

**PENERAPAN ALGORITMA C4.5 UNTUK
MEMPREDIKSI PROFIT PENJUALAN**

SKRIPSI



**Oleh:
Calvin
181510006**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
2022**

PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PROFIT PENJUALAN

SKRIPSI

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



**Oleh:
Calvin
181510006**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
2022**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Calvin
NPM/NIP : 181510006
Fakultas : Teknik dan Komputer
Program Studi : Sistem Informasi

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

Penerapan Algoritma C4.5 Untuk Memprediksi Profit Penjualan

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, 06 Agustus 2022



Calvin

181510006

**PENERAPAN ALGORITMA C4.5 UNTUK PROFIT
PENJUALAN**

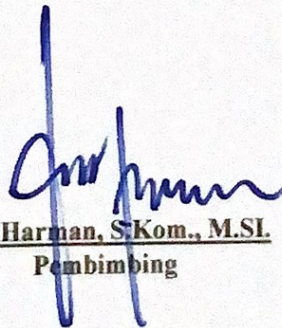
Oleh
Calvin
181510006

SKRIPSI

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

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 06 Agustus 2022



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



Universitas Putera Batam

ABSTRAK

Tak bisa dihindari kecanggihan teknologinya dan informasinya telah menggiring perkembangan di keseluruhan bidang, mencakup berkembangnya bisnis usaha. Pesatnya serta berkembang teknologinya dan informasinya kini mencuri perhatiannya sekelompok besar masyarakatnya. Berubahnya perkembangan komputer dari waktu ke waktu dapat menghasilkan informasi dengan cepat, akurat dan tanpa memakan waktu lama untuk mencari informasi yang mereka butuhkan, termasuk pada toko-toko yang menjual barang-barang pakaian, dimana usahanya masih memerlukan perhitungan informasi kritis, penjualan yang mereka kelola setiap tahun. Yanyan Boutique ialah tokonya yang bergerak dipenjualan *online* yang menyediakan produknya seperti pakaian, dompet, dan tas. Yanyan Boutique berdiri ditahun 2018 dan berlokasi di kios Pasir Putih blok A No. 20, Batam. Di karenakan berubahnya aturan bea cukai yang memengaruhi pengirimannya, tiap pengirimannya dari Batam di kenakan pajaknya yang besar. Hal inilah menyebabkan turunnya penjualannya yang begitu besar. Didalam menganalisa inilah penelitiannya menerapkan teknik *data mining* dengan pengklasifikasian algoritma C4.5 yang di lakukan dengan menerapkan perhitungannya secara manual dan menerapkan aplikasi WEKA 3.9.5 didalam menelusuri hasil prediksinya profit penjualan di Yanyan Boutique. Penelitiannya memakai 2 variabel keputusannya yakni Beli dan Tidak Beli.

Kata Kunci: Algoritma C4.5, Analisis,*Data Mining*, Profit.

ABSTRACT

The development of technology and information has changed all fields, including the development of the business world. It's hard to deny that today's rapidly evolving technology is attracting the attention of many people, including online marketers. An online store business needs to calculate all the important information to run the business plan they manage every year. Yanyan Boutique is a store engaged in online sales or an online store that provides products such as men's and women's clothing, men's and women's wallets, men's and women's bags. Yanyan Boutique started operating in 2018 and is located at the Pasir Putih kiosk block A no 20, Batam Center, Batam City, Riau Islands. Due to changes in customs regulations affecting shipments, every shipment from Batam is subject to heavy taxation, This results in a huge drop in sales. To perform this analysis, the researcher uses data mining techniques with the classification of the C4.5 algorithm which is carried out using manual calculations and the researcher uses the WEKA 3.9.5 application to find out the prediction results of sales profit at Yanyan Boutique. Researchers used 2 decision variables, namely Buy and Not Buy.

Keywords: *Algorithm C4.5, Analysis, Data Mining, Profit.*

KATA PENGANTAR

Dengan mengucapkan puji dan syukur kepada Tuhan yang Maha Esa yang telah melimpahkan segala rahmat dan karunianya, sehingga penulis dapat menyelesaikan skripsi yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terimakasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI, selaku Rektor Universitas Putera Batam.
2. Bapak Welly Sugianto, S.T., M.M, selaku Dekan Fakultas Teknik Dan Komputer Universitas Putera Batam.
3. Bapak Muhammad Rasyid Ridho, S.Kom., M.SI., selaku Ketua Program Studi Sistem Informasi Universitas Putera Batam.
4. Bapak Rika Harman, S.Kom., M.SI., sebagai pembimbing skripsi yang telah menyediakan waktu dalam memberikan bimbingan dan petunjuk sampai selesainya skripsi ini.
5. Seluruh dosen dan staff Universitas Putera Batam.
6. Dan pihak-pihak yang telah memberikan dukungannya hingga tersusunnya penelitian ini yang tidak dapat disebutkan oleh penulis satu persatu.

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

Batam, 06 Agustus 2022

Penulis



Universitas Putera Batam

DAFTAR ISI

ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	ix
DAFTAR RUMUS	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	6
1.3 Batasan Masalah.....	6
1.4 Rumusan Masalah	6
1.5 Tujuan Penelitian.....	7
1.6 Manfaat Penelitian.....	8
1.6.1 Manfaat Teoritis	8
1.6.2 Manfaat Praktis	8
BAB II TINJAUAN PUSTAKA.....	9
2.1 Teori Dasar	9
2.1.1 Teori Umum.....	9
2.1.2 Teori Khusus	19
2.2 Penelitian Terdahulu.....	22
2.3 Kerangka Pemikiran	25
BAB III METODE PENELITIAN	26
3.1 Desain Penelitian	26
3.2 Objek Penelitian	28
3.2.1 Lokasi Penelitian.....	28
3.2.2 Jadwal Penelitian.....	28
3.3 Populasi dan Sampel	29
3.3.1 Populasi.....	29
3.3.2 Sampel.....	29

3.4	Variabel Penelitian	29
3.5	Teknik Pengumpulan Data	30
3.6	Metode Penelitian	31
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Hasil Penelitian.....	32
4.2	Pembahasan	32
4.2.1	Melakukan Pra-Proses Data Penelitian	33
4.2.2	Analisis Data	34
4.2.3	Seleksi Data.....	35
4.2.4	Transformasi Data	35
4.2.5	Pohon Keputusan	38
4.2.6	Pengujian Menggunakan Aplikasi WEKA	60
BAB V SIMPULAN DAN SARAN		65
5.1	Simpulan.....	65
5.2	Saran	66
DAFTAR PUSTAKA		67

DAFTAR GAMBAR

	Halaman
Gambar 1.1 Data penjualan Toko Yanyan Boutique.....	4
Gambar 2.1 Proses KDD.....	11
Gambar 2.2 Model Pohon Keputusan.....	17
Gambar 2.3 Kerangka Pemikiran.....	25
Gambar 3.1 Desain Penelitian.....	26
Gambar 4.1 Pohon keputusan pada node pertama.....	46
Gambar 4.2 Pohon keputusan pada node kedua.....	52
Gambar 4.3 Pohon keputusan pada node ketiga.....	57
Gambar 4.4 Data Uji.xls.....	59
Gambar 4.5. Data di Notepad.csv.....	59
Gambar 4.6 Tampilan WEKA 3.9.5.....	60
Gambar 4.7 Open file pada WEKA 3.9.5.....	60
Gambar 4.8 Tampilan atribut WEKA 3.9.5.....	61
Gambar 4.9 Tampilan <i>classify</i> WEKA 3.9.5.....	61
Gambar 4.10 Tampilan <i>classify</i> J48 WEKA 3.9.5.....	62
Gambar 4.11 Tampilan setelah <i>classify</i> WEKA 3.9.5.....	62
Gambar 4.12 Hasil <i>Visualize Tree</i> WEKA 3.9.5.....	63

DAFTAR TABEL

	Halaman
Tabel 3.1 Jadwal Penelitian.....	29
Tabel 4.1 Format Data Penjualan.....	33
Tabel 4.2 Klasifikasi Atribut Ketersediaan.....	35
Tabel 4.3 Klasifikasi Atribut Kualitas.....	35
Tabel 4.4 Klasifikasi Atribut Harga.....	36
Tabel 4.5 Klasifikasi Atribut Pengiriman.....	36
Tabel 4.6 Klasifikasi Atribut Promosi.....	37
Tabel 4.7 Pra Proses Setelah Klasifikasi.....	37
Tabel 4.8 Hasil perhitungan node pertama.....	45
Tabel 4.9 Hasil perhitungan node kedua.....	51
Tabel 4.10 Hasil perhitungan node ketiga.....	57

DAFTAR RUMUS

	Halaman
Tabel 2.1 Perhitungan Entropy.....	18
Tabel 2.2 Perhitungan Gain.....	18



Universitas Putera Batam

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tak bisa dihindari kecanggihan teknologinya dan informasinya telah menggiring perkembangan di keseluruhan bidang, mencakup berkembangnya bisnis usaha. Pesatnya serta berkembang teknologinya dan informasinya kini mencuri perhatiannya sekelompok besar masyarakatnya. Berubahnya perkembangan komputer dari waktu ke waktu dapat menghasilkan informasi dengan cepat, akurat dan tanpa memakan waktu lama untuk mencari informasi yang mereka butuhkan, termasuk pada toko-toko yang menjual barang-barang pakaian, dimana usahanya masih memerlukan perhitungan informasi kritis, penjualan yang mereka kelola setiap tahun. Perkembangan di era globalisasi hingga saat ini memang sangat pesat, mendesak semua bidang bisnis yang ada untuk bisa saling bersaing. Keberadaan sistem operasional sangat bermanfaat, meskipun tidak semuanya dapat memberikan informasi yang berharga seperti yang diharapkan para perusahaan (Wijaya & Fauzi, 2020).

Prakiraan ataupun ramalan penjualan (*forecast*) adalah perhitungan untuk memprediksi kondisi masa depan dengan menguji kondisi dimasa lampau. Melakukan peramalan penjualannya dimasa depan artinya menetapkan prakiraan volume penjualannya dan bahkan menetapkan berpotensi atau tidakkah penjualannya hingga ukuran pasarnya yang akan di kuasai dimasa depan. Salah satu fungsi peramalan adalah untuk membantu pemilik bisnis membuat keputusan

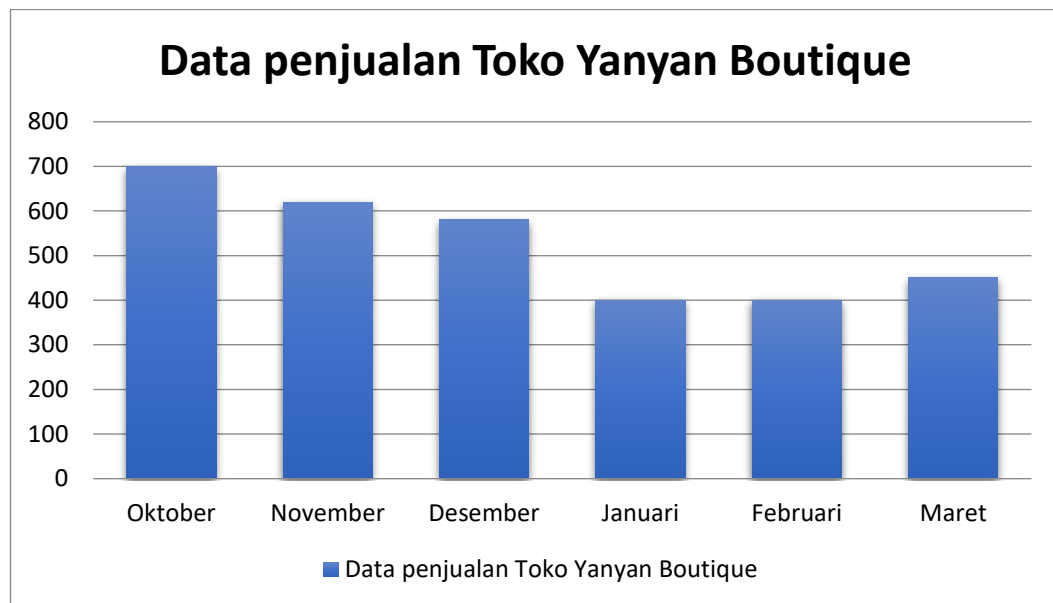
tentang berapa banyak untuk memasok bisnis mereka. Selanjutnya, peramalan dapat membantu rencana bisnis persediaan stok, sebab peramalan ini bisa memerikan hasilnya yang terbaik. Oleh karena itu, di harapkan risiko kesalahannya atas perencanaannya dapat diminimalkan. Prediksi diperlukan karena sering digunakan untuk mencari informasi dari sejumlah besar data mining (Amalia, 2018).

Datanya sebuah penjualan yang tersedia dan hendak di olah ataupun dianalisa didalam menelusuri derajat cenderungnya konsumen dimasing-masing tujuannya didalam menjualkan produknya difaktor keminatan. Dari olah datanya bisa di dapatkan pola didalam mengonsumsi yang lebih jelas dari produk perusahaannya. Tersedianya data yang cukup, butuhnya informasi didalam menyokong penentuan keputusannya didalam melahirkan penyelesaian bisnisnya dan infrastruktur pendukungnya dibidang teknologinya ialah munculnya data mining yang memiliki tujuan didalam memerikan langkah yang tepat didalam penentuan ketetapan bisnisnya didalam memperluas bisnisnya (Sutrisno, Afriyudi, & Widiyanto, 2013).

Data mining ialah kegiatan pengumpulan serta penggunaan data masa lalunya didalam mencari aturannya, polanya, serta hubungannya didalam perkumpulan datanya yang besar. Hasil penambangan datanya ini dapat digunakan didalam meningkatkan penentuan keputusannya dimasa mendatang (Swatisna, 2013). Data mining di definisikan rangkaian prosesi didalam secara manual mencari penilaian tambahan mencakup pengetahuannya yang tak di ketahui dari sebuah perkumpulan datanya. Dalam data mining adanya sejumlah

cara didalam menelusuri informasinya pada sebuah perkumpulan datanya dengan menerapkan metode serta algoritma tertentu (Wajhillah & Yulianti, 2017). Informasi yang ada dalam model dapat digunakan sebagai pengambilan keputusan dan solusi pengembangan bisnis lainnya di dunia bisnis. (Wijaya & Fauzi, 2020). Salah satu proses yang menggunakan data mining adalah klasifikasi data, yaitu mengklasifikasikan beberapa record data yang sudah ada dan yang terjadi sebelumnya. Data ini disebut data pelatihan. Data ini hanya untuk keperluan klasifikasi untuk menentukan pola dari data latih. Dalam penelitian ini adalah mencari pohon keputusan. Kemudian gunakan model untuk mengklasifikasikan data baru yang tidak terklasifikasi (Hasmin & Aisa, 2019).

Toko Yanyan Boutique adalah toko yang bergerak di bidang penjualan pakaian. Dalam operasionalnya, toko sering menghadapi kendala dalam menentukan profit penjualannya. Agar toko dapat mengambil keputusan yang tepat untuk memprediksi profit penjualan pakaian dan membantunya dalam bidang penjualan, makanya penulisnya akan menerapkan data mining dengan Metode C4.5 (Decision tree) berdasar kebutuhannya terkait penjualannya. Sebab dengan menerapkan metode ini, datanya yang terkumpulkan tentunya di gunakan dan di olah sedemikian rupa hingga semua informasinya terstruktur yang ada didalam Decision tree serta di peroleh modelnya yang terbaik (Fikri & Verina, 2020). Berikut adalah data penjualan pakaian dalam 6 bulan terakhir:



Gambar 1.1 Data penjualan Toko Yanyan Boutique

Sumber: Toko Yanyan Boutique

Berdasarkan gambar dari grafik penjualan produk yang dijual oleh Toko Yanyan Boutique Tahun 2021-2022, pada bulan Oktober hingga November mengalami penurunan dalam penjualan produk yaitu sekitar 80 produk. Kemudian pada bulan Desember ke bulan Januari, terjadi penurunan penjualan produk lagi yang cukup besar yaitu 180 produk. Pada bulan Februari hingga bulan Maret 2022, terjadi kenaikan penjualan produk hingga bisa di simpulkan artinya ditahun 2022 mulai Oktober-Maret Toko Yanyan Boutique rata-rata mendapati penyusutan didalam penjualan produknya tiap bulannya ditahun 2022.

Dari hasil evaluasi yang dilakukan Toko Yanyan Boutique, hasil penjualan produk tersebut dapat dicapai karena Toko Yanyan Boutique menjual produk dengan kualitas yang baik terhadap setiap pelanggan yang ada di Kota Batam. Namun dalam penjualan produk yang dilakukan oleh Toko Yanyan Boutique pada

tahun 2022 selalu mengalami penurunan pada bulan tertentu sehingga karena permasalahan tersebut, penulis ingin memprediksi profit penjualan pada Toko Yanyan Boutique dari segi pelayanan dan evaluasi produk yang selama ini dirasakan oleh pelanggan.

Sebuah pohon keputusan dapat diimplementasikan sebagai struktur percabangan di mana tiap cabangnya memperoleh node internalnya yang menggambarkan atributnya, tiap cabangnya yang menggambarkan keluaran dari atribut yang diselidiki dibuat, serta keputusannya untuk mendefinisikannya dibuat berdasarkan daun keputusan class. Pohon keputusan terstruktur bekerja dari akar ke daunnya, apabila diperiksa dengan data pengujian, misalnya saja datanya tanpa class tidak diketahui, tentunya pohon keputusannya membuat grafik dari akar ke simpul dan tiap penilaian atributnya sejalan dengan data pengujian di periksa apakah sejalan dengan peraturan modelnya pohon keputusan. Algoritma C4.5 serta pohon keputusannya ialah dua modelnya yang tak bisa dipisahkan, dikarenakan didalam menumbuhkan pohon keputusannya diperlukan algoritma C4.5 (Hasmin & Aisa, 2019).

Berdasar latar belakangnya di atas yang sudah teruraikan, makanya permasalahan yang tersedia yakni belum diterapkannya *data mining* didalam mengolah data penjualannya terutama pada penjualan pakaian, makanya diusulkan penelitian yang berjudul **“Penerapan Algoritma C4.5 untuk Memprediksi Profit Penjualan”**.

1.2 Identifikasi Masalah

Berdasar latar belakangnya yang sudah teruraikan makanya bisa di simpulkan sejumlah permasalahannya berikut:

1. Kesulitan toko dalam hal menentukan pakaian yang paling banyak diminati oleh *customer*.
2. Adanya keluhan beberapa *customer* pada pakaian yang dibeli tidak sesuai dengan pesanan para *customer*.
3. Tidak adanya pemanfaatan data pada toko Yanyan Boutique mengenai penjualan dan *customer* yang membeli.

1.3 Batasan Masalah

Untuk membuat pembahasannya lebih terfokus, penelitiannya ini difokuskan:

1. Penelitian ini dilakukan untuk membahas tentang prediksi profit penjualan produk pakaian pada Toko Yanyan Boutique.
2. Datanya yang di ambil hanyalah data penjualannya produk pakaian terhitung daribulan April 2021 sampai dengan bulan September 2021.
3. Data dalam penelitiannya ini diolah memakai metode algoritma C4.5 menggunakan software weka dan perhitungan manual untuk pembandingan menggunakan Microsoft excel.

1.4 Rumusan Masalah

Berdasar sejumlah identifikasi permasalahan yang ada, permasalahannya yang hendak di bahas ialah:

1. Bagaimana melakukan prediksi profit penjualan produk pakaian pada Toko Yanyan Boutique dengan menggunakan algoritma C4.5?
2. Apakah metode pengklasifikasian memakai algoritma C4.5 bisa menetapkan pola pohon keputusan yang dapat diperoleh saat memprediksi keuntungan penjualan pakaian pada toko Yanyan Boutique?
3. Bagaimana menerapkan teknik pengklasifikasian algoritma C4.5 sebagai sarana yang membantu guna mendapati pola yang di gunakan untuk menetapkan keuntungan penjualan pakaian di toko Yanyan Boutique?
4. Bagaimana tindakan yang akan dilakukan dengan hasilnya yang di peroleh yang di lakukan terkait penjualan?

1.5 Tujuan Penelitian

Tujuannya yang hendak di capai yang telah diuraikan ialah:

1. Penerapan data mining terutama metode pengklasifikasian memakai algoritma C4.5 untuk memprediksikan laba penjualannya pakaian di toko Yanyan Boutique.
2. Pengujian menggunakan metode pengklasifikasian serta algoritma C4.5 dari penerapan data mining yang di gunakan didalam memprediksi profit penjualannya pakaian di Toko Yanyan Boutique.
3. Memprediksi profit penjualan pakaian pada toko Yanyan Boutique dengan melakukan algoritma C4.5 memakai *software* data mining WEKA 3.9.5.

4. Menganalisa dan menguji penerapan data mining teknik pengklasifikasian dengan memakai algoritma C4.5 didalam memprediksikan profit penjualannya pakaian di toko Yanyan Boutique.

1.6 Manfaat Penelitian

Manfaat penelitiannya ini didalam melangsungkan penelitiannya ialah:

1.6.1 Manfaat Teoritis

Berikut manfaat teoritisnya ialah:

1. Agar dapat dijadikan sebagai informasi tambahan ataupun pengetahuannya yang terbaik bagi pembaca.
2. Agar bisa di jadikan acuan bagi peneliti yang menginginkan menentukan profit penjualan dengan melakukan penelitian menggunakan metode algoritma C4.5.

1.6.2 Manfaat Praktis

Berikut manfaat praktisnya ialah:

1. Dengan menggunakan aplikasi algoritma C4.5, semoga dapat mempermudah dalam menentukan profit penjualan di toko Yanyan Boutique.
2. Sebagai wawasan atau pengetahuan tambahan toko Yanyan Boutique agar tidak terjadi kekeliruan atau hal yang merugikan toko lagi dimasa mendatang.



Universitas Putera Batam

BAB II

TINJAUAN PUSTAKA

2.1 Teori Dasar

Penulisnya memaparkan berbagai teorinya yang dijadikan sebagai acuan yang mendasar didalam melakukan penelitiannya dengan menjabarkan definisi serta istilah yang terkait dengan judul penelitiannya.

2.1.1 Teori Umum

Penulisnya menganalisis berbagai teorinya yang umum untuk melakukan penelitian yang mencakup definisi dan istilahnya yang umum.

2.1.1.1 Sistem Pendukung Keputusan (SPK)

Sebagian dari sistem informasinya terkomputerisasi seperti sistem yang basisnya pengetahuan dan manajemen pengetahuannya, dan di gunakan didalam memotivasi penentuan keputusannya sebuah perusahaan. Pada dasarnya, DSS adalah evolusi dari sistem informasinya suatu manajemen yang di komputerisasikan serta di rancang untuk saling melakukan interaksi sesama pengguna. Sistem ini mencakup sejumlah komponennya yang utama, yakni database, model base, dan dialog.

2.1.1.2 KDD (Knowledge Discovery In Database)

KDD ialah langkah didalam memperoleh informasi melalui penggunaan data dari database atau data yang disimpan. Setelah informasi ini ditemukan,

digunakan sebagai basis pengetahuan dalam penentuan keputusannya. Berikut tahapannya didalam KDD, yakni:

1. *Data Selection*

Fasenya ini berjalan diawal prosesi KDD. Artinya, dilakukan dengan mengumpulkan informasinya dengan cara pemilihan datanya yang di jadikan didalam sumber datanya yang diakhirnya diproses ditahap data mining.

2. *Pre-processing atau Cleaning*

Tahapan ini di lakukan didalam membuang dan memvalidasi sejumlah duplikasinya pada datanya yang tidak memenuhi, sebab prosesnya didalam entry datanya haruslah mencukupi keperluan penelitiannya.

3. *Transformation*

Langkah didalam mengubah serta mengadaptasi model menyimpan database.

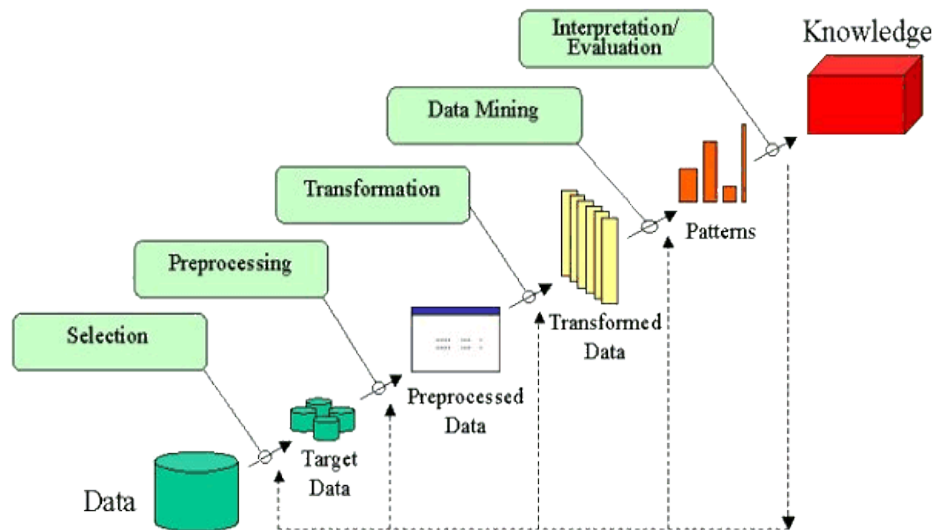
4. *Data Mining*

Suatu langkah didalam menyediakan caranya untuk menemukan serta menyerupai pola peraturan didalam memperoleh informasinya didalam keputusannya berdasar tujuan penelitiannya.

5. *Interpretation*

Hal ini membuat hasilnya sebuah model peraturan data mining mudah di pahami. Secara khusus, ini menampilkan informasi yang bertentangan dengan hipotesis penelitian.

Langkah didalam KDD yakni:



Gambar 2.1 Proses KDD

2.1.1.3 Data Mining

Langkah didalam menemukan informasinya yang sebelumnya tak di ketahui dari kumpulan datanya yang besar. Data mining ialah rangkaian prosesi yang menggunakan 1 ataupun lebih tekniknya didalam mempelajari komputer didalam menganalisa serta mengekstrakkan pengetahuannya dengan otomatis, ataupun rangkaian prosesi yang menemukan penilaian dari sekumpulan datanya sebagai pengetahuannya. Hanya mereka yang tak di definisikan dengan manual. Data mining di gunakan menggambarkan menemukan pengetahuannya didalam database. Data mining ialah prosesi mengekstraksikan serta mengidentifikasi informasi dan pengetahuannya yang bermanfaat dengan menerapkan teknik statistika, matematika, kecerdasan buatanya, serta mempelajari mesin. Dari database yang besar. Di lihat dari definisinya mining artinya mengekstraksi dari data stack. Penumpukan datanya di sini ialah datanya yang sudah di kumpulkan sebelumnya. Tujuannya dari data mining ialah didalam mendapatkan sekumpulan

informasinya ataupun pengetahuannya yang dapat disimpan selama data mining serta digunakan oleh komunitas dan organisasi itu sendiri (Yulia & Azwanti, 2018).

2.1.1.4 Pengelompokkan Data Mining

Apabila pembagiannya di dasarkan berdasar tugasnya serta langkahnya yang memiliki basisnya didalam prosesnya yang di lakukan didalam data mining, bisa di bagi menjadi sejumlah tahapan berikut:

1. Deskripsi atau *Description*

Fase ini dimulai dengan seorang analisanya yang mencoba didalam melakukan klasifikasi atas jenisnya yang terkandung didalam datanya serta sumber informasinya. Hal ini bisa di jelaskan apabila staff yang bertanggung jawab untuk menghitung aktivitas didalam memungut suaranya didalam mengkampanyekan pemilu tak diketahui ataupun memperoleh informasinya yang lebih jelas terkait keterampilannya serta pengalaman para kandidatnya yang terdaftar dalam pemilu. Oleh sebab itu, penjabaran permasalahannya ialah pengupayaan didalam memerikan penjabatannya terkait kemampuannya pihak yang ada didalam sistem pemilihannya.

2. Estimasi atau *Estimation*

Fase ini aktivitasnya yang di lakukan dari estimasi berfungsi sama dengan pengklasifikasiannya, namun berbeda jika proses estimasi atau variabel objektif yang terdapat pada fase ini mengandung angka. nilai. data. Pola yang dibuat bersumber dari kumpulan datanya yang berisikan angka

ataupun variabelnya yang ada di dalamnya yang di gunakan sebagai bahan penetapan nilainya didalam proses memprediksikan. Lalu dilanjutkan dengan memberikan estimasinya yang diturunkan dari nilainya yang ada divariabel yang diturunkan dari nilai prediksinya. Hal ini bisa di jelaskan dengan nilainya yang termasuk didalam IPK ataupun nilai rata-rata (S.1) mahasiswa yang sedang mengikuti program pendidikan di universitas tersebut. Estimasi dapat ditampilkan dengan menggunakan nilai IPK atau Indeks Prestasi Mahasiswa selama proses pengajaran setiap semester.

3. Prediksi atau *Prediction*

Proses atau fase peramalan juga hampir mirip dengan proses dan fase klasifikasi dan juga selama proses estimasi, tetapi tidak termasuk apakah ramalan nilai yang ada ditujukan untuk masa depan.

4. Klasifikasi atau *Classification*

Aktivitas yang dikategorikan menggunakan beberapa target untuk variabel berdasarkan kategori yang ada. Inilah ketika mengklasifikasikan pendapatan karyawan suatu perusahaan, pendapatan karyawan dibagi menjadi tiga kategori: karyawan berpenghasilan sangat tinggi, karyawan berpenghasilan menengah atau berpenghasilan normal, sangat rendah.

5. Pengklusteran atau *Clustering*

Langkah yang didalamnya ada pengelompokkan ialah kegiatannya yang melakukan pengelompokkan hasil pengumpulan datanya, serta datanya di kelompokkan kembali berdasar datanya yang menunjukkan kesamaan ataupun pola kesamaan. Kelompok yang bersangkutan memiliki hasil

pencatatan pada satu kelompok yang memiliki pola yang sama atau serupa, dan kelompok yang lain tidak memiliki pola yang sama atau serupa pada hasil pencatatan yang ada. Satu atau lebih proses dalam sebuah cluster berbeda dari satu atau lebih proses dalam klasifikasi karena cluster tak mempunyai target. Algoritma komputasi yang di gunakan didalam clustering berfokus pada mengelompokkan datanya yang tersedia berdasar kesamaannya yang berfokus model, sehingga pekerjaan yang dilakukan pada tahap dan proses clustering adalah estimasi, prediksi, dan tidak melakukan proses yang berhubungan dengan klasifikasi dan untuk evaluasi. Kemiripan atau kemiripan data yang disimpan dalam kelompok mencapai maksimum atau sebaliknya. Jika kemiripan ataupun kesamaan datanya yang tersedia sangatlah rendah, makanya hasil dari penilaian ini cukup kecil.

6. Asosiasi atau *Association*

Asosiasi tersebut melalui proses memperoleh informasi lengkap atau pendukung yang terkandung dalam suatu item data.

2.1.1.5 Pohon Keputusan (Decission Tree)

Pohon keputusan menerapkan representasi pohon dimana tiap node mewakili atributnya, cabang mewakili nilai atributnya, serta daunnya mewakili kelasnya. Node teratas pada pohon keputusannya di sebut root. Pohon keputusannya ialah metode kategorisasi yang umumnya di gunakan. Selain cukup cepat didalam penyetulan, hasilnya dari modelnya yang di bangun juga mudah di pahami.

Data pohon keputusannya umumnya di representasikan berbentuk tabel atributnya dan catatannya. Atributnya mewakili parameternya yang di sebut karakteristik bangunan pohon. Keuntungan utamanya menerapkan pohon keputusan ialah anda dapat memecah proses keputusan yang awalnya rumit menjadi lebih ringan. Hal ini memungkinkan penentuan keputusannya menginterpretasikan solusinya untuk masalah tersebut. Jika Anda memutuskan untuk bermain tenis, kriterianya yang perlu dipertimbangkan ialah cuacanya, anginnya, dan suhunya. Sebuah atribut disebut atribut hasil, yang mewakili data solusi untuk setiap item data.

Ada 3 jenis node dalam pohon keputusannya, yakni:

1. *Root node* ialah node teratas, tak adanya penginputan pada nodenya serta mungkin tak mempunyai outputnya ataupun mempunyai lebih dari satu outputnya.
2. *Internal Node* ialah node cabangnya, dengan nodenya ini hanyalah mempunyai 1 inputnya dan mempunyai setidaknya 2 output.
3. *Leaf Node* ataupun Terminal Node ialah nodenya yang terakhir, hanyalah ada 1 input serta tak adanya outputnya di node ini. (Amalia, 2018).

Selanjutnya berikut penjelasan mengenai kelebihan dan kekurangannya dari pohon keputusannya, yakni:

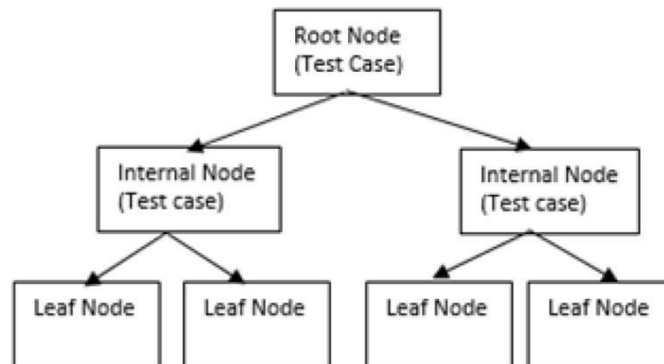
1. Kelebihan
 - a. Dalam penggunaan pohon keputusannya, ketika ada cakupan data kompleks yang telah di modifikasi supaya lebih memudahkan disaat di gunakan nantinya dianggap keputusan kebijakannya.

- b. Bila membuang metodenya didalam perhitungannya yang tak diperlukan, sehingga hanyalah berkriteria yang memenuhi kebutuhannya didalam pengujian yang digunakan dalam proses pengujian.
- c. Jika anda memilih integritas atau deskripsi data saat menggunakan pohon keputusan, hasilnya dari pemilihannya terintegritas ataupun keakurasian datanya ialah perbedaan peringkat datanya berdasar kriterianya yang lain.
- d. Membantu menghindarkan permasalahan lainnya dengan memakai beberapa keterbatasan jenis kelompoknya ataupun kategorinya tanpa melakukan pengorbanan kualitasnya dari hasil keputusannya yang terakhir.

2. Kekurangan

- a. Disaat memakai pohon keputusannya, sejumlah besar hasil keputusannya untuk dihasilkan serta kriterianya yang di gunakan akan meningkatkan waktunya yang di butuhkan memeroleh keputusannya serta perlu sistemnya dengan dukungan memorinya yang tinggi.
- b. Pohon keputusannya mempunyai jumlah sekumpulan kesalahannya yang ada ditiap bagian strukturnya yang ada dipohon keputusannya.
- c. Sulit didalam mendirikan struktur pohon keputusannya yang memerikan hasilnya yang terbaik.

- d. Disaat pohon keputusannya digunakan, hasilnya ataupun kualitasnya yang di hasilkan sangatlah di tentukan oleh modelnya ataupun strukturnya yang diimplementasikan.



Gambar 2.2 Model Pohon Keputusan

2.1.1.6 Algoritma C4.5

Algoritma C4.5 membantu didalam mengklasifikasikan datanya ke dalam angkanya ataupun kategorinya. Sesudah tahapan prosesnya pengklasifikasiannya selesai, makanya akan dihasilkan aturan pola yang berbeda yang bisa di gunakan didalam melangsungkan tahapan prosesnya didalam memprediksikan didalam memakai nilai prediksinya yang di peroleh dengan merekam datanya yang lama ke datanya yang terbaru.

Algoritma C4.5 di gunakan didalam mendirikan pohon keputusannya. Sejumlah pengembangannya yang di lakukan oleh di C4.5 termasuk melebihi nilai yang hilang, melebihi data yang berkelanjutan, dan pemangkasan. (Tukino, 2019). Menurut (Hendra & Harman, 2020) Ada sejumlah langkah didalam membangun pohon keputusannya pada algoritma C4.5:

1. Persiapkan data pelatihannya. Datanya latih umumnya di ambil dari data historisnya yang sudah dibangkitkan sebelum nya ataupun data historis yang bisa mereferensikan data sebelum nya serta dikategorikan kedalam kelas tertentu.
2. Perhitungan akar pohonnya. Akar diekstraksi dari atributnya yang di pilih memakai metode penghitungan perolehan *gain* untuk tiap atributnya dengan perolehan *gain* yang paling tinggi. Dengan kata lain, ini adalah akar pertama dari pohon keputusan. Langkah pertama adalah menghitung perolehan *entropy* sebelum menghitung perolehan *gain*.

Perhitungan I untuk menghitung *entropy*:

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

Rumus 2.1 Perhitungan Entropy

Keterangan:

1. Simbol *S* ialah himpunan
2. Simbol *A* ialah keterangannya yang di gunakan.
3. Simbol *n* ialah banyaknya partisi yang ada diketerangan bagian *S* *p_i* ialah proporsinya dari *S_i* kepada *S*.
4. Dalam perhiutungan nilai *gain* menggunakan persamaan II:

$$\mathbf{Gain (A) = Entropi (S) - \sum_i^n \frac{|S_i|}{|S|} x Entropi (S_i)}$$

Rumus 2.2 Perhitungan Gain

Keterangan:

1. Simbol S ialah himpunan
2. Simbol A ialah keterangan yang di gunakan.
3. Simbol n ialah jumlah partisi yang ada diketerangan bagian A .
4. Simbol $|S_i|$ ialah banyaknya dari kasusnya yang ada didalam partisinya yang ke- i .
5. Simbol $|S|$ ialah banyaknya dari kasusnya yang ada didalam S .

2.1.2 Teori Khusus

Penulisnya menjabarkan sejumlah teori khususnya untuk melakukan penelitiannya mencakup istilah serta definisi tertentu.

2.1.2.1 Prediksi

Prediksi ialah aktivitas memprakirakan ataupun memprediksikan hal yang hendak terjadi dimasa mendatang didalam waktunya yang singkat. Prognosis ialah keadaan yang kemungkinan terjadi di masa mendatang.

1. Jenis Prediksi

Tergantung pada jenis prediksi itu dapat dibagi menjadi beberapa jenis.

Menurut perencanaan pengoperasiannya dimasa mendatang, ramalan terbagi kedalam 3 jenis, yakni:

- a. Prediksi ekonomi, menggambarkan siklusnya sebuah perusahaan dengan memprediksikan tingkatan inflasinya, adanya dana, kebutuhan dananya untuk membangun rumah, dan lainnya.

- b. Prediksi teknologi, membahas majunya teknologi yang bisa membawa dan menarik produknya yang baru ke pasar yang memerlukan pabriknya serta peralatannya yang baru.
- c. Prediksi permintaan, ialah prakiraan permintaannya didalam layanannya ataupun produknya perusahaannya.

2. Metode Prediksi

Analisi *Time Series* ialah hubungannya diantara variabelnya yang Anda cari dengan variabelnya yang bisa mereka pengaruhi, terkait dengan waktunya misalnya mingguan, bulanan, triwulanan, triwulanan, setengah tahunan atau tahunan.

Didalam analisa *time series*, variabel yang di cari ialah waktu.

Metode prediksinya mencakup:

- a. Metode *Smoothing* adalah beberapa tipe ramalannya yang berjangka pendek, mencakup rencana keuangannya, rencana persediaannya. Tujuan menerapkan metodenya ini ialah demi mendeteksikan ketidak teraturan didata masa lalunya, misalnya musiman.
- b. Metode *Box Jenkins* ialah pengaturan waktunya yang digunakan untuk prediksi jangka pendek dengan menggunakan model matematika.
- c. Metode proyeksi tren regresi ialah metodenya yang di gunakan didalam jangkanya yang pendek maupun panjang. Metodenya

ini sebagai rumusan garis trend didalam persamaannya yang matematis.

2.1.2.2 Penjualan

Definisi penjualan berarti mikro berupa organisasinya yang memprakirakan kebutuhannya para pelanggan serta mengatur arus barangnya dan jasanya didalam mencukupi kebutuhan produsennya serta konsumennya sedangkan mikro bersifat sosial, yaitu mengatur kegiatan untuk mencapai tujuan kita. Proses membimbing arus barangnya dan jasanya dalam perekonomian dari produsennya ke konsumennya, seefektif mungkin didalam mencocokkan penawarannya serta permintaannya, dan meraih tujuan sosialnya.

Penjualan ialah aktivitas yang memiliki tujuannya didalam memenuhi kebutuhannya dan keinginannya pelanggan dengan cara menukarkan produk dan prospek di dalam perusahaan.

2.1.2.3 Kualitas Produk

Menurut (Umboh, Tumbel, & Soepeno, 2015) pengertian kualitas produk perlu dijadikan perhatiannya produsen atau perusahaannya, sebab kualitasnya dari produk yang berkaitan dengan kepuasan konsumennya yang dijadikan tujuannya dari aktivitas pemasarannya. Kotler (2005: 49) menjabarkan kualitas produk sebagai karakteristik dan ciri khusus umum dari sebuah produknya ataupun jasanya didalam berkemampuan didalam memuaskan kebutuhannya yang dinyatakan / tersirat.

2.2 Penelitian Terdahulu

Penelitinya juga telah menemukan sejumlah temuan sebelumnya berkaitan analisis menggunakan algoritma C4.5 yang disimpulkan yaitu:

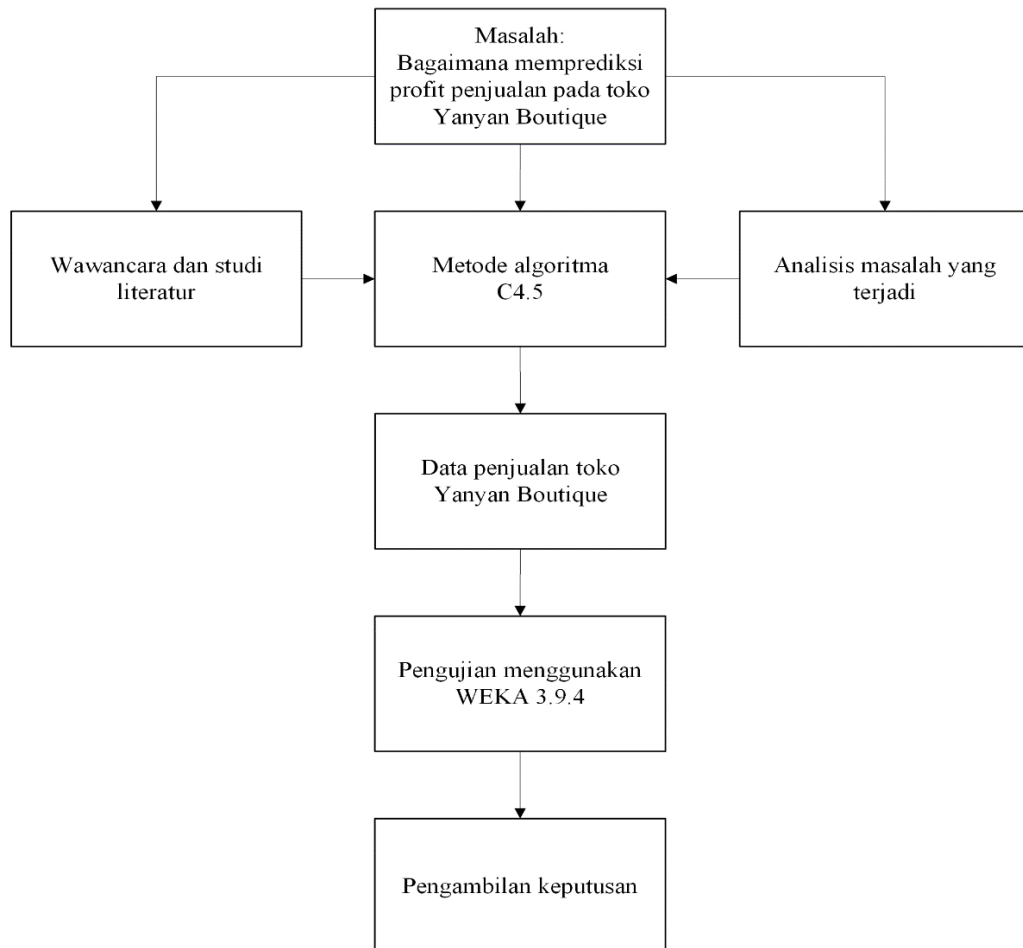
1. Dengan judul “Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa” yang di lakukan oleh (Hasmin & Aisa, 2019) penelitian ini menunjukkan bahwa mengidentifikasi memakai teknik data mining terutama algoritma C4.5 dapat sangat membantu didalam pengambilan keputusan.
2. Dengan judul “Penerapan Algoritma C4.5 Untuk Memprediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam” yang dilakukan oleh (Yulia & Azwanti, 2018) didalam penelitiannya ini, hasil perhitungan dengan memakai algoritma C4.5 memperhatikan lima atribut yang digunakan yaitu pendidikan, jumlah keluarga, dan nilai *gain* tertinggi dari semua produk elektronik menunjukkan bahwa konsumsi dapat diprediksi dari jumlah pengguna, lama tinggal di rumah, luas bangunan rumah.
3. Dengan judul “Penerapan Data Mininh Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT. Murni Indah Sentosa” yang dilakukan oleh (Fikri & Verina, 2020) penelitian ini menunjukkan bahwa algoritma *data mining* C4.5 ini, PT. Murti Indah Sentosa digunakan untuk memberikan kemudahan dalam mengolah data penjualan alat kesehatan dengan cara cepat mendokumentasikan data penjualan dan menyimpannya dalam database.

4. Dengan judul “Data Mining Algoritma C4.5 Untuk Memprediksi Penjualan Baterai Di PT Varta Microbattery Indonesia” yang dilakukan oleh (Wijaya & Fauzi, 2020) penelitian ini menunjukkan bahwa Pengujiannya dengan memakai software RapidMiner begitu efektif di karenakan pohon keputusannya yang di dapat hasilnya serupa dengan perhitungannya secara manual. Oleh sebab itu memakai RapidMiner begitu membantu didalam pengujiannya yang di lakukan.
5. Dengan judul “Penerapan Data Mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada PT Remicon Widyaprima” yang di lakukan oleh (Hendra & Harman, 2020) penelitiannya ini menggunakan perhitungan pengklasifikasian untuk referensi data transaksi penjualan PT Remicon Widyaprima atau sebagai panduan untuk membangun pohon keputusan disaat memakai teknik *data mining* dengan algoritma C4.5 didalam membuat prakiraan penjualan ready mix.
6. Dengan judul “Penerapan Algoritma C4.5 didalam Memprediksikan Keuntungan Pada PT SMOE Indonesia” yang dilakukan oleh (Tukino, 2019) penelitiannya ini menggunakan algoritma C4.5 sebagai metode klasifikasi dan mempertimbangkan perolehan *gain* yang paling tinggi, hingga menguntungkan PTSMOE Indonesia.
7. Dengan judul “Analisis Kualitas Produk, Brand Image Dan Life Style Terhadap Keputusan Pembelian Pakaian Wanita Di Mississippi Manado Town Square” yang di lakukan oleh (Umboh, Tumbel, & Soepeno,

- 2015) penelitian ini menunjukkan bahwa temuan tersebut dapat digunakan untuk mengembangkan ilmu manajemen pemasaran.
8. Dengan judul “Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa” yang di lakukan oleh (Swatisna, 2013) penelitiannya ini menunjukkan Algoritma Decision Tree C4.5 akurat di terapkan didalam menentukan kesesuaiannya jurusan mahasiswa.
 9. Dengan judul “Analisis Kepuasan Pelanggan Dengan Algoritma C4.5 Pada PT Wook Global Technology” yang dilakukan oleh (Amalia, 2018) penelitian ini menunjukkan bahwa penelitian ini dilakukan pemodelan memakai datanya yang di olah berdasar tahapan KDD.
 10. Dengan judul “Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus PT. Indomarco Palembang” yang dilakukan oleh (Sutrisno, Afriyudi, & Widiyanto, 2013) penelitian ini menunjukkan bahwa penerapan *data mining* memakai aplikasinya yang di bangun bisa membantu PT. Indomarco sebagai contoh proses penentuan keputusannya suatu perusahaannya untuk memperoleh sampel penjualan produknya.

2.3 Kerangka Pemikiran

Didalam kerangka pemikirannya ini ialah langkah yang di lakukan didalam menyelesaikan permasalahan penelitiannya yang di buat, yakni:



Gambar 2.3 Kerangka Pemikiran

Sumber: (Penulis,2022)



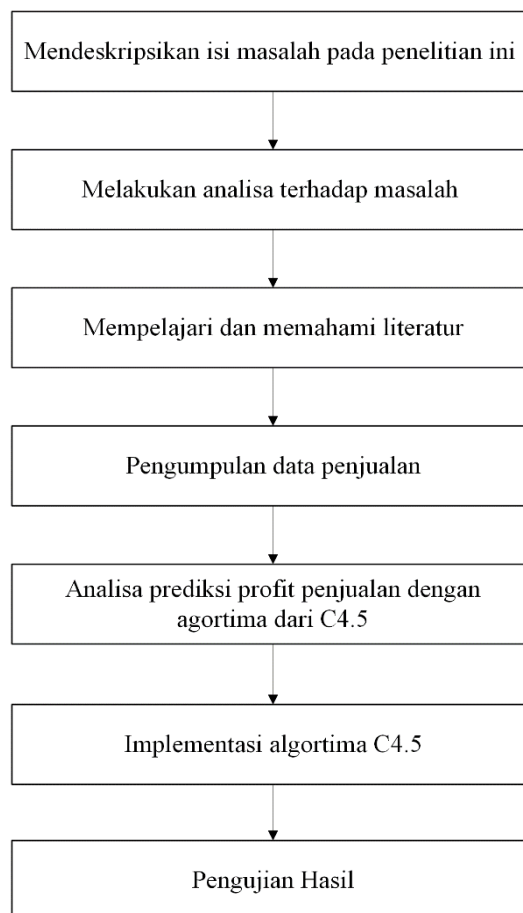
Universitas Putera Batam

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Pada desain penelitiannya ini merupakan prediksi profit penjualan toko Yanyan Boutique yang berlokasi di kota Batam. Gambar desain penelitiannya ini berikut:



Gambar 3.1 Desain Penelitian

Sumber: (Penulis, 2022)

Pada desain penelitiannya ini merupakan prediksi profit penjualan toko Yanyan Boutique yang berlokasi di kota Batam. Desain penelitian ini sebagai berikut:

1. Melakukan pendeskripsian isi permasalahannya dipenelitian ini

Didalam mendeskripsikan permasalahan, peneliti berusaha keras memerikan penjabarannya terkait masalah yang hendak di teliti, terkait memprediksikan profit penjualannya memakai algoritma C4.5.

2. Menganalisa permasalahannya

Tahapannya ini ialah kegiatan yang dilakukan didalam mendeskripsikan permasalahannya. Pada tahapan ini peneliti menganalisa masalah yang telah dideskripsikan untuk memahami masalah yang telah diidentifikasi.

3. Mengertikan literatur

Penulisnya berusaha didalam meneliti serta mengerti berbagai sumber teorinya yang membantu didalam meneliti jurnalnya hingga bisa di jadikan pedomannya bagi didalam meneliti.

4. Pengumpulan data penjualan

Mengumpulkan datanya dengan melangsungkan observasinya dan mewawancarai dilokasi yang hendak di amati. Peneliti menyelesaikan kegiatannya terkait prediksi penjualan pada Toko Yanyan Boutique.

5. Analisa prediksi profit penjualan dengan algoritma dari C4.5

Datanya yang di peroleh dari penelitiannya yang selanjutnya dianalisis serta diolah memakai algoritma C4.5 dari prosesi terbentuknya *decision tree* untuk menghasilkan datanya yang di inginkan peneliti.

6. Implementasi algoritma C4.5

Pada tahapan ini peneliti akan menetapkan atributnya sebagai akar pohon keputusannya serta menghitung keuntungannya yang hendak di gunakan didalam membentuk struktural pohon keputusannya.

7. Pengujian Hasilnya

Ditahap terakhir ini, peneliti hendak mengujikan serta hasil perancangannya menerapkan software **WEKA 3.9.5**.

3.2 Objek Penelitian

Pada subbab objek penelitiannya, peneliti akan menjabarkan lokasi penelitiannya dan penjelasannya jadwal sepanjang aktivitas penelitiannya berikut.

3.2.1 Lokasi Penelitian

Penelitiannya ini berlokasi di Toko Yanyan Boutique yang berlokasi di kios Pasir Putih blok A no 20, Batam Center, Kota Batam, Kepulauan Riau.

3.2.2 Jadwal Penelitian

No	Kegiatan	Feb-22	Mar-22	Apr-22	Mei-22	Juni-22	Juli-22
1	Menentukan Judul						
2	Bimbingan						
3	Identifikasi Permasalahan						
4	Studi Literatur						
5	Mengumpulkan Data						
6	Mengolah Data						

7	Mengujikan Hasil			
8	Menyusun Laporan			

Tabel 3.1 Jadwal Penelitian

Sumber: (Penulis,2022)

3.3 Populasi dan Sampel

Peneliti hendak menjabarkan sejumlah keterangannya yang ada dari populasinya dan sampelnya yang di jadikan bahan didalam melaksanakan penelitiannya berikut.

3.3.1 Populasi

Populasi penelitiannya ini ialah datanya dari hasil penjualannya yang terbanyak di minati konsumennya di Toko Yanyan Boutique, datanya yang hendak di ambil yakni pada bulan Maret - Mei 2022.

3.3.2 Sampel

Datanya yang akan di jadikan sampel didalam penelitiannya ini, penelitiannya hendak memperoleh data penjualannya secara random hingga 50 data sampelnya dimana seluruh datanya di ambil dari semua datanya dipopulasi pada bulan Maret - Mei 2022.

3.4 Variabel Penelitian

Dalam penelitiannya ini di gunakan variabel penjualannya pakaian di Toko Yanyan Boutique, dimana data penjualannya pakaian ialah penjumlahan dari sekian banyaknya kelas yang terbanyak di minati oleh konsumennya.

Datanya yang di gunakan didalam penelitiannya, terutama terkait variabel prediksi penjualannya, ialah:

1. Ketersediaan
2. Kualitas
3. Harga
4. Pengiriman
5. Promosi

3.5 Teknik Pengumpulan Data

Mengumpulkan datanya di lakukan didalam memperoleh informasinya yang di perlukan didalam meraih tujuan penelitiannya, ataupun didalam mengungkapkan informasinya ataupun keadaan lokasi penelitiannya sejalan dengan kebutuhan penelitiannya. Teknik didalam mengumpulkan datanya ialah:

1. Studi literatur

Pendeskripsian yang singkat tentang topiknya yang sudah di pelajari, dibahas, serta diidentifikasi, umumnya di susun berdasar waktunya ataupun topiknya. Penelitinya memelajari sejumlah jurnal yang berhubungan terkait permasalahan yang dimaksud.

2. Observasi

Observasinya secara langsung peneliti lakukan di lapangan, penelitinya langsung mendatangi Toko Yanyan Boutique untuk melihat transaksi penjualan di Toko Yanyan Boutique.

3. Dokumentasi

Peneliti mengkaji informasi tertulis dilokasi penelitiannya yang berkenaan dengan permasalahan yang sudah ada ataupun melalui serangkaian proses menggali datanya lalu dikelompokkan sejalan dengan kebutuhan penelitiannya yang hendak di lakukan. Peneliti melaukan pengamatan serta pengkajiannya data penjualan di Toko Yanyan Boutique.

4. Wawancara

Aktivitas bertanya dan menjawab didalam penelitiannya yang melangsungkan penelitiannya secara lisan dimana 2 pihak ataupun lebih mendengarkan langsung pernyataannya. Peneliti mengajukan sejumlah pertanyaan pada pemilik dan pekerja Toko Yanyan Boutique.

3.6 Metode Penelitian

Didalam metode penelitiannya ini, memakai teknik *data mining* dan teknik klasifikasi dari algoritma C4.5. Berikutnya, klasifikasi berdasarkan kategori dilakukan pada tahap awal penelitiannya ini. Hasilnya ditahap awal diperhitungkan dengan algoritma C4.5 demi memperoleh *entropy* dan *gain* sehingga membentuk pohon keputusan (*Decision Tree*). Prosesnya ini diulangi sampai *node* di pohon keputusannya tak bisa di turunkan lagi, sampai pola aturannya yang mendasar untuk mengevaluasi prediksi penjualannya di Toko Yanyan Boutique.



Universitas Putera Batam