

**ANALISIS KRITERIA KAMAR KOS LAYAK HUNI
MENGUNAKAN ALGORITMA C4.5
DI KOTA BATAM**

SKRIPSI



**Oleh
Ratih Yuninda Hutagalung
181510059**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
2023**

**ANALISIS KRITERIA KAMAR KOS LAYAK HUNI
MENGUNAKAN ALGORITMA C4.5
DI KOTA BATAM**

SKRIPSI
Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana



Oleh
Ratih Yuninda Hutagalung
1815100159

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
2023

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini saya:

Nama : Ratih Yuninda Hutagalung
NPM : 181510059
Fakultas : Teknik dan Komputer
Program Studi : Sistem Informasi

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

Analisis Kriteria Kamar Kos Layak Huni Menggunakan Algoritma C4.5 Di Kota Batam

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 30 Januari 2023



Ratih Yuninda Hutagalung
181510059

**ANALISIS KRITERIA KAMAR KOS LAYAK HUNI
MENGUNAKAN ALGORITMA C4.5 DI KOTA
BATAM**


SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar Sarjana**

**Oleh
Ratih Yuninda Hutagalung
181510059**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera dibawah ini**

Batam, 30 Januari 2023


Rika Harman, S.Kom., M.SI.
Pembimbing

ABSTRAK

Tempat tinggal yang layak huni merupakan kebutuhan dasar manusia. Rasa nyaman serta aman menjadi patokan untuk hunian yang layak. Kamar kos/ *dormitory* yang dekat dengan tempat bekerja menjadi pilihan para karyawan yang belum memiliki rumah maupun kendaraan, selain menghemat waktu juga dapat mengurangi biaya kebutuhan. Dikatakan layak apabila kamar kontrakan memenuhi standar kelayakan menurut para penghuni, mulai dari saluran air, MCK, ventilasi udara, ukuran kamar dan sesuai dengan harga yang diberikan. Untuk mengetahui apakah suatu kamar kos/ *dormitory* layak maka penulis menemukan solusi menggunakan algoritma C4.5. Algoritma ini dipilih karena merupakan salah satu metode *decision tree/* pohon keputusan yang mudah dimanfaatkan untuk prediksi suatu kasus. Proses perhitungan menggunakan *software* WEKA. Peneliti menyebarkan hak angket pada penghuni kos sebanyak 50 orang setelahnya data yang diterima akan diolah melalui proses penghitungan secara manual dan dimasukkan kedalam tabulasi dan sesuai variabelnya kemudian dites kembali dengan aplikasi WEKA untuk menyesuaikan kebenaran data. Peneliti menggunakan 2 variabel keputusan yaitu Layak Huni dan Tidak Layak Huni

Kata Kunci: ***Data Mining, Algoritma C4.5, Kamar Kos, Node, WEKA.***

ABSTRACT

Comfort residence is a basic human need. A sense of comfort and security is a benchmark for decent housing. Boarding rooms/ dormitories that are close to the place of work are the options of employees who do not have a house or vehicle, besides saving time they can also reduce the cost of necessities. It is said to be appropriate if the rented room meets the eligibility standards according to the occupants, starting from the drains, toilets, air ventilation, room size and according to the price given. To find out whether a boarding room/dormitory is feasible, the authors find a solution using the C4.5 algorithm. This algorithm was chosen because it is a decision tree method that is easy to use for predicting a case. This study also uses the WEKA application version 3.9. The researchers distributed questionnaires to 50 boarding house occupants after which the data received would be processed through a manual counting process and entered into tabulations and according to the variables then tested again with the WEKA application to adjust the correctness of the data. Writer used 2 decision variables, namely Decent and Undecent.

Keywords: *Data Mining, Algorithm C4.5, Rent House, nodes, WEKA*

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi Universitas Putera Batam.

Masih banyak terdapat kekurangan dari skripsi ini suatu kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari skripsi ini terwujud dari beberapa dukungan, bimbingan, nasehat dan bantuan dari beberapa pihak. Maka, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. selaku Rektor Universitas Putera Batam.
2. Bapak Muhammat Rasid Ridho, S.Kom., M.SI. selaku Ketua Program Studi Sistem Informasi Universitas Putera Batam.
3. Bapak Rika Harman, S.Kom., M.SI. selaku dosen pembimbing skripsi pada program studi sistem informasi yang telah memberi support kepada penulis untuk menyelesaikan skripsi ini.
4. Dosen dan Staff Universitas Putera Batam.
5. Kedua orangtua yang selalu mendoakan, memberi dukungan dan semangat hingga penulis bisa menyelesaikan penelitian ini tepat waktu.
6. Teman-teman seperjuangan yang bersedia membagikan ilmunya.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 30 Januari 2023

Ratih Yuninda Hutagalung

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iii
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	1
1.1. Latar belakang	1
1.2. Identifikasi masalah.....	3
1.3. Batasan Masalah.....	3
1.4. Rumusan masalah.....	4
1.5. Tujuan penelitian.....	5
1.6. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	8
2.1 Teori	8
2.1.1 Teori Umum	8
2.1.1.1 <i>Data Mining</i>	8
2.1.1.2 Algoritma C4.5.....	12
2.1.1.3 Pohon Keputusan (<i>decision tree</i>)	14
2.1.2 Teori Khusus	15
2.1.2.1 Rumah Layak Huni	15
2.1.2.2 Indekos	16
2.1.2.3 Kelayakan.....	16
2.1.2.4 Software Weka	17

2.2	Kerangka Pemikiran.....	18
2.3	Hipotesis Penelitian.....	19
2.4	Penelitian Terdahulu.....	19
DAFTAR ISI		
BAB III METODE PENELITIAN		23
3.1	Desain Penelitian.....	23
3.2	Objek Penelitian	25
3.2.1	Lokasi Penelitian	25
3.2.2	Jadwal Penelitian.....	25
3.3	Populasi dan Sampel	25
3.3.1	Populasi	25
3.3.2	Sampel.....	26
3.4	Variabel Penelitian	26
3.4.1	Teknik Pengumpulan Data	26
3.4.2	Model penelitian.....	27
BAB IV HASIL DAN PEMBAHASAN.....		28
4.1	Hasil Penelitian	28
4.2	Pembahasan.....	29
4.3	Analisis Data	33
4.3.1	Pembersihan Data.....	33
4.3.2	Seleksi Data.....	34
4.3.3	Transformasi Data	34
4.3.4	Pohon Keputusan.....	38
4.3.4.1	Node 1	39
4.3.4.2	Node 2	45
4.3.4.3	Node 3	50
4.3.5	Hasil Perhitungan Entropy dan Gain.....	54
4.3.7	Pengujian menggunakan <i>Software WEKA</i>	57
BAB V SIMPULAN DAN SARAN.....		65
5.1	Kesimpulan.....	65
5.2	Saran.....	66
DAFTAR PUSTAKA		67
LAMPIRAN 1 DATA KUISIONER.....		70
LAMPIRAN 2 DAFTAR RIWAYAT HIDUP		74
LAMPIRAN 3 SURAT IJIN PENELITIAN		76

LAMPIRAN 4 HASIL TURNITIN 77

DAFTAR ISI

DAFTAR GAMBAR

	Halaman
Gambar 1.1 Jumlah Penduduk Kota Batam Pada Tahun 1980-2020	1
Gambar 2.1 Tahapan Proses KDD	8
Gambar 2.2 Representasi Proses Data Mining	11
Gambar 2.3 Kerangka Pemikiran	17
Gambar 3.1 Desain Penelitian	23
Gambar 4.1 Pohon Keputusan	28
Gambar 4.2 Pohon Keputusan <i>Node 1</i>	43
Gambar 4.3 Pohon Keputusan <i>Node 2</i>	48
Gambar 4.4 Pohon Keputusan <i>Node 3</i>	52
Gambar 4.5 <i>Node 1</i> dengan <i>Microsoft Excel</i>	53
Gambar 4.6 <i>Node 2</i> dengan <i>Microsoft Excel</i>	54
Gambar 4.7 <i>Node 3</i> dengan <i>Microsoft Excel</i>	54
Gambar 4.8 Format Data <i>Microsoft Excel</i>	57
Gambar 4.9 Format Data Notepad.....	56
Gambar 4.10 <i>Software WEKA</i>	57
Gambar 4.11 Data Pada <i>Menu Explorer</i>	57
Gambar 4.12 Memilih Tabel Yang Akan di Input.....	58
Gambar 4.13 Data Pada <i>Menu Classify</i>	59
Gambar 4.14 <i>Classify Output</i> pada <i>WEKA</i>	60
Gambar 4.15 <i>Visualize Tree</i> pada <i>WEKA</i>	61
Gambar 4.16 Hasil <i>Visualize Tre</i> pada <i>WEKA</i>	61

DAFTAR TABEL

	Halaman
Tabel 3.1 Jadwal Penelitian.....	25
Tabel 4.1 Data Kriteria Kamar Kos.....	29
Tabel 4.2 Klasifikasi Atribut Ketersediaan Air.....	33
Tabel 4.3 Klasifikasi Atribut Ventilasi Udara.....	33
Tabel 4.4 Klasifikasi Atribut Ukuran Kamar.....	34
Tabel 4.5 Klasifikasi Atribut Sarana MCK.....	34
Tabel 4.6 Klasifikasi Atribut Harga.....	34
Tabel 4.7 Pra Proses Data.....	35
Tabel 4.8 Hasil Perhitungan <i>Node 1</i>	42
Tabel 4.9 Perhitungan Hasil <i>Node 2</i>	47
Tabel 4.10 Perhitungan Hasil <i>Node 3</i>	51

DAFTAR RUMUS

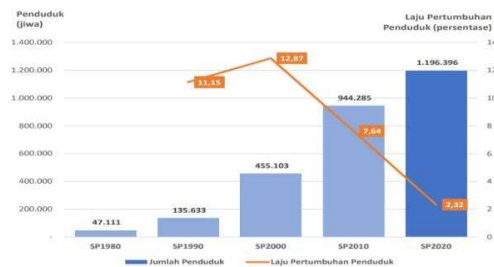
	Halaman
Rumus 2.1 Rumus <i>Gain</i>	12
Rumus 2.2 Rumus Nilai <i>Entropy</i>	12

BAB I

PENDAHULUAN

1.1. Latar belakang

Batam merupakan salah satu kota industri di Indonesia. Pada tahun 1970-an, Batam ditetapkan dengan Keputusan Presiden no. 41 Tahun 1973 ditetapkan sebagai syarat untuk aksi industrial. Kota Batam sangat diuntungkan dengan lokasinya yang strategis sebagai salah satu kawasan industri. Artinya, Batam secara geografis berbatasan dengan Singapura dan Malaysia, serta terletak di Selat Malaka, alur pelayaran tersibuk di dunia. Kehadiran industrial di Kota Batam terletak di kawasan bisnis. Kawasan industri Batam terbagi menjadi dua bagian, kawasan industri ringan dan kawasan industri berat. Industri ringan meliputi industri pengolahan, industri elektronika, industri jahit, industri plastic dan lain-lain. Industri berat di Kota Batam terutama adalah pembuatan kapal, baja, dan logam.



Gambar 1.1 Jumlah penduduk Kota Batam pada tahun 1980-2020

Sumber: Berita Resmi Statistik No. 14/3/2171.Th.1,2 Maret 2021

Minat masyarakat untuk bekerja di Kota Batam, akibatnya pekerja membutuhkan akomodasi dan bantuan untuk kegiatan sehari-hari mereka. Kondisi tersebut mendorong warga kota Batam memanfaatkan peluang untuk memenuhi kebutuhan tenaga kerja tersebut, hal ini dibuktikan dengan adanya lokasi, duplikasi dan persewaan rumah bahkan usaha makanan. Kos merupakan salah satu hunian vertikal dengan tarif tertentu yang berguna sebagai tempat tinggal sementara, baik itu yang menempuh pendidikan di luar tempat tinggal atau bekerja pada lokasi yang tidak menentu (Heryanto, Hartono, and Ronyastra 2021). Namun terdapat beberapa masyarakat yang memaksakan mendirikan rumah kos dengan fasilitas seadanya. Sehingga terdapat rumah kos yang seadanya, seperti ukuran kamar yang sempit, tidak tersedianya lahan parkir, sirkulasi udara yang kurang pada kamar, dan lainnya. Namun bagi sebagian karyawan yang berpenghasilan kurang, banyak yang tidak mempermasalahkan hal tersebut karena selain harga yang murah dan lokasi rumah kos yang dekat dengan tempat kerja mereka.

Membuka usaha rumah kos membutuhkan perencanaan yang matang. Langkah pertama adalah melakukan analisis kelayakan bisnis. Lakukan analisis kelayakan bisnis untuk mengetahui apa yang layak diselidiki. Langkah pertama adalah penjualan dan pemasaran untuk menentukan apakah ada target pasar. Langkah kedua adalah menentukan apakah lokasi tersebut secara teknis cocok untuk bisnis rumah kos. Langkah ketiga adalah kelompok keuangan untuk menentukan apakah bisnis tersebut menguntungkan dan berkelanjutan. (Saksono 2018).

Setelah menjelaskan mengenai permasalahan yang terjadi pada rumah kos yang ada di Batam dan penulis beberapa kali menyebutkan jurnal penelitian, mengulang dan menegaskan bahwa penulis ingin melakukan kegiatan penelitian kajian efisiensi rumah kos di kota Batam, dan terakhir menyatakan laporan kerja penelitian penulis sebagai berikut. “Analisis Kriteria Rumah Kos Layak Huni Menggunakan Algoritma C4.5” di Kota Batam.

1.2 Identifikasi masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, penulis mengidentifikasi masalah – masalah yang ada dalam penelitian sebagai berikut, yaitu:

1. Adanya keluhan dari para pekerja di kota Batam tentang kesulitan untuk menentukan tempat tinggal atau kamar kos yang layak huni.
2. Tidak adanya acuan atau standar dalam menentukan tingkat kelayakan kamar kos yang ditempati oleh para pekerja di Batam.
3. Kesulitan para pemilik kamar kos untuk menentukan hunian yang paling banyak diminati para pekerja.

1.3 Batasan Masalah

Untuk memperjelas pembahasan dalam penelitian ini, maka penulis membatasi pertanyaan yang akan diteliti, dengan tetap memfokuskan penilitan, yaitu:

1. Data untuk penelitian ini diambil pada Kota Batam (dormitori) khususnya bagi para karyawan yang sedang menentukan kamar kos.

2. Penelitian ini menggunakan algoritma C4.5 dan *software* Weka versi 3.9.4 dalam pengujian untuk menentukan pohon keputusan yang nantinya dibutuhkan untuk pengambilan keputusan.
3. Penelitian ini juga membutuhkan *microsoft excel* 2019 dalam hal proses pencarian data secara manual untuk nilai *gain* dan *entropy* hingga bisa di bentuk pohon keputusan yang diinginkan.

1.4 Rumusan masalah

Berdasarkan identifikasi masalah, maka rumusan masalah yang akan dibahas antara lain:

1. Bagaimana proses dari data mining dan proses klasifikasi decision tree dengan perhitungan algoritma dari C4.5 dalam menganalisa kriteria kamar kos layak huni di Kota Batam?
2. Apakah teknik klasifikasi menggunakan algoritma C4.5 dapat meneukan pola (pohon keputusan yang dihasilkan) dalam menganalisis kriteria kamar kos layak huni pada Kota Batam?
3. Bagaimana cara mengimplementasikan teknik klasifikasi algoritma C4.5 sebagai alat untuk menemukan pola yang digunakan dalam menganalisis kriteria kamar kos layak huni?
4. Bagaimana tindakan kedepannya dengan adanya hasil yang diperoleh dari hasil penelitian yang dilakukan mengenai kriteria kamar kos layak huni?

1.5 Tujuan penelitian

Berikut merupakan tujuan penelitian berdasarkan rumusan masalah yang sudah disusun sebelumnya:

1. Menerapkan data mining khususnya teknik klasifikasi menggunakan algoritma C4.5 dalam hal menganalisis kriteria kamar kos layak huni pada Kota Batam.
2. Pengujian teknik klasifikasi dan algoritma C4.5 dalam wujud penerapan dari data mining yang digunakan dalam menganalisis kriteria kamar kos layak huni pada Kota Batam.
3. Pengujian Algoritma C4.5 menggunakan *software* data mining Weka versi 3.9.5 untuk menganalisis kriteria kamar kos layak huni pada Kota Batam.
4. Menguji dan menganalisa penerapan data mining teknik klasifikasi menggunakan algoritma C4.5 dalam menganalisis kriteria kamar kos layak huni pada Kota Batam.

1.6 Manfaat Penelitian

Demikian penjelasan penulis mengenai manfaat penelitian yang diharapkan dapat diterima oleh pihak lain atau peneliti lainnya, yaitu:

1.6.1 Manfaat Teoritis

Demikian penjelasan penulis mengenai manfaat teoritis yang diharapkan dapat diterima oleh pihak lain atau peneliti lainnya, yaitu:

1. Hasil penelitian ini dapat menambah dan memperkaya referensi dan teori yang berkaitan dengan penggunaan teknik data mining dengan perhitungan algoritma dari C4.5 dalam menganalisis kelayakan kamar kos.
2. Hasil dari penelitian dapat dijadikan pedoman saat peneliti lain ingin melakukan penelitian dengan teknik data mining dengan perhitungan algoritma C4.5.

1.6.2 Manfaat Praktis

Demikian penjelasan penulis mengenai manfaat teoritis yang diharapkan dapat diterima oleh penulis sendiri, yaitu:

1. Bagi penulis
 - a. Penerapan ilmu yang diperoleh penulis selama kuliah di Universitas Putera Batam, khususnya pada kajian teknik data mining dan komputasi algoritmik C4.5.
 - b. Untuk melengkapi persyaratan agar bisa lulus dari skripsi Sistem Informasi.
2. Bagi Mahasiswa
 - a. Mendiskusikan metode perhitungan algoritmik dan metode pensiun, khususnya di bidang real estat, sesuai panduan belajar bagi pengguna data tambahan dan teknik penambangan data.
3. Bagi Masyarakat yang mendirikan rumah kos
 - a. Bagi masyarakat yang mendirikan rumah kos yang berada di Kota Batam, dapat menjadi sumber berupa informasi yang akurat mengenai kriteria rumah kos layak huni.

- b. Membantu masyarakat yang mendirikan rumah kos dalam melakukan perbaikan rumah kos yang didirikan.

BAB II

TINJAUAN PUSTAKA

2.1 Teori

Dalam subbab ini kita membahas teori, menafsirkannya sebagai sekumpulan bagian atau variabel yang saling terkait, definisi dan proposal yang bertujuan untuk menjelaskan fenomena alam dengan membangun hubungan antara variabel yang dipelajari dan memberikan penjelasan fenomena secara sistematis.

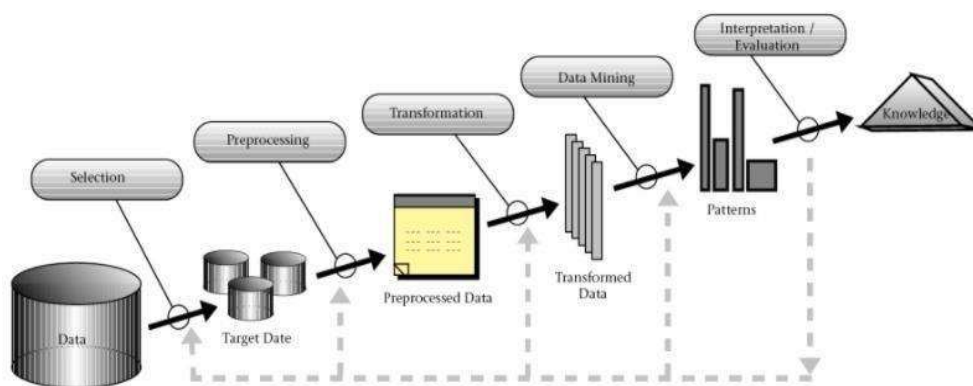
2.1.1 Teori Umum

Sub Bab ini akan membahas teori utama di balik karya penelitian ini, termasuk ide-ide utama.

2.1.1.1 *Data Mining*

Penambangan data adalah sebuah proses mencari pola-pola yang ada dan data dari sejumlah besar data yang disimpan dalam basis data atau sumber data lainnya. perbedaan; Sangat berguna untuk mengekstrak wawasan yang berguna seperti struktur dan cacat. Penambangan data banyak digunakan dalam waktu beberapa tahun terakhir karena ketersediaan informasi yang didapat dalam jumlah besar dalam bentuk elektronik dan kebutuhan untuk mengubah informasi ini menjadi informasi dan pengetahuan yang berguna dalam sistem besar. Aplikasi ini dapat ditemukan di berbagai bidang seperti kecerdasan buatan, pembelajaran mesin, analisis pasar, statistik, sistem basis data, manajemen bisnis, dan dukungan

keputusan. Penambangan data adalah proses menganalisis dan mengekstraksi pengetahuan secara otomatis menggunakan satu atau lebih teknik pembelajaran mesin. Tujuan dari data mining adalah untuk menemukan pola global dalam data yang ada. Untuk mengekstraksi pengetahuan dari konsensus yang ada, perlu menggunakan metode ilmiah yang dikenal sebagai *Knowledge Discovery in Database* (KDD) dalam database. (Sugara et al. 2018).



Gambar 2.1 Tahapan proses KDD

Sumber: Sugara 2018

Proses analisis data (KDD) terdiri dari 5 (lima) langkah, yang dapat diuraikan berikut ini :

1. *Selection*

Kesatu, dapat memilih data yang akan diolah dari beberapa sumber data dengan membersihkan data yang bertentangan atau nilai yang hilang.

2. *Preprocessing*

Kedua, mengintegrasikan semua sumber informasi yang dikumpulkan sebelum melanjutkan ke proses selanjutnya.

3. *Transformation*

Ketiga, ubah data menjadi format yang lebih cocok untuk penambangan data.

4. *Data Mining*

Gunakan pemrosesan data dan sistem untuk membangun model dari data yang dikumpulkan.

5. *Interpretation/ Evaluation*

Kelima, interpretasikan dan evaluasi pola yang diterima sehingga mengidentifikasi atau mewakili pola yang ingin terima.

Saat menggunakan KDD, teknik analisis data adalah mengekstraksi pola dari data. Skema dapat dibuat dalam operasi basis data. Secara umum, ada dua jenis metode penambangan data: metode analisis data, yang menggambarkan karakteristik utama dari data yang ada, dan metode penambangan data deskriptif, yang berupaya membuat prediksi berdasarkan data yang ada. Ekstraksi informasi dapat dilakukan dalam bentuk data kuantitatif, tekstual atau multimedia.

Aplikasi penambangan data dapat menggunakan parameter yang berbeda untuk menganalisis data. Ini termasuk korelasi (bagaimana satu faktor berhubungan dengan yang lain), konvergensi atau analisis jalur (bagaimana satu faktor mengarah ke yang lain), diferensiasi (menemukan desain baru dengan tujuan tertentu), dan integrasi (pengelompokan satu atau serupa)..

Dalam penambangan data, ekstraksi data dapat mencakup beberapa langkah dasar, yang dijelaskan di bawah ini:

1. *Problem definition*

Langkah pertama adalah menentukan tujuan. Bergantung pada tujuannya, seperangkat alat yang tepat dapat diterapkan pada data untuk memodelkan perilaku.

2. *Data exploration*

Jika kualitas data tidak cukup tinggi untuk model yang akurat, buat rekomendasi untuk data mendatang. Pada tahap ini teknik pengumpulan dan penyimpanan dapat diterapkan. Untuk keperluan analisis, semua data harus dikumpulkan agar dapat diolah secara konsisten.

3. *Data preparation*

Tujuan dari proses ini adalah untuk membersihkan dan memodifikasi untuk memperhitungkan nilai yang hilang dan salah, dan memperlakukan semua fakta yang diketahui secara konsisten untuk analisis yang lebih kuat.

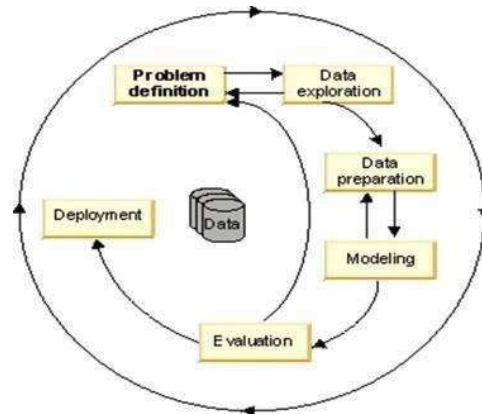
4. *Modeling*

Berdasarkan data dan hasil yang diinginkan, pilih algoritma penambangan data atau kombinasi algoritma untuk analisis. Algoritme ini berkisar dari teknik klasik seperti statistik, lingkungan, dan pengelompokan hingga teknik generasi berikutnya seperti pohon keputusan, jaringan, dan algoritme berbasis aturan. Pemilihan algoritma tertentu bergantung pada tujuan spesifik yang ingin dicapai dan kualitas data yang dianalisis.

5. *Evaluation and Deployment*

Berdasarkan hasil algoritma data mining, analisis dilakukan untuk menentukan kesimpulan utama dari analisis dan mengembangkan

seperangkat rekomendasi untuk dipertimbangkan.



Gambar 2.1 Representasi proses data mining

Sumber: Sugara 2018

2.1.1.2 Algoritma C4.5

Algoritma C4.5 adalah program yang menerima sinyal dan menghasilkan algoritma keputusan. Pohon keputusan ini divalidasi terhadap data yang belum diuji untuk pelatihan. Program membuat aturan klasifikasi menggunakan pohon keputusan untuk kumpulan data tertentu. Sistem C4.5 merupakan perpanjangan dari sistem ID3 asli yang dikembangkan oleh Quinlan. C4.5 adalah salah satu kurva pembelajaran. Algoritma C4.5 menggunakan konsep entropi informasi untuk membangun pohon keputusan dari set pelatihan, mirip dengan algoritma ID3. C4.5 Klasifikasi statistik. (Sugara et al. 2018)

Algoritma dapat memecahkan masalah dalam sistem dengan membuat keputusan berdasarkan langkah-langkah berikut:

1. Pilih perangkat sebagai *root*.
2. Buat cabang untuk setiap rekaman aset.
3. Bagilah benda-benda menjadi cabang-cabang,

4. Ulangi proses untuk setiap cabang sampai semua kasus di cabang terselesaikan dengan baik.

Nilai atribut dihitung dengan memilih atribut sebagai baris akar. Nilai keuntungan maksimum digunakan sebagai dasar dari pohon keputusan. Rumus ini digunakan untuk menghitung nilai keuntungan:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Rumus 1. Rumus *gain*

Keterangan:

S= Himpunan Kasus

A= Atribut

n= Jumlah partisi atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut ini.

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i$$

Rumus 2. Perhitungan Nilai *Entropy*

Keterangan:

S = Himpunan Kasus A = Fitur

n = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

Setelah perhitungan selesai, pohon keputusan dibuat, yang dapat diubah menjadi aturan dengan aturan *IF-THEN-ELSE*. Aturan-aturan tersebut dijelaskan

sebagai berikut:

1. *IF-THEN Rule*

Aturan Induktif: Proses menurunkan aturan "jika maka" yang berguna dari data berdasarkan signifikansi statistik. Sistem berbasis aturan membuat seperangkat aturan jika-maka. Pengetahuan sudah merupakan kumpulan bentuk di mana strategi penyimpanan dapat dirancang. Semua data yang digunakan untuk analisis harus dikumpulkan sehingga dapat diproses secara terus menerus.

2. *IF* kondisi *THEN* Kesimpulan

Aturan ini memiliki dua bagian. Perintah (klausa *IF*) mendefinisikan satu atau lebih kondisi di sekitar nilai prediksi yang berisi prediksi tentang nilai properti target berdasarkan atribut perintah (klausa *THEN*). Evaluasi yang akurat dari nilai atribut objektif meningkatkan pengambilan keputusan. Aturan umum untuk estimasi *IF-THEN* dalam data mining. Mereka memberikan abstraksi ekstrim dari fungsi pencarian. Metode penalaran berbasis aturan dapat membuat prediksi dengan kasus yang dikecualikan.

2.1.1.3 Pohon Keputusan (*decision tree*)

Pohon keputusan adalah representasi pengetahuan yang diatur oleh node dan cabang di pohon, sehingga setiap node non-teks menggunakan atributnya. Label domain dari node internal berisi nilai atribut dari node tersebut. Setiap node disebut kelas (nilai target karakteristik). Model berbasis pohon, termasuk klasifikasi dan regresi pohon, adalah aplikasi umum model induktif. Model pohon keputusan paling cocok untuk pemrosesan data. Mereka lebih mudah untuk dibangun, lebih

mudah diinterpretasikan, lebih mudah diintegrasikan dengan sistem basis data, dan memiliki akurasi implementasi yang sama atau lebih baik. Ada banyak algoritme pohon keputusan seperti algoritme HUNTS (ini adalah salah satu algoritme paling awal), CART, ID3, C4.5 (versi terbaru dari algoritme ID3), SLIQ, SPRINT. (Sugara et al. 2018)

Dalam pembelajaran pohon keputusan, ID3 (*Iterative Dichotomizer 3*) adalah algoritma yang dikembangkan oleh Ross Quinlan untuk mengekstraksi pengetahuan keputusan dari dataset. ID3 sering digunakan dalam pembelajaran mesin dan pemrosesan bahasa alami. Metode pengurangan termasuk membangun pohon yang menunjukkan proses klasifikasi.. Setelah pohon dibangun, itu diterapkan ke setiap grup dalam database dan membuat taksonomi untuk grup itu. Masalah yang dihadapi oleh sebagian besar algoritma pohon keputusan adalah sebagai berikut:

1. Memilih atribut pemisahan
2. Memesan atribut pemisahan
3. Jumlah perpecahan yang harus diambil
4. Saldo struktur pohon dan pemangkasan
5. Menghentikan kriteria

2.1.2 Teori Khusus

Dalam sub bagian ini, kami membahas prinsip-prinsip khusus yang menarik terkait dengan pelaksanaan penelitian ini.

2.1.2.1 Rumah Layak Huni

Untuk menentukan apa yang dimaksud dengan habitat layak huni, peneliti

menerapkan prinsip-prinsip Hukum Perpajakan No. 1 tahun 2011 Bagian 1 tentang *real estat*, yang mendefinisikan *real estat* sebagai rencana penting, pengelolaan perumahan, pengelolaan perumahan, pemeliharaan dan perbaikan. Pencegahan dan perbaikan permukiman kumuh dan kualitas hidup di permukiman kumuh, kebijakan pertanahan, pembiayaan dan sistem pembayaran serta peran masyarakat. UUD No. 1 Pasal 19 Tahun 2011 perumahan dan pengenalan perumahan dalam rangka memenuhi kebutuhan perumahan, salah satu kebutuhan dasar manusia, untuk memajukan kesejahteraan dan pemerataan masyarakat, sebagaimana disebutkan dalam pasal 1, pemerintah ini, yang dikelola oleh suatu daerah. Pemerintah dan/atau seluruh rakyat harus menikmati hidup dalam lingkungan yang sehat, aman, serasi, dan tertib serta merupakan hak setiap warga negara.

2.1.2.2 Indekos

Indekos adalah tempat tinggal sementara yang disewakan kepada wisatawan atau orang yang tinggal jauh dari tempat tujuan biasanya, terutama kepada pekerja atau pekerja yang jauh dari tempat kerjanya. Seiring pertumbuhan populasi, mereka yang ingin tinggal di rumah sering menghadapi tantangan. Banyak kendala internal adalah banyaknya pemangku kepentingan. Di luar itu, pencari kerja di pesantren kebanyakan adalah orang yang baru pindah rumah, berasal dari berbagai daerah dan belum mengenal daerah Batam, sehingga perlu diketahui secara lengkap tentang pesantren tersebut. (Rianto 2019).

2.1.2.3 Kelayakan

Studi kelayakan atau feasibility study ini adalah proses analisis mendalam terhadap suatu proyek atau perusahaan atau bisnis untuk menentukan apakah bisnis dapat dilanjutkan.

Kelayakan berarti analisis mendalam untuk menentukan apakah manfaat bisnis lebih besar daripada biayanya. Dengan kata lain, kapasitas dapat diartikan bahwa usaha yang dikelola akan memberikan keuntungan finansial dan non finansial sesuai dengan tujuan yang dituju. Ditafsirkan dengan benar di sini, bermanfaat tidak hanya bagi perusahaan yang mengelolanya, tetapi juga bagi investor, pemberi pinjaman, pemerintah, dan masyarakat secara keseluruhan. (R Nurmalina, T Sarianti 2018).

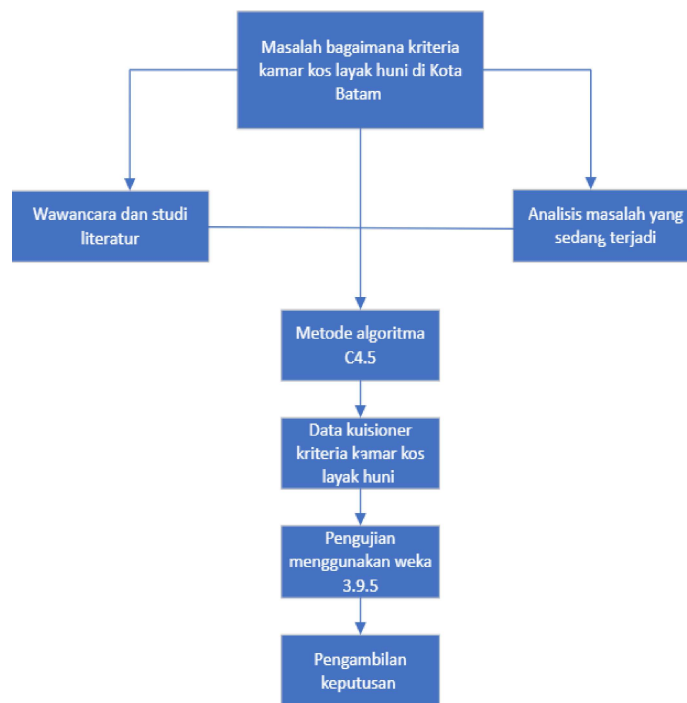
2.1.2.4 Software Weka

Weka (*Waikato Environment for Knowledge Analysis*) Lingkungan Analisis Pengetahuan Waikato dikembangkan oleh Universitas Waikato di Selandia Baru untuk penelitian, pendidikan, dan berbagai aplikasi. Weka mampu memecahkan masalah penambangan data dunia nyata, terutama masalah klasifikasi, yang merupakan inti dari teknik pembelajaran mesin. Perangkat lunak ini ditulis menggunakan pendekatan berorientasi objek dalam hierarki kelas Java dan berjalan di hampir semua platform. Weka mudah digunakan dan diimplementasikan pada beberapa tingkatan. Ada implementasi algoritme pembelajaran modern yang dapat diterapkan ke database dari baris perintah. WEKA meliputi preprocessing data, klasifikasi, regresi, pengelompokan, aturan asosiasi, dan visualisasi. Pengguna dapat memproses data terlebih dahulu, memasukkannya ke dalam program pembelajaran, dan menganalisis serta melakukan klasifikasi yang dihasilkan tanpa

menulis kode program apa pun. Inferensi adalah salah satu dari banyak teknik data mining yang dapat digunakan di pohon keputusan (Cobit and Utami 2019)

2.2 Kerangka Pemikiran

Model kerangka pemikiran berdasarkan penelitian yang dilakukan, yaitu sebagai berikut:



Gambar 2.2 Kerangka pemikiran

Sumber: (Penulis, 2022)

2.3 Hipotesis Penelitian

Berikut ini adalah hipotesis dari hasil penelitian berdasarkan dugaan penulis penelitian ini:

1. Teknik data mining dengan perhitungan algoritma C4,5 dapat menganalisa kriteria kamar kos layak huni di Kota Batam.
2. Hasil dari penelitian ini dapat mengetahui kriteria kamar kos yang diminati oleh karyawan di Kota Batam.

2.4 Penelitian Terdahulu

Berikut ini adalah informasi yang berupa referensi dari penelitian terdahulu yang berkaitan tentang penelitian ini yang di rangkum sebagai berikut:

1. Dengan judul “Penerapan *data mining* menggunakan algoritma c4.5 dalam menentukan kelayakan penerima bantuan bedah rumah pada desa tiga dolok” yang dilakukan oleh (V.M. et al. 2021) penelitian ini menunjukkan bahwa menentukan penerima bantuan bedah rumah pada desa tiga dolok dengan menggunakan algoritma c4.5 akan bermanfaat sekali dalam proses menentukan penerima bantuan bedah rumah.
2. Dengan judul “Penerapan data mining untuk prediksi penjualan Readymix menggunakan metode algoritma c4.5 pada PT.Remicon widyaprima” yang dilakukan oleh (Harman 2020) penelitian ini menunjukkan bahwa Mencari perkiraan penjualan yang telah disiapkan dihitung menurut algoritma c4.5 menggunakan metode data mining, menggunakan data penjualan di PT. Remicon Vidiaprim sebagai referensi dan panduan yang memungkinkan untuk pohon keputusan.

3. Dengan judul “Penerapan algoritma c4.5 untuk deteksi dini gangguan autisme pada anak” yang dilakukan oleh (Sugara et al. 2018) penelitian ini menunjukkan deteksi dini autisme pada anak dapat diprediksi dengan menggunakan metode data mining yang menggunakan algoritma C 4.5 untuk memprediksi autisme pada anak berdasarkan gejala yang dialami anak tersebut.
4. Dengan judul “Perbandingan algoritma C4.5 dan *Naïve Bayes* untuk menentukan kelayakan penerima bantuan program keluarga harapan” (Fitriani 2020) penelitian ini menunjukkan bahwa algoritma C4.5 merupakan metode yang cukup baik dalam memprediksi kelayakan warga dalam menerima bantuan Program Keluarga Harapan.
5. Dengan judul “Membangun model algoritma C4.5 dalam menentukan kelayakan rumah ibadah berdasarkan indicator epidemiologi pada masa pandemic covid-19” (Amaliyyah 2021) penelitian ini menunjukkan bahwa berdasarkan dataset perkembangan kasus covid-19 perkelurahan, maka terbentuklah model C4.5 dalam bentuk rule dan pohon keputusan yang dapat digunakan untuk menentukan kluster terkait Kelayakan rumah ibadah.
6. Dengan judul “Aplikasi berbasis web untuk klasifikasi kelayakan pemberian kredit pada koperasi Bumdes Mitra Baru menggunakan algoritma C4.5 (Ulfa 2020) penelitian ini menunjukkan bahwa akurasi model pohon keputusan yang di hasilkan algoritma C4.5 sangat akurat.
7. Dengan judul “Komparasi algoritma C4.5 dan *naïve bayes* untuk

menganalisa kelayakan pengajuan kredit” (Khasanah 2019) penelitian ini menunjukkan bahwa untuk penentuan kelayakan kredit apakah calon nasabah akan menjadi nasabah yang pembayarannya lancar atau bermasalah dengan menggunakan algoritma C4.5 dan naïve bayes.

8. Dengan judul “Penerapan algoritma C4.5 untuk penentuan kelayakan pemberian kredit” (Santoso and Sekardiana 2019) penelitian ini menunjukkan bahwa dengan hasil evaluasi penelitian bahwa algoritma C4.5 mendapatkan nilai akurasi data training serta data testing menggunakan algoritma C4.5.
9. Dengan judul “Penerapan algoritma C4.5 untuk kelayakan penerimaan uang dan beras” (Rika Harman 2018) penelitian ini menunjukkan bahwa dihasilkannya keputusan yaitu berupa perhitungan jumlah kepala-kepala keluarga yang menerima bantuan uang dan beras dan yang tidak menerima bantuan uang dan beras yang mana dalam proses perhitungan node cukup satu kali perhitungan saja.
10. Dengan judul “Analisis Algoritma C4.5 untuk menentukan factor pembelian sepeda bekas pada took sepeda Batam” (Internasional and Ananda 2021) penelitian ini menunjukkan bahwa dihasilkannya keputusan yaitu berupa sejumlah opsi aktivasi dapat dibuat untuk mendemonstrasikan hal ini, termasuk: a). Jika roda sepeda bengkok, sepeda tidak akan dibawa; b) Jika roda sepeda tidak bengkok atau normal dan setang sepeda berayun, sepeda tidak akan dibeli; dan c). Jika roda sepeda bengkok atau patah, dan setang sepeda tidak bergerak atau terasa

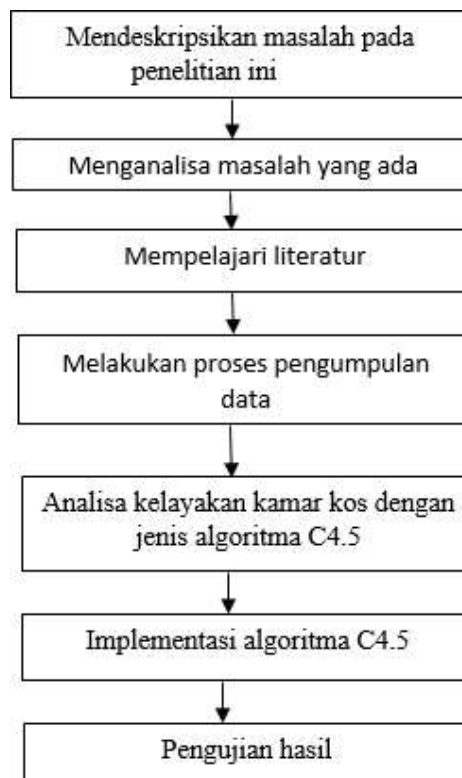
nyaman, maka sepeda akan dibeli.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Dalam penelitian ini perkembangan tersebut terkait dengan efisiensi hotel di kota Batam. Garis besar desain penelitian adalah sebagai berikut:



Gambar 3.1 Desain penelitian

Sumber: (Penulis,2022)

1. Mendeskripsikan Isi Dari Masalah

Para peneliti mencoba menggunakan metode algoritma pada C4.5 untuk menjelaskan pertanyaan penelitian tentang kelayakan ruang tamu.

2. Melakukan Analisa Terhadap Masalah

Pada tahap ini, peneliti menganalisis masalah yang diuraikan dan memahami masalah yang diidentifikasi.

3. Mempelajari Literatur

Peneliti harus mencoba membaca dan memahami berbagai model yang mendukung penelitian jurnal sehingga penulis dapat memandu penelitian mereka.

4. Melakukan Proses Pengumpulan Data

Pengumpulan data dilakukan melalui teknik wawancara dan pemberian kuesioner. Peneliti melakukan kegiatan dengan memberikan pertanyaan dan menyebarkan link kuisisioner kepada karyawan yang tinggal di dormitory dan karyawan yang sedang menentukan kamar kos.

5. Analisa Kelayakan Kamar kos Dengan Jenis Algoritma Dari C4.5

Data yang diperoleh diolah dengan algoritma C4.5, mulai dari proses penentuan pohon keputusan hingga pembuatan data yang dibutuhkan oleh peneliti.

6. Implementasi Jenis Algoritma Dari C4.5

Penulis memeriksa kelengkapan data yang digunakan sebagai sumber pohon keputusan kemudian menghitung nilai kelengkapan yang paling berguna untuk semua data yang digunakan untuk membangun pohon keputusan.

7. Pengujian Hasil

Pada tahap akhir ini dilakukan pengecekan hasil pengolahan data secara

manual menggunakan algoritma C4.5 dengan WEKA versi 3.9.5.

3.2 Objek Penelitian

Pada subbab objek penelitian, peneliti ingin memperjelas gambaran tentang lokasi penelitian sebagai berikut:

3.2.1 Lokasi Penelitian

Lokasi penelitian ini dilakukan di Kota Batam, Kepulauan Riau terkhusus di daerah dormitori dan para karyawan yang sedang menentukan kamar kos.

3.2.2 Jadwal Penelitian

Tabel 3.1 Jadwal penelitian

Uraian	Bulan					
	Sep-22	Okt-22	Nov-22	Des-22	Jan-23	Feb-23
1 Pengajuan judul	■					
2 Bimbingan skripsi		■	■	■	■	■
3 Studi literatur	■					
4 Pengajuan izin penelitian		■				
5 Pengumpulan data		■	■	■		
6 Pengelolahan data					■	■
7 Penyusunan laporan	■	■	■	■	■	■

3.3 Populasi dan Sampel

Pada subbab ini, peneliti akan membahas suatu populasi dari sampel yang didapat sebagai literatur dalam penelitian ini:

3.3.1 Populasi

Populasi penelitian ini adalah data hasil wawancara dan penyebaran kuesioner terhadap karyawan yang tinggal di dormitori ataupun yang sedang menentukan kamar kos yang ada di Kota Batam.

3.3.2 Sampel

Data yang akan dijadikan sampel pada penelitian ini adalah sebanyak 70 koresponden yang merupakan karyawan yang bekerja di Kota Batam.

3.4 Variabel Penelitian

Pengolahan data dilakukan sesuai data karyawan yang tinggal di kamar kos. Beberapa indikator yang digunakan dalam penelitian khususnya yang berkaitan dengan variabel dari kelayakan rumah kos, yaitu :

1. Ukuran kamar
2. Ventilasi udara
3. Ketersediaan air
4. Sarana MCK
5. Harga

Untuk hasil akhir keputusannya adalah layak huni dan tidak layak huni.

3.4.1 Teknik Pengumpulan Data

Mengumpulkan data untuk mendapatkan informasi yang diperlukan bagi peneliti untuk mencapai tujuan penelitian ini yaitu:

1. Studi literatur

Materi penelitian dari penelitian sebelumnya, namun masih ada makalah penelitian yang ingin penulis temukan.

2. Kuesioner

Proses kegiatan dengan memberikan kuesioner berisi pertanyaan menggunakan indikator dari variabel kelayakan kamar kos karyawan yang

bekerja dan menentukan kamar kos di kota Batam. Yang akan menjadi acuan dalam menentukan perhitungan.

3. Wawancara

Melakukan kegiatan tanya jawab kepada karyawan yang sedang menentukan kamar kos di Kota Batam.

3.4.2 Model penelitian

Dalam model metode penelitian ini, peneliti menggunakan teknik data mining dan teknik klasifikasi algoritma C4.5. Selain itu, pada tahap pendahuluan ini akan dilakukan klasifikasi sesuai dengan kategori penelitian ini. Dapat menggunakan metode Algoritma C4.5 untuk menghitung hasil langkah pertama yang dilakukan, mendapatkan nilai entropi terlebih dahulu, dan membangun pohon keputusan. Proses ini dilakukan secara iteratif, dalam hal ini kita dapat menggunakan perangkat lunak Weka dan menggunakan Microsoft Excel untuk menghitungnya, hingga akhirnya kita dapat menemukan simpul-simpul pada pohon keputusan, jika koridor salah satu parameter aturan dasar yang akan dievaluasi adalah contoh kota Batam. Berdasarkan pola itulah kesimpulan dari permasalahan dapat diambil dan solusi bisa didapat.