AC
EC6	AC	A	ABC	AC	X

2.2.4 Discernibility Matrix Modulo D

Selian itu juga dapat menggunakan Discernibility Matrix Modulo D. Discrnibility Matrix Modulo Dini merupakan sekumpulan atribut yang berbeda antara objek ke-i dan objek ke-jbeserta dengan atribut hasilnya seperti terlihat pada tabel di di bawah ini. Adapun penulis menggunakan Discrnibility Matrix Sebagai acuan untuk melakukan proses reduction.

2.2.5 Reduction

Adapun penulis menggunakan Discernibility Matrixsebagai acuan untuk melakukan proses Reduction. Untuk data yang jumlah variabel yang sangat besar sangat tidak mungkin mencari seluruh kombinasi variabel yang ada, karena jumlah indiscernibilty yang dicari - 2 (2 n-1-1). Oleh karena itu dibuat satu tekhnik pencarian kombinasi atribut yang dikenal dengan QuickReduct, yaitu dengan cara:

1. Nilai indiscernibiltyyang pertama dicari adalah indiscernibiltyyang kombinasi atribut yang terkecil yaitu 1.

2. Kemudian lakukan proses pencarian dependency attributes. Jika nilai dependency attributesyang didapat 1, maka indiscernibiltyuntuk himpunan minimal variabel adalah variabel tersebut.

3. Jika pada proses pencarian kombinasi atribut tidak ditemukan dependency attributes 1, maka lakukan pencarian kombinasi yang lebih besar, di mana kombinasi variableyang dicari adalah kombinasi dari variabel di tahap sebelumnya yang nilai dependency attributespaling besar.

Lakukan proses (3), sampai didapat nilai dependency attributes 1.

Berdasarkan hasil yang diperoleh dari proses Discernibility Matrixberikut ini adalah proses Reduction-nya (riski amalia,2020).

Tabel 2.5 Proses Reduction

Objec	CNF OF Booelean Function	Prime implicant	Reduct
EC1	C^(Bvc(AvBVc)^(AvC)^(AvC)	C(AvC)	{c}
EC2	C^(Bvc(AvBVc)^(AvC)^A	A^c	{A,C}
EC3	(AvC^(BvC)^A^(AvBvC)	A^(BvC)	{A,B} {A,C}
EC4	(AvBvC^AvBvC)6A^(AvB^(BvB)	A^(BvC)	{A,B} {A,C}
EC5	(AvC^(AvC^(AvB)^(AvB^(AvC)	Av BC	{A,B} {A,C}

EC6	(AvC)^A^(AvBvC^(BvC)^(AvC)	A	{A}

2.2.6 Generation Rules

Setelah didapatkan hasil dari Reduction, maka langkah terakhir untuk menentukan General Rulesnya. Adapun General Rulesdari hasil Reductionyang dideskripsikan pada penyeleksian tipe rumah idaman sesuai selera konsumen.

2.1. Model Klasifikasi

Model klasifikasi merupakan salah satu bentuk penerapan data mining yang digunakan untuk mempelajari pola dan hubungan dari berbagai atribut atau variabel guna memprediksi nilai atau kelas dari suatu data [38]. Model ini bekerja dengan memanfaatkan data pelatihan (training data) yang terdiri dari berbagai contoh atau sampel dengan atribut dan label yang sudah diketahui, sehingga dapat digunakan untuk "belajar" pola yang muncul dari data tersebut. Setelah pola dapat diidentifikasi, model klasifikasi dapat digunakan untuk memprediksi nilai atau kelas dari data yang belum pernah terlihat sebelumnya, berdasarkan pola yang telah dipelajari dari data pelatihan [39]. Berbagai algoritma klasifikasi yang digunakan dalam praktik data mining termasuk *Naïve Bayes*, Decision Tree, Support Vector Machine, K-Nearest Neighbors, dan lain sebagainya, masing-masing dengan kelebihan dan karakteristik tertentu. Pemilihan algoritma klasifikasi yang tepat dapat berdampak signifikan terhadap tingkat akurasi dan relevansi model dengan kebutuhan analisis, khususnya dalam konteks pemodelan pola transaksi dan

perilaku konsumen [40].

Penelitian ini memanfaatkan model klasifikasi, khususnya algoritma *Naïve Bayes*, untuk menganalisis pola transaksi di Oreen Studio berdasarkan nilai diskon dan bentuk promosi yang digunakan sebagai atribut data [41]. Model ini digunakan untuk mempelajari pola dari data pemesanan historis yang terdiri dari berbagai contoh nilai diskon, bentuk promosi, jumlah pemesanan, dan atribut lain yang terkait, guna memprediksi pola atau pola hubungan yang dapat digunakan sebagai dasar pengambilan keputusan bisnis. Dengan penerapan model klasifikasi tersebut, Oreen Studio dapat mengidentifikasi pola pemesanan yang paling signifikan terkait dengan nilai diskon dan bentuk promosi, serta mendapatkan gambaran yang lebih jelas mengenai atribut yang memengaruhi peningkatan jumlah pemesanan. Hasil dari analisis ini memungkinkan Oreen Studio untuk merumuskan strategi promosi dan diskon yang lebih tepat guna, berdasarkan pola dan informasi yang dapat diandalkan dari data yang ada [42].

2.2. Alat Bantu Program Aplikasi Orange

Orange merupakan salah satu perangkat lunak open source untuk analisis data dan pembelajaran mesin (*Machine Learning*) yang dikembangkan guna mempermudah proses pengolahan, visualisasi, dan pemodelan data bagi berbagai kalangan, mulai dari akademisi, peneliti, hingga pelaku bisnis [43]. Aplikasi ini menawarkan antarmuka grafis yang intuitif dan dapat digunakan oleh pengguna dengan berbagai tingkat pengalaman, dari pemula hingga ahli, tanpa memerlukan penulisan kode yang kompleks. Berbagai metode analisis data dapat dijalankan melalui Orange, termasuk klasifikasi, clustering, regresi, dan analisis asosiasi, yang

dapat dikombinasikan dengan berbagai fitur visualisasi seperti scatter plot, bar chart, heat map, dan lain sebagainya. Salah satu keunggulan dari Orange adalah ketersediaan berbagai widget yang dapat digunakan untuk membaca data dari berbagai format, memproses data, membuat model, mengevaluasi kinerja model, hingga membuat laporan dari hasil analisis.

Gambar 2. 4. Aplikasi Orange

Orange digunakan sebagai alat bantu untuk memodelkan pola hubungan dari atribut diskon dan promosi yang diterapkan oleh Oreen Studio terhadap pola pemesanan yang terjadi [44]. Dengan memanfaatkan Orange, data transaksi dapat diolah dan dianalisis guna mempelajari pola hubungan antaratribut yang terdiri dari nilai diskon, jenis promosi, jumlah pemesanan, dan atribut lain yang relevan. Melalui berbagai metode klasifikasi yang tersedia, termasuk algoritma *Naïve Bayes*, Orange memungkinkan identifikasi pola dengan tingkat akurasi yang dapat diukur, sehingga pola atau aturan yang paling signifikan dapat digunakan sebagai bahan pertimbangan bagi Oreen Studio dalam membuat strategi promosi yang lebih efektif. Visualisasi dan evaluasi yang dihasilkan dari Orange juga dapat digunakan sebagai bahan pelaporan dan pertimbangan bagi pihak manajemen guna mengoptimalkan bentuk promosi yang digunakan, dengan tujuan akhir untuk

meningkatkan jumlah pemesanan dan daya saing usaha.

2.3. Kelebihan dan Kekurangan Metode Naive Bayes

Metode *Naïve Bayes* merupakan salah satu algoritma klasifikasi yang sangat populer dalam data mining dan *Machine Learning*, khususnya untuk pengklasifikasian pola dengan jumlah atribut yang relatif banyak. Salah satu kelebihan dari metode ini ialah kesederhanaannya dalam implementasi dan kebutuhan waktu pemrosesan yang sangat efisien, bahkan untuk jumlah data yang besar [45]. *Naïve Bayes* bekerja dengan asumsi bahwa setiap atribut dalam data independen satu sama lain, sehingga memungkinkan algoritma ini untuk membuat prediksi dengan cepat berdasarkan nilai probabilistik dari masing-masing atribut. Hal ini menjadikan metode *Naïve Bayes* sangat sesuai untuk diterapkan pada berbagai studi pola, mulai dari analisis pola transaksi, klasifikasi teks, hingga deteksi spam, karena dapat menangani data dengan struktur kompleks secara efisien [46]. Namun, asumsi kemandirian antaratribut ini juga dapat menjadi titik lemah dari *Naïve Bayes*, terutama bila atribut yang digunakan memiliki tingkat korelasi yang tinggi. Dalam kondisi tersebut, akurasi model dapat terpengaruh, mengingat asumsi independensi sulit terpenuhi sepenuhnya. Meskipun demikian, dengan pemilihan atribut yang tepat dan jumlah data pelatihan yang memadai, *Naïve Bayes* tetap dapat menjadi metode yang handal dan efektif untuk berbagai kebutuhan klasifikasi.

Naïve Bayes digunakan sebagai metode klasifikasi untuk menganalisis pola pemesanan berdasarkan nilai diskon dan jenis promosi yang digunakan oleh Oreen Studio [47]. Kecepatan dan kesederhanaan metode ini memungkinkan pengolahan

data transaksi dalam jumlah yang besar dengan waktu yang efisien, sehingga dapat digunakan untuk memprediksi pola pembelian dengan tingkat akurasi yang memadai. Kemampuan *Naïve Bayes* dalam menangani berbagai atribut terkait nilai diskon dan pola promosi sangat berguna bagi Oreen Studio, mengingat jumlah data pemesanan yang terus tumbuh dari waktu ke waktu dan terdiri dari berbagai variasi bentuk promosi yang digunakan. Walaupun asumsi independensi antaratribut belum sepenuhnya dapat dijamin, jumlah data transaksi yang memadai dan pemilihan atribut yang relevan dapat meminimalkan dampak dari asumsi tersebut, sehingga model yang dihasilkan tetap dapat digunakan sebagai acuan bagi pengambil keputusan untuk mengoptimalkan nilai dan bentuk promosi guna meningkatkan jumlah pemesanan secara signifikan.

2.4. Evaluasi Model Naïve Bayes

Evaluasi model *Naïve Bayes* merupakan tahap krusial dalam proses penerapan algoritma klasifikasi guna memastikan bahwa pola dan aturan yang dihasilkan dapat diandalkan untuk pengambilan keputusan. Model ini dievaluasi dengan menghitung berbagai metrik kinerja, seperti tingkat akurasi (*accuracy*), presisi (*precision*), sensitivitas atau recall, dan nilai *F1-Score*, yang masing-masing memberikan gambaran berbeda terkait kemampuan model dalam mengklasifikasi data dengan benar [48]. Akurasi mengukur persentase total prediksi yang benar dari semua data uji, sedangkan precision memberikan gambaran seberapa banyak dari sampel yang diprediksi sebagai positif memang benar-benar positif. Recall digunakan untuk menilai kemampuan model dalam mengidentifikasi seluruh sampel positif dari data yang tersedia, sedangkan F1-Score menggabungkan nilai

precision dan recall ke dalam ukuran tunggal untuk mengukur performa model secara menyeluruh. Evaluasi ini juga dapat dilengkapi dengan confusion matrix yang memberikan rincian jumlah prediksi yang masuk kategori *True Positive*, *True Negative*, *False Positive*, maupun *False Negative*. Dengan pengukuran dan interpretasi yang lengkap ini, model dapat dikaji lebih mendalam untuk memastikan kualitas dan relevansinya sebelum digunakan sebagai landasan pengambilan keputusan bisnis.

Evaluasi model *Naïve Bayes* digunakan untuk mengukur sejauh mana pola pemesanan di Oreen Studio dapat diklasifikasikan dengan tingkat akurasi yang tinggi berdasarkan atribut nilai diskon dan bentuk promosi. Model dikembangkan dari data historis pemesanan yang terdiri dari berbagai nilai dan pola promosi, kemudian divalidasi dengan metode pengujian yang terdiri dari data pelatihan dan data pengujian guna memperoleh nilai-nilai akurasi, *precision, recall, dan F1-Score* yang relevan. Hasil dari evaluasi ini memberi gambaran mengenai tingkat keandalan model dalam memprediksi pola pemesanan di masa mendatang, sehingga dapat digunakan untuk merumuskan strategi promosi yang lebih terarah dan efisien bagi Oreen Studio [49]. Dengan nilai evaluasi yang memadai, model *Naïve Bayes* dapat dijadikan sebagai alat bantu bagi pihak manajemen untuk mengoptimalkan nilai dan bentuk promosi guna meningkatkan jumlah pemesanan dan daya saing usaha di tengah perkembangan bisnis jasa fotografi dan videografi yang semakin kompetitif.

2.5. Kelebihan Penelitian

Kelebihan dari penelitian ini terletak pada pendekatan yang digunakan dalam menganalisis pola pemesanan dengan memanfaatkan metode klasifikasi data mining, khususnya algoritma *Naïve Bayes* [50]. Berbeda dari berbagai metode analisis deskriptif biasa, penerapan *Naïve Bayes* memungkinkan peneliti untuk tidak hanya memotret pola data transaksi yang sudah terjadi, tetapi juga memprediksi pola pemesanan yang belum terlihat dari data historis. Keunggulan metode ini ialah kemampuannya untuk menangani berbagai atribut atau variabel yang digunakan, termasuk nilai diskon, jenis promosi, jumlah transaksi, maupun pola waktu pemesanan. Dengan kemampuan tersebut, analisis dapat dilakukan dengan tingkat akurasi yang tinggi, bahkan ketika jumlah data yang digunakan relatif terbatas. Hal ini memberikan nilai lebih bagi pelaku bisnis jasa, khususnya Oreen Studio, untuk dapat memahami pola pemesanan pelanggan berdasarkan nilai diskon dan promosi yang digunakan, sehingga dapat dijadikan landasan pengambilan keputusan yang lebih matang dan strategis.

Penelitian ini memberikan nilai lebih bagi Oreen Studio karena dapat menjembatani kebutuhan bisnis dengan penerapan teknologi data mining yang relevan dan efisien. Dengan memanfaatkan metode *Naïve Bayes*, Oreen Studio dapat mengidentifikasi pola dari berbagai atribut transaksi guna mengetahui bentuk diskon dan promosi yang paling berdampak signifikan terhadap peningkatan jumlah pemesanan. Hasil dari penelitian ini memungkinkan Oreen Studio untuk merumuskan strategi promosi yang lebih terarah dan efisien, mengurangi risiko pengeluaran biaya promosi yang tidak memberikan efek positif, serta

memaksimalkan nilai dari setiap bentuk promosi yang dijalankan.

2.6. Penelitian Terdahulu

Pada tabel dibawah ini merupakan penelitian terdahulu tentang metode Naive Bayes yang digunakan pada penelitian ini. Untuk tabel nya yaitu sdebagai berikut.

Referensi Penelitian	1
Judul	Penerapan Data Mining Dalam Proses Hukum Pidana Bagi Pelaku Kekerasan Pada Wanita Menggunakan Algoritma C4.5
Nama Penulis	Lidya Priscila Simanjuntak
Tahun	2024
Hasil	Permasalahan selama ini yang terjadi pada POLDASU dalam menangani proses hukum pidana kekerasan pada wanita ditangani oleh Kanit PPA 1(Perlindungan Perempuan dan Anak) dengan cara mengumpulkan identitas pelaku kekerasan, identitas korban, dan identitas saksi, barang bukti

	Pohon keputusan (decision t ree) merupakan salah satu teknik terkenal dalam data mining dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasu
Referensi Penelitian	2
Judul	Clustering Tingkat Kekerasan pada Anak Dan Perempuan Berdasarkan Jenis Kekerasan Yang Dialami di Sumatera Utara 2023
Nama Penulis	Marnis Nasution
Tahun	2025
Hasil	roses pengumpulan data diambil diambil dari data jumlah korban kekerasan terhadap anak berdasarkan jenis kekerasan yang dialami dan juga data jumlah koban kekerasan terhadap Perempuan berdasakan jenis kekerasan yang dialami. Dimana data diperoleh dari dinas perlindungan Perempuan dan anak. Jenis kekerasan yang dialami oleh anak dan perempuan adalah fisik, psikis, seksual, ekploitasi, tindak pidana perdagangan orang, penelantaran dan lainnya. Data berupa jumlah korban ditiap jenis kekerasan ditiap-tiap kabupaten.