

**ANALISIS ALGORITMA C4.5 UNTUK
MEMPREDIKSI PENJUALAN PADA *DEALER*
MOTOR HONDA**

SKRIPSI



**Oleh:
Nolaskus Sinaga
150210162**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2020**

**ANALISIS ALGORITMA C4.5 UNTUK
MEMPREDIKSI PENJUALAN PADA *DEALER*
MOTOR HONDA**

SKRIPSI

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



**Oleh:
Nolaskus Sinaga
150210162**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2020**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini saya:

Nama : Nolaskus Sinaga
Npm : 150210162
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa Skripsi yang saya buat dengan judul:

ANALISIS ALGORITMA C4.5 UNTUK MEMPREDIKSI PENJUALAN PADA DEALER MOTOR HONDA 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 tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka. Apabila ternyata di dalam naskah skripsi ini digugurkan dan skripsi 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, 21 Februari 2020
Yang membuat pernyataan,

Nolaskus Sinaga
150210162

**ANALISIS ALGORITMA C4.5 UNTUK
MEMPREDIKSI PENJUALAN PADA *DEALER*
MOTOR HONDA**

SKRIPSI

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

**Oleh :
Nolaskus Sinaga
150210162**

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

Batam, 21 februari 2020

**Sunarsan Sitohang, S.Kom., M.TI.
Pembimbing**

ABSTRAK

Motor menjadi pilihan setiap orang sebagai andalan *transportasi* karena dari harga yang terjangkau, perawatan yang mudah serta dapat digunakan dalam jangkah waktu yang cukup lama. Di Indonesia sendiri, motor yang paling laris dipasaran adalah motor merk Honda. Untuk wilayah kota batam ada banyak perusahaan yang menjual motor Honda salah satunya adalah PT. Capella Dinamik Nusantara Cabang Muka Kuning. Wilayah penjualan sangat mempengaruhi terhadap penjualan motor selain ke showroom langsung, beberapa tempat ramai seperti mall dapat digunakan sebagai promosi penjualan motor karna tempat yang strategis dan mudah dijangkau masyarakat. Data penjualan yang banyak akan sangat berpengaruh terhadap pengambilan keputusan serta sangat sulit untuk dianalisa. Dari data penjualan yang begitu banyak, maka dilakukan metode *data mining* dengan menggunakan algoritma C4.5. algoritma C4.5 adalah salah satu *data mining* yang digunakan untuk melakukan klasifikasi dan segmentasi yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar. Dari penelitian yang dilakukan terdapat 3 root (akar) dari 5 variabel yang digunakan sebagai sampel yaitu segmen, harga, tahun produksi semester dan penjualan via. Hasil dari kegiatan *data mining* ini diharapkan mampu memberikan sebuah pohon keputusan untuk membantu pihak PT. Capella Dinamik Nusantara Cabang Muka Kuning dalam melihat pola prediksi perilaku konsumen dalam membeli motor.

Kata Kunci : Algoritma C4.5, *Data mining*, Klasifikasi, Pohon keputusan.

ABSTRACT

Motorbike is the choice of everyone as a mainstay of transportation because of the affordable price, easy maintenance and can be used in a long time. In Indonesia alone, the best-selling motorbike in the market is the Honda motorcycle brand. For the city of Batam there are many companies that sell Honda motorcycles, one of which is PT. Capella Dinamik Nusantara Yellow Front Branch. The sales area is very influential on motorcycle sales other than the direct showroom, some crowded places like malls can be used as motorcycle sales promotions because of the strategic location and easy to reach by the public. A lot of sales data will greatly affect the decision making and very difficult to analyze. From so many sales data, the data mining method is performed using the C4.5 algorithm. C4.5 algorithm is one of the data mining used to do classification and segmentation which aims to find valuable patterns from relatively large to very large data. From the research conducted there are 3 root (roots) of 5 variables used as samples, namely segment, price, semester production year and sales via. The results of data mining activities are expected to be able to provide a decision tree to help PT. Capella Dinamik Nusantara, Branch of Muka Kuning in looking at predictive patterns of consumer behavior in buying a motorcycle.

Keywords: *C4.5 Algorithm, Classification, Data mining, Decision Tree.*

KATA PENGANTAR

Puji dan syukur kepada tuhan yang maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika 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 terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Ketua Program Studi bapak Andi Maslan, S.T., M.SI.
3. Sunarsan Sitohang, S.Kom., M.TI. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
4. Dosen dan Staf Universitas Putera Batam.
5. Orang tua penulis yang selalu mendoakan dan memberi semangat.
6. Fatul fikri, Oda, irwansyah, Sopar sebagai sahabat yang selalu memberi support.
7. Teman-teman seangkatan yang berjuang bersama dan saling memberi semangat dan motivasi.
8. Serta berbagai pihak yang telah membantu dalam menyelesaikan skripsi ini secara langsung maupun tidak langsung.

Batam, Februari 2020

Penulis

DAFTAR ISI

	HALAMAN
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	3
1.3. Batasan Masalah	4
1.4. Perumusan Masalah	4
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	5
1.6.1. Manfaat Teoritis	5
1.6.1. Manfaat Praktis.....	
BAB II KAJIAN PUSTAKA	
2.1. <i>Knowledge Discovery In Database (KDD)</i>	7
2.2. <i>Data mining</i>	8
2.3. Metode Klasifikasi.....	11
2.4. <i>Software</i> Pendukung	15
2.4.1. <i>RapidMiner</i>	15
2.5. Penelitian Terdahulu	16
2.6. Kerangka Pemikiran	19
BAB III METODE PENELITIAN	
3.1. Desain Penelitian	21
3.2. Teknik Pengumpulan Data	23
3.3. Operasional Variabel	24
3.4. Metode Analisis dan Rancangan Sistem.....	26
3.5. Lokasi dan Jadwal Penelitian	28
3.5.1. Lokasi Penelitian	28
3.5.2. Jadwal Penelitian	28

BAB IV HASIL DAN PEMBAHASAN	
4.1. Analisis <i>Data mining</i> Menggunakan Algoritma C4.5	29
4.1.1. Seleksi dan Pemilihan Atribut	29
4.1.2. Melakukan Pra-Proses	30
4.2. Menghitung <i>Entropy dan Gain Node 1</i>	33
4.2.1. Menghitung <i>Entropy Total</i>	33
4.2.2. Menghitung <i>Entropy Semester</i>	34
4.2.3. Menghitung <i>Entropy Segmen</i>	36
4.2.4. Menghitung <i>Entropy Tahun Produksi</i>	37
4.2.5. Menghitung <i>Entropy Harga</i>	39
4.2.6. Menghitung <i>Entropy Penjualan Via</i>	41
4.3. Menghitung <i>Entropy dan Gain Node 1.1</i>	46
4.3.1. Menghitung <i>Entropy Total</i>	46
4.3.2. Menghitung <i>Entropy Semester</i>	47
4.3.3. Menghitung <i>Entropy Segmen</i>	48
4.3.4. Menghitung <i>Entropy Tahun Produksi</i>	49
4.3.5. Menghitung <i>Entropy Harga</i>	50
4.4. Menghitung <i>Entropy dan Gain Node 1.2</i>	54
4.4.1. Menghitung <i>Entropy Total</i>	54
4.4.2. Menghitung <i>Entropy Semester</i>	55
4.4.3. Menghitung <i>Entropy Segmen</i>	57
4.4.4. Menghitung <i>Entropy Tahun Produksi</i>	58
4.4.5. Menghitung <i>Entropy Harga</i>	60
4.5. Hasil Pengujian.....	64
BAB V SIMPULAN DAN SARAN	
5.1. Simpulan	72
5.2. Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN	
Lampiran 1. Pendukung Penelitian	
Lampiran 2. Daftar Riwayat Hidup	
Lampiran 3. Surat Keterangan Penelitian	

DAFTAR GAMBAR

	HALAMAN
Gambar 2.1 Pengelompokan Teknik Klasifikasi	12
Gambar 2.2 Kerangka Pemikiran	20
Gambar 3.1 Desain Penelitian.....	21
Gambar 3.2 Rancangan Sistem	26
Gambar 4.1 Pohon Keputusan <i>Node 1</i>	45
Gambar 4.2 Pohon Keputusan <i>Node 1.1</i>	53
Gambar 4.3 Pohon Keputusan <i>Node 1.2</i>	63
Gambar 4.4 <i>Welcome To Rapidminer Studio</i>	65
Gambar 4.5 <i>Import Data Read Excel</i>	65
Gambar 4.6 <i>Data Import Select Step 1</i>	66
Gambar 4.7 <i>Data Import Select Step 3</i>	67
Gambar 4.8 <i>Data Import Select Step 4</i>	67
Gambar 4.9 <i>Data Import Finish</i>	68
Gambar 4.10 <i>Import Operator Decesion Tree</i>	69
Gambar 4.11 Menghubungkan Operator	70
Gambar 4.12 Pohon Keputusan	70

DAFTAR TABEL

	HALAMAN
Tabel 3.1 Jadwal Penelitian	28
Tabel 4.1 Penjualan Tahun 2019	30
Tabel 4.2 Klasifikasi Semester	30
Tabel 4.3 Klasifikasi Tahun Produksi	31
Tabel 4.4 Klasifikasi Penjualan Via.....	31
Tabel 4.5 Klasifikasi Harga	32
Tabel 4.6 Format Akhir Data Penjualan.....	32
Tabel 4.7 Hasil Perhitungan <i>Node 1</i>	44
Tabel 4.8 Hasil Perhitungan <i>Node 1.1</i>	46
Tabel 4.9 Hasil Perhitungan <i>Node 1.2</i>	52
Tabel 4.10 Data Penjualan <i>Node 1.1</i>	53
Tabel 4.11 Data Penjualan <i>Node 1.2</i>	62

DAFTAR RUMUS

	HALAMAN
Rumus 2.1 <i>Entropy</i>	15
Rumus 2.2 <i>Gain</i>	15

BAB I

PENDAHULUAN

1.1 Latar Belakang

Semakin banyak produsen motor membuat persaingan bisnis semakin tajam dalam strategi pemasaran, banyak strategi pemasaran yang digunakan untuk mengenalkan produk kepada masyarakat seperti promosi, *event* berhadiah, serta menjadi donator beberapa acara. Motor menjadi salah satu pilihan banyak orang untuk dijadikan *transportasi*, selain praktis, harga motor yang terjangkau menjadi alternatif membantu menjalankan berbagai macam aktivitas dalam lingkungan masyarakat sehingga berdampak kepada pasar motor nasional meningkat. Di Indonesia sampai saat ini ada 5 merek sepeda motor yang bertarung dipasar dalam negeri, yaitu Yamaha, Honda, Suzuki, Kawasaki dan TVS, dari 5 merek tersebut hanya 3 merek yang benar-benar bersaing, yakni Honda, Yamaha dan Suzuki.

PT. Capella Dinamik Nusantara merupakan perusahaan yang bergerak dibidang penjualan sepeda motor bermerek honda dengan berbagai tipe seperti, *cub, matic dan sport*. Selain di showroom perusahaan ini memiliki banyak pameran atau *stand* sebagai promosi yang mudah dijangkau oleh masyarakat. Kendala yang sering terjadi pada saat konsumen ingin membeli motor, *stock* tipe atau warna pada dealer tersebut tidak ada maka *head officer* akan menghubungi cabang PT.Cappela Dinamik Nusantara yang lain kota batam untuk mengetahui ketersediaan unit, apabila masih tidak ada *head officer* membuat *pre-order*(PO) ke gudang KRO batam *center*, yang menjadi permasalahan saat *stock* motor diguda

ng juga tidak ada (*indent*) yang berarti dealer harus menunggu waktu lebih kurang 2 bulan agar *stocknya* tersedia.

Hal tersebut tentunya membuat konsumen kecewa dan memilih dealer lain untuk mencari sepeda motor yang di inginkan, dan yang akan dirugikan adalah perusahaan itu sendiri. Sehubungan dengan masalah tersebut maka dibutuhkan metode ilmiah untuk memprediksi penjualan motor yang laris dan tidak laris ditahun berikutnya dengan menganalisis data penjualan ditahun sebelumnya, sehingga kita dapat mengetahui seberapa banyak *stock* yang dibutuhkan.

Proses pendistribusian motor untuk penjualan tentunya tidak hanya di showroom, tetapi juga ditempat-tempat strategis yang sering dikunjungi masyarakat, sehingga membutuhkan suatu metode untuk merencanakan dan menyiapkan strategi penjualan agar kinerja penjualan selalu berada dalam posisi yang baik dan meminimalisir masalah penjualan yang ada dilapangan. Jumlah data yang sangat banyak tentunya akan mempengaruhi dalam mengambil keputusan karna sulit di analisis. Analisis diperlukan untuk mengetahui pola-pola penjualan dari data yang sudah didapat untuk memprediksi penjualan motor yang nantinya akan berguna sebagai panduan untuk pendistribusian motor dibeberapa wilayah.

Data mining sering digunakan untuk menguraikan penemuan pengetahuan baru yang bisa didapat dari sebuah gudang data. *Data mining* merupakan sebuah teknik statistik dan *machine learning* untuk mengentrik informasi yang berkaitan dengan berbagai *database* besar supaya bermanfaat dan dapat digunakan dalam mengambil keputusan. Menurut (D. W. T. Putra, 2016). *Data mining* dapat menganalisis data dari sudut pandang yang berbeda kemudian membentuknya

menjadi informasi baru untuk dapat memahami substansi hubungan antara data. Salah satu algoritma yang bisa digunakan sebagai metode untuk mengklasifikasi data yang sudah ada adalah *algoritma C4.5*.

Algoritma *C4.5* adalah algoritma yang dapat membangun sebuah pohon keputusan (*decision tree*). Menurut (Arifin & Fitriyah, 2018) Algoritma *C4.5* memiliki kelebihan mengolah data numerik dan diskret, dapat menangani nilai-nilai atribut yang terhapus, membentuk *rule* yang mudah dipahami dan diinterpretasi serta kemampuannya sebagai yang tercepat dibandingkan dengan algoritma lain. Proses memperbaharui bentuk data menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule merupakan pembentukan pohon keputusan. Dalam algoritma *C4.5* nantinya akan diambil informasi dari data-data yang pasti dalam membantu pemakai dalam mengambil keputusan dari masalah ini. (D. W. T. Putra, 2016). Berdasarkan uraian diatas peneliti tertarik untuk melakukan sebuah penelitian yang berjudul “ANALISIS ALGORITMA C4.5 UNTUK MEMPREDIKSI PENJUALAN PADA DEALER MOTOR HONDA”

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang permasalahan yang dihadapi adalah kesulitan dalam hal melakukan pengolahan dari data transaksi penjualan yang sangat berfungsi sebagai informasi penjualan untuk periode selanjutnya dan sebagai pertimbangan strategi pemasaran.

1.3 Batasan Masalah

Agar penelitian ini tidak menimbulkan kekeliruan dan untuk menghindari meluasnya permasalahan yang dibahas, maka dalam penelitian ini dilakukan pembatasan masalah. Batasan masalah pada penelitian sebagai berikut.

1. Penelitian ini hanya sebagai alat bantu bagi perusahaan untuk memprediksi penjualan sebagai acuan untuk pendistribusian *stock* motor.
2. Data yang digunakan dalam penelitian ini diambil dari PT. Capella Dinamik Nusantara Cabang Muka Kuning berupa data penjualan tahun 2019
3. Penelitian ini menggunakan aplikasi *RapidMiner* untuk mengolah data penelitian.
4. Hasil dari penelitian hanya berupa pohon keputusan tanpa membangun sebuah aplikasi.
5. Penelitian ini menggunakan 5 variabel yaitu, atribut semester, segmen, tahun produksi, harga dan penjualan via.

1.4 Perumusan Masalah

Berdasarkan latar belakang yang sudah didapat, maka dapat diambil rumusan masalah sebagai berikut :

1. Bagaimana algoritma C4.5 dapat memprediksi penjualan motor ?
2. Bagaimana implementasi algoritma C4.5 *classification* dengan menggunakan *software RapidMiner*?

1.5 Tujuan Penelitian

Berdasarkan pada perumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk memprediksi penjualan dengan menganalisis data menggunakan algoritma C4.5 metode klasifikasi.
2. Untuk mengimplementasikan algoritma C4.5 dalam mengolah data dengan software *RapidMiner*.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat membawa manfaat bagi yang membaca, adapun sebagai berikut :

1.6.1 Aspek Teoritis

Hasil penelitian yang dilakukan diharapkan dapat memberikan manfaat dalam aspek teoritis seperti berikut :

1. Meningkatkan wawasan dan pengetahuan mengenai *data mining* algoritma C4.5.
2. Menjadi sumber referensi bagi penelitian selanjutnya.

1.6.2 Aspek Praktis

Hasil penelitian diharapkan dapat memberikan manfaat dalam aspek praktis sebagai berikut:

1. Bagi penulis, mendapat pengetahuan dan lebih memahami kinerja dari algoritma C4.5 untuk pengambilan keputusan.

-
2. Untuk pembaca, mendapat pengetahuan mengenai data warehouse yang digunakan sebagai acuan dalam melakukan penelitian.

BAB II

KAJIAN PUSTAKA

2.1 *Knowlegde discovery in database (KDD)*

Menurut (Luvia, Windarto, Solikhun, & Hartama, 2017) *data mining* dan *knowledge discovery in databases (KDD)* dapat digunakan secara bergantian yang berguna untuk menjelaskan proses mengekstraksi data yang tersembunyi dari *database* besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi saling terhubung dan saling berhubungan. Tahapan keseluruhan proses KDD adalah *data mining*.

(Mardi, 2017) *Knowledge Discovery In Database (KDD)* merupakan metode untuk mendapat dan memperoleh pengetahuan dari sebuah dari *database* yang ada. Dalam database terdapat record-record yang saling berhubungan satu sama lain. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk pengambilan keputusan dan penggalian informasi tersembunyi dalam suatu *database*.

Menurut (D. W. T. Putra, 2016) terdapat 4 tahapan menggunakan proses KDD, antara lain :

1. Pengumpulan data

Sumber data utama yang digunakan dalam melakukan penelitian merupakan data yang diperoleh dari data penjualan motor PT. Capella Dinamik Cabang Muka Kuning pada tahun 2019, dimana data tersebut akan dianalisis untuk memperoleh

masalah-masalah yang ada untuk digunakan sebagai sampel dalam melanjutkan penelitian.

2. Penyeleksi data

Pada tahapan ini analisis penelitian diperoleh dari 15 jurnal serta 2 judul buku sebagai referensi sehingga pada tahapan penyeleksi sumber data, data yang akan diambil harus sesuai dengan referensi yang sudah ditetapkan. Pada tahapan ini data akan dikelompokkan sesuai kelas untuk difokuskan pada subset variable dimana variabel tersebut akan menjadi data penelitian yang akan digunakan pada proses *data mining* lalu disimpan dalam sebuah berkas yang akan dipisahkan dari basis data yang tidak dibutuhkan.

3. *Preprocessing / Cleaning*

Tujuan utama dari *Preprocessing / Cleaning* adalah untuk memperoleh sampel data yang akan digunakan sebagai acuan dalam penelitian, membersihkan keseluruhan data yang tidak dibutuhkan, memeriksa isi data serta memperbaiki semua kesalahan yang ada pada data.

4. Transformasi data

Transformasi ini sangat berguna untuk mengintegrasikan data yang belum valid sehingga menjadikan data yang belum mempunyai entitas akan diubah keseluruhan menjadi sebuah data yang valid karena dalam proses *data mining* harus menggunakan data yang valid.

2.2 *Data mining*

Proses mendapatkan suatu informasi yang berguna dari sebuah database merupakan manfaat *data mining*, dari sebuah database yang besar akan di ekstrak dan diambil informasi baru yang berguna untuk membantu dalam pengambilan suatu keputusan, sering juga disebut *knowledge discovery* (Haryati, Sudarsono, & Suryana, 2015).

Menurut (Luvia et al., 2017), berikut ini adalah tahap-tahap *data mining* :

1. Pembersihan data (*data cleaning*)

data cleaning merupakan proses membuang data yang tidak konsisten, *noise* dan data yang tidak relevan.

2. Integrasi data (*data integration*)

data integration merupakan proses menggabungkan data dari beberapa *database* kedalam satu *database* baru yang lebih efektif.

3. Seleksi data (*data selection*)

Hanya data yang dibutuhkan saja yang perlu diambil dari database untuk dianalisis karena tidak semua data bisa digunakan, untuk itu dilakukan seleksi data.

4. Transformasi data (*data transformation*)

data transformation bertugas untuk mengubah dan menggabungkan data menjadi nilai kedalam format yang sesuai untuk selanjutnya diproses kedalam *data mining*, untuk data yang cakupannya terlalu luas akan dilakukan pengelompokan menjadi beberapa kelompok kecil (Putri & Waspada, 2018).

5. Proses *mining*

Menemukan pengetahuan baru yang berharga dan tersembunyi dari berbagai *database* yang digabungkan merupakan istilah dari proses *mining*.

6. Evaluasi pola

Untuk mengidentifikasi pola-pola menarik yang didapatkan dari *database* kedalam *knowledge based* yang ditemukan.

7. Persentasi pengetahuan

Visualisasi dan penyajian pengetahuan tentang metode yang dapat digunakan untuk memperoleh pengetahuan yang didapat pengguna.

Menurut (Mardi, 2017) pengelompokan *data mining* terdiri dari beberapa kelompok berdasarkan metode dan tugas pemrosesannya adalah sebagai berikut.

1. Deskripsi

Deskripsi merupakan sebuah cara dalam menggambarkan dan menemukan sebuah pola dalam sebuah data. Dalam hal ini sering kali peneliti memiliki kebiasaan hanya menjelaskan suatu pola saja, hal ini dikarenakan karna sulitnya menentukan keterangan atau fakta mengenai data yang sudah diperoleh.

2. Estimasi

Sebuah model yang sudah dibangun menggunakan record data yang lengkap, estimasi hampir sama seperti klasifikasi, perbedaannya hanya divariabel target estimasi yang lebih cenderung mengarah ke numerik dibanding ke arah kategori.

3. Prediksi

Prediksi merupakan suatu keadaan dimana nilai sebuah data dihasilkan belum diketahui kepastiannya, mirip dengan klasifikasi dan estimasi, kecuali dalam prediksi nilai dari hasil akan ada dimasa yang akan datang.

4. Klasifikasi

Teknik klasifikasi digunakan sebagai pendekatan secara sistematis untuk mengelompokan data berdasarkan dari kemiripan tertentu dengan menggunakan target variabel berfungsi sebagai pemisah atau sebagai penentu golongan pada tiap kategori (Gunawan, 2016). Misalnya, penggolongan pendapatan dipisahkan kedalam tiga kategori/kelas, yaitu pendapatan rendah, pendapatan sedang, dan pendapatan tinggi.

5. Pengklasteran

Pengklasteran bisa dikatakan sebagai pengelompokan record, memperhatikan atau pengamatan objek-objek yang mempunyai kemiripan yang hampir sama dengan objek lain, tetapi didalam pengklasteran tidak memiliki variabel target seperti pada klasifikasi.

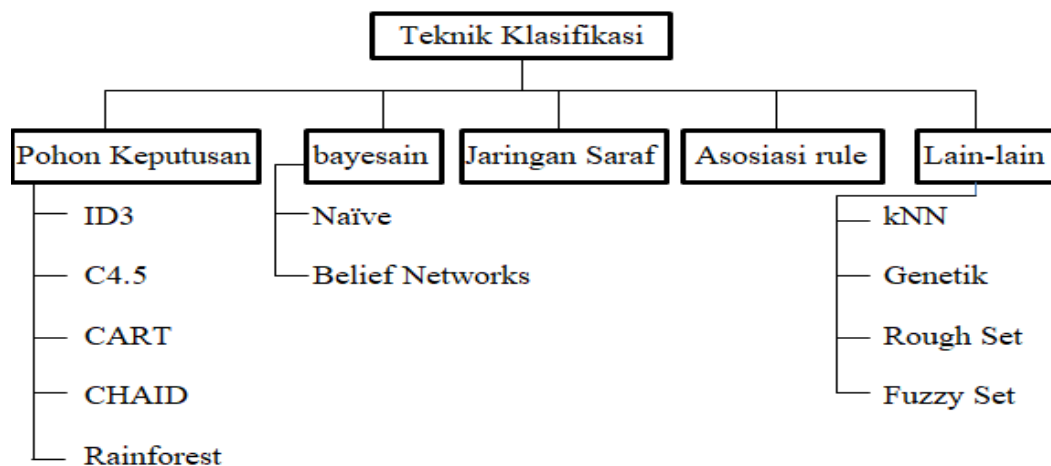
6. Asosiasi

Asosiasi berfungsi untuk menemukan suatu atribut atau mengidentifikasi hubungan antara berbagai atribut yang muncul dalam suatu waktu.

2.3 Metode klasifikasi

Klasifikasi adalah proses untuk mengetahui suatu objek yang sudah didefinisikan sebelumnya. Klasifikasi berfungsi untuk memprediksi kelas dari

suatu objek yang atribut belum diketahui. Dengan memanipulasi data yang sudah didapatkan teknik ini dapat memberikan klasifikasi pada data baru yang telah diklasifikasi dan hasilnya dapat digunakan untuk memberikan sejumlah aturan (Eka Ratnawati, ., & Muflikhah, 2014). Klasifikasi pertama kali diterapkan pada bidang tanaman seperti yang dilakukan oleh Carolus von Linne merupakan peneliti pertama yang mengklasifikasi spesies tanaman berdasarkan karakteristik fisik. Setelah penelitian tersebut dia dikenal sebagai bapak klasifikasi(Mardi, 2017).



Gambar 2.1 Pengelompokan Teknik klasifikasi

Sumber : Data Penelitian, 2019

Algoritma klasifikasi berfungsi untuk menemukan hubungan antara nilai-nilai prediksi dan nilai atribut target (Brevik et al., 2016). Metode klasifikasi mempunyai banyak algoritma, apabila algoritma klasifikasi berbeda berarti harus menggunakan teknik yang juga berbeda untuk mencari hubungan antara nilai-nilai prediksi. Klasifikasi model dapat diuji dengan membandingkan nilai-nilai diprediksi nilai atribut target dikenal dalam satu set data uji (Han, Kamber, & Pei,

2012). Data untuk klasifikasi biasanya dibagi menjadi dua set data, satu untuk membangun model, yang lain untuk pengujian model. (P. Putra, Informatika, & 2018, 2018)

1. Pohon Keputusan

Didalam algoritma C4.5 ada beberapa metode yang dapat digunakan untuk klasifikasi tetapi yang paling sering digunakan adalah pohon keputusan (*decision tree*). Pohon keputusan merupakan sebuah metode yang dapat mengubah fakta yang sangat besar yang sudah didapatkan menjadi sebuah pohon keputusan yang merepresentasikan aturan (*rule*) yang dapat dengan mudah dipahami dengan bahasa alami (Mardi, 2017).

Manfaat utama yang didapat dengan menggunakan pohon keputusan adalah kemampuannya mengambil keputusan yang kompleks dan mudah dipahami menjadi lebih simpel (*break downproces*) sehingga pengambilan keputusan dapat menghasilkan solusi permasalahan. Algoritma C4.5 dan pohon keputusan merupakan dua model yang saling berhubungan dan tidak terpisahkan, karena untuk membangun sebuah pohon keputusan harus membutuhkan algoritma C4.5 (Haryati et al., 2015).

Secara umum ada beberapa langkah untuk membangun pohon keputusan (Rani, 2015) yakni :

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap atribut
3. Bagi kasus dalam cabang

Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2. Algoritma C4.5

Algoritma C4.5 pertama kali diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma ID3 digunakan untuk membuat sebuah pohon keputusan yang berguna untuk mengeksplorasi data dan menemukan hubungan-hubungan tersembunyi antara sejumlah calon variabel-variabel input dengan variabel target. Pada dasarnya konsep dari algoritma C4.5 digunakan untuk mengubah data menjadi pohon keputusan dan menghasilkan aturan-aturan (*rule*) keputusan (Mukminin & Riana, 2017).

Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian akar atau node dari sebuah pohon keputusan. Tahapan dalam membuat pohon keputusan dengan algoritma C4.5 (Septiani, 2017) yaitu:

1. Mempersiapkan data training, dapat diambil dari data histori atau data penjualan yang pernah terjadi sebelumnya serta mewawancarai pihak yang terkait dalam penelitian guna mendapatkan sampel data yang dibutuhkan, selanjutnya data yang sudah didapatkan kemudian dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon keputusan yang didapatkan dengan menghitung nilai gain yang tertinggi dari masing-masing *entropy* yang didapatkan dari tiap-tiap partisi atribut yang akan dijadikan sebagai akar dalam pohon keputusan dengan rumus sebagai berikut :

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

Dimana :

- S : himpunan kasus
- A : fitur
- N : jumlah partisi
- P_i : proporsi dari S_i terhadap S

3. Hitung nilai gain dengan menggunakan rumus sebagai berikut :

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

Dimana :

- S : himpunan kasus
- A : atribut
- N ; jumlah partisi ke A
- S_i : jumlah kasus pada partisi ke_i
- S : jumlah kasus dalam S

4. Ulangi kembali langkah ke-2 hingga semua record terpartisi mendapatkan nilai entropy dan gain. Proses dari partisi pohon keputusan akan selesai dan berhenti apabila:

- a. Semua tupel dalam record dalam simpul N mendapat kelas yang sama.
- b. Tidak ada atribut dalam record yang dipartisi lagi.
- c. Tidak ada record di dalam cabang yang kosong.

2.4 Software Pendukung

2.4.1 Rapidminer

Data mining mempunyai beberapa perangkat lunak aplikasi atau software pendukung seperti *WEKA*, *Orange*, *Microsoft Analysis Service*, *Oracle Data Mining*, dan *RapidMiner* dimana aplikasi-aplikasi ini dapat membantu dalam pengolahan *data mining*. Pada penelitian ini, peneliti akan menggunakan aplikasi *RapidMiner* untuk melakukan pengujian.

RapidMiner adalah *software* yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan Ralf Klinkenberg dari *rapid-i.com* dengan tampilan GUI sehingga memudahkan pengguna dalam mengolah data menggunakan perangkat lunak ini.

Perangkat lunak ini bersifat open source yang dapat dijalankan di sistem operasi manapun dan dibuat dengan menggunakan program *Java*. Dengan menggunakan *RapidMiner*, tidak dibutuhkan kemampuan menggunakan bahasa pemrograman, karena semua fasilitas sudah disediakan. *RapidMiner* dikhususkan untuk penggunaan *data mining*. Model yang disediakan, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain. (Haryati et al., 2015)

2.5 Penelitian Terdahulu

Terdapat beberapa penelitian yang digunakan peneliti sebagai referensi dalam melakukan penelitian yang berupa jurnal, referensi-referensi tersebut diambil karena adanya permasalahan atau metode yang berhubungan dengan penelitian yang dibahas didalamnya , antara lain sebagai berikut:

1. Pada jurnal referensi yang ditulis oleh (P. Putra et al., 2018) yang berjudul **PENGEMBANGAN APLIKASI PERHITUNGAN PREDIKSI STOCK MOTOR MENGGUNAKAN ALGORITMA C 4.5 SEBAGAI BAGIAN DARI SISTEM PENGAMBILAN KEPUTUSAN**, meneliti tentang inventaris sepeda motor sesuai dengan kebutuhan konsumen. Berdasarkan proses perhitungan yang telah dilakukan beberapa tahap selama implementasi pemrosesan data C4.5 diperoleh pohon keputusan dengan 12 aturan-aturan (*rule*) dalam menentukan prediksi jumlah persediaan stock motor pada dealer saudara motor.
2. Pada jurnal referensi yang ditulis oleh (Harahap, 2015) yang berjudul **PENERAPAN DATA MINING DALAM MEMPREDIKSI PEMBELIAN CAT FITRIANA**, meneliti tentang Departement Penjualan Home Smart Medan mengolah data perusahaan hanya menggunakan aplikasi Microsoft excel sehingga terkadang tidak dapat bersaing dengan perusahaan lain dalam mengolah data untuk dijadikan informasi sebagai strategi untuk penjualan. Berdasarkan hasil penelitian yang dilakukan bahwa pembelian cat dengan menggunakan metode *data mining* khususnya algoritma C4.5 akan bermanfaat sekali dalam proses pengambilan keputusan dalam pembelian cat pada Home Smart Medan.
3. Pada jurnal referensi yang ditulis oleh (Nababan & Tanlim, 2019) yang berjudul **ANALIS PERSEDIAAN STOK BARANG MENGGUNAKAN ALGORITMA C4.5 PADA CV HARAPAN JAYA**, meneliti tentang jumlah penjualan yang fluktuatif mengakibatkan stok barang yang tersedia

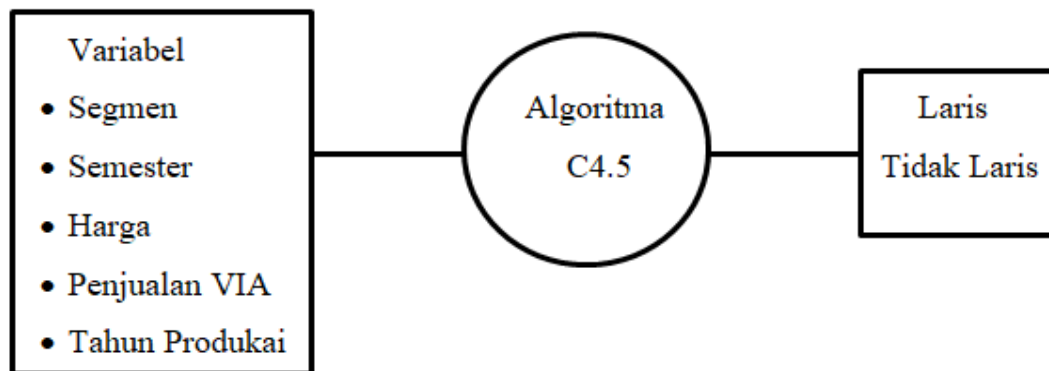
tidak stabil dan dapat berdampak langsung ke retailer. Berdasarkan penelitian dan perhitungan menggunakan algoritma C4.5 maka diperoleh pohon keputusan dengan 9 aturan-aturan (*rule*) dalam menentukan prediksi jumlah persediaan stok barang pada CV Harapan Jaya.

4. Pada jurnal referensi yang ditulis oleh (Jurnal & Informasi, 2016) yang berjudul **PENERAPAN DATA MINING UNTUK MEMREDIKSI PENJUALAN WALLPAPER MENGGUNAKAN ALGORITMA C4.5** meneliti tentang cara mempermudah penjual memilih wallpaper mana yang banyak diminati konsumen agar disediakan stok maka perlu dilakukan prediksi untuk penjualan wallpaper terbanyak dengan metode klasifikasi. Berdasarkan penelitian yang dilakukan yang menjadi faktor utama yang mempengaruhi penjualan adalah faktor jumlah model wallpaper sedangkan faktor harga, ukuran, kualitas bahan dan warna tidak mempengaruhi pembelian.
5. Pada jurnal referensi yang ditulis oleh (Permasalahan & Masalah, 2017) yang berjudul **PERANCANGAN APLIKASI TREND PENJUALAN DAN STOK BARANG MENGGUNAKAN WAREHOUSE DAN DATA MINING** meneliti tentang penggunaan data warehouse perusahaan untuk mendapatkan ramalan penjualan dengan *data mining*. Dari hasil penelitian yang dilakukan, penggunaan aplikasi analisa *trend* penjualan ini dapat dijadikan acuan pada perusahaan dalam mengambil kebijakan dan strategi tepat yang dibutuhkan agar target penjualan semakin meningkat dimasa mendatang.

6. Pada jurnal referensi yang ditulis oleh (Elisa, 2017) yang berjudul **ANALISA DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK MENGIDENTIFIKASI FAKTOR-FAKTOR PENYEBAB KECELAKAAN KERJA KONTRUKSI PT.ARUPADHATU ADISESANTI** meneliti tentang penggunaan algoritma C.45 untuk mengidentifikasi penyebab kecelakaan kerja agar dapat digunakan sebagai panduan untuk menghindari resiko kecelakaan (zero accident), dari hasil penelitian yang dilakukan diambil sebuah kesimpulan bahwa metode Algoritma C4.5 atau pohon keputusan lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian.
7. Pada jurnal referensi yang ditulis oleh (Jamhur, 2016) yang berjudul tentang **PENERAPAN DATA MINING UNTUK MENGANALISA JUMLAH PELANGGAN AKTIF DENGAN MENGGUNAKAN ALGORITMA C4.5** meneliti tentang tingkat kepuasan pelanggan dengan mengetahui jumlah pelanggan aktif Untuk itu perlu adanya pengolahan data tentang pelanggan aktif yaitu dengan menggunakan algoritma C4.5. Hasil dari penelitian dapat menjadi kriteria evaluasi untuk pelanggan aktif dan tidak aktif menggunakan algoritma C4.5 dan membuat aturan yang dapat menggambarkan proses yang terkait dengan pelanggan aktif dan tidak aktif.

2.6 Kerangka Pemikiran

Penelitian dilakukan menggunakan tahapan-tahapan kegiatan dengan mengikuti kerangka pemikiran yang meliputi metode pengumpulan data, analisis data, dan pengujian hasil. Berikut ini adalah kerangka pemikiran penulis dalam melakukan penelitian.



Gambar 2.2 Kerangka Pemikiran
Sumber : Data Penelitian, 2019

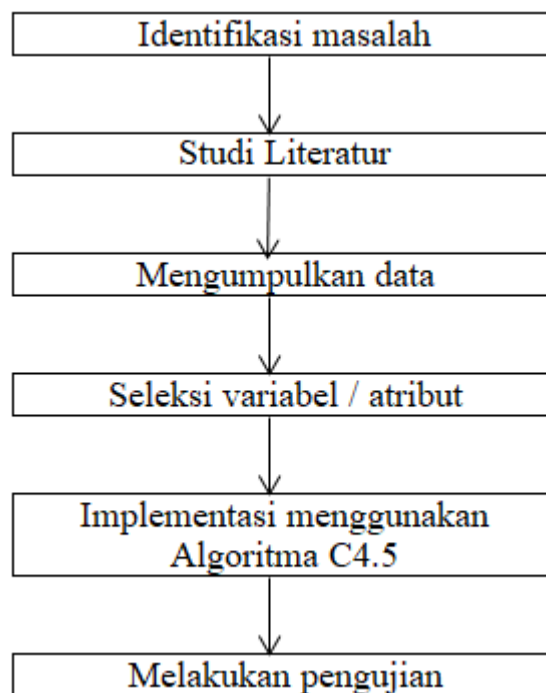
Pada gambar 2.2, penelitian ini menggunakan 4 variabel sebagai input data yaitu segmen, Harga, semester, penjualan VIA dan tahun produksi kemudian data tersebut akan diolah menggunakan *algoritma C4.5* dengan teknik klasifikasi untuk mendapat *output* laris dan tidak laris, dari situ kita akan membuat pohon keputusan (*decision tree*) yang akan menghasilkan rule atau aturan berdasarkan pohon keputusan untuk mendapatkan hasil dan menyimpulkan penelitian yang telah dilakukan.

BAB III

METODE PENELITIAN

2.1 Desain Penelitian

Metode penelitian ini dilakukan secara sistematis dan dapat berfungsi sebagai panduan bagi peneliti untuk melakukan penelitian dari awal hingga selesai, sehingga hasil yang diperoleh tidak menyimpang dari tujuan yang diinginkan dan dapat dilakukan dengan tepat dan sesuai dengan tujuan yang ditetapkan. Bentuk desain penelitian adalah sebagai berikut



Gambar 3.1 Desain Penelitian
Sumber : Data Penelitian, 2019

1. Identifikasi Masalah

Pada dasarnya semua penelitian dilakukan karna adanya sebuah masalah, identifikasi masalah dalam hal ini berguna untuk peneliti agar fokus pada titik permasalahan, dalam kasus ini peneliti ingin menganalisa dengan melakukan observasi klasifikasi data penjualan PT Capella dinamik nusantara dengan menggunakan algoritma C4.5 sekaligus ingin mengetahui bagaimana pola dan cara kerja klasifikasi dalam *data mining*.

2. Studi Literatur

Studi Literatur dilakukan dengan tujuan untuk mencari jurnal, buku dan artikel yang berhubungan dengan argoritma C4.5 klasifikasi, decision tree, dan *data mining* yang relevan dengan penelitian dan semuanya digunakan sebagai bahan referensi untuk memperdalam pemahaman teori-teori dan penerapan data.

3. Pengumpulan Data

Data-data penelitian didapat dari PT. Capella Dinamik Nusantara cabang muka kuning, dimana data yang diambil adalah data penjualan sepeda motor dalam satuan unit dan menganalisa kasus-kasus yang terjadi pada data yang sudah didapatkan untuk dijadikan sebagai sampel dalam penelitian.

4. Seleksi atribut

Setelah data dianalisa, selanjutnya data akan diolah menjadi data yang berbentuk informasi, data-data tersebut akan diseleksi variabel dan atributnya berdasarkan kebutuhan penelitian dan kemudian melakukan tahapan cleaning untuk menghilangkan atribut-atribut atau variabel yang tidak berguna atau yang tidak dibutuhkan.

5. Implementasi menggunakan algoritma C4.5

Pengolahan data sangat menentukan keberhasilan dari penelitian, pada proses ini peneliti akan mengimplementasikan data atau kasus-kasus yang didapatkan dengan menggunakan metode algoritma C4.5 klasifikasi mulai dari menentukan *dataset*, atribut, dan variabel harus sangat diperhatikan. Data-data tersebut akan diolah sehingga akan menghasilkan pohon keputusan.

6. Melakukan pengujian

Pengujian ini bertujuan untuk mengetahui untuk bagaimana cara kerja dari *algoritma C4.5* mengklasifikasikan data kedalam kelas yang ditentukan, maka peneliti menggunakan *tools* atau aplikasi *RapidMiner*. jika pengujian belum sesuai dengan kebutuhan yang diinginkan maka tahapan penelitian kembali ke pengolahan data sampai pengujian berhasil.

3.2 Teknik Pengumpulan data

Ada 3 metode yang digunakan peneliti dalam proses pengumpulan data penelitian, antara lain sebagai berikut :

1. Observasi

Peneliti melakukan observasi langsung ke lapangan untuk mengamati dan mengumpulkan data-data mengenai penjualan yang diperlukan dalam penelitian.

2. Wawancara

Wawancara dilakukan peneliti untuk menanyakan kepada marketing saudara khususnya tentang persediaan unit sepeda motor karena data tersebut dapat dijadikan pertimbangan untuk pengambilan keputusan.

3. Kajian Pustaka

Kajian pustaka sangat diperlukan peneliti sebagai bahan penelitian untuk menjelaskan secara tepat penelitian yang dilakukan, kajian pustaka bisa dari media seperti buku, jurnal-jurnal yang berhubungan dengan metode dan topik penelitian yang dilakukan sebagai bahan referensi untuk memperdalam pemahaman teori-teori dan penerapan data.

3.3 Operasional Variabel

Operasional variabel yang digunakan dalam penelitian ini adalah data penjualan tahun 2019 dengan menggunakan variabel sebagai berikut.

1. Penjualan berdasarkan segmen

Data-data yang digunakan dalam operasional variabel ini sebanyak 23 data yang diambil dari penjualan perbulannya yang terbagi menjadi 3 segmen, yaitu : *cub*, *MATIC*, *sport*. Sepeda motor *cub* memiliki ukuran mesin dalam kisaran antara 50 cc sampai 125 cc dengan kecepatan 100-125 km/jam, atau sering juga disebut motor bebek. Motor *MATIC* adalah jenis motor dengan sekali akselerasi. Sedangkan motor *sport* dikhususkan untuk penggunaan balap atau mempunyai kecepatan yang tinggi dengan bodi yang cukup besar dengan kecepatan minimal 150 km/jam.

2. Penjualan VIA

Data-data yang digunakan dalam operasional variabel ini terbagi menjadi 4 variabel berdasarkan lokasi penjualannya yaitu penjualan via showroom, sp plaza, panbil, dan kepri mall, masing-masing memiliki 23 data yang diambil dari data

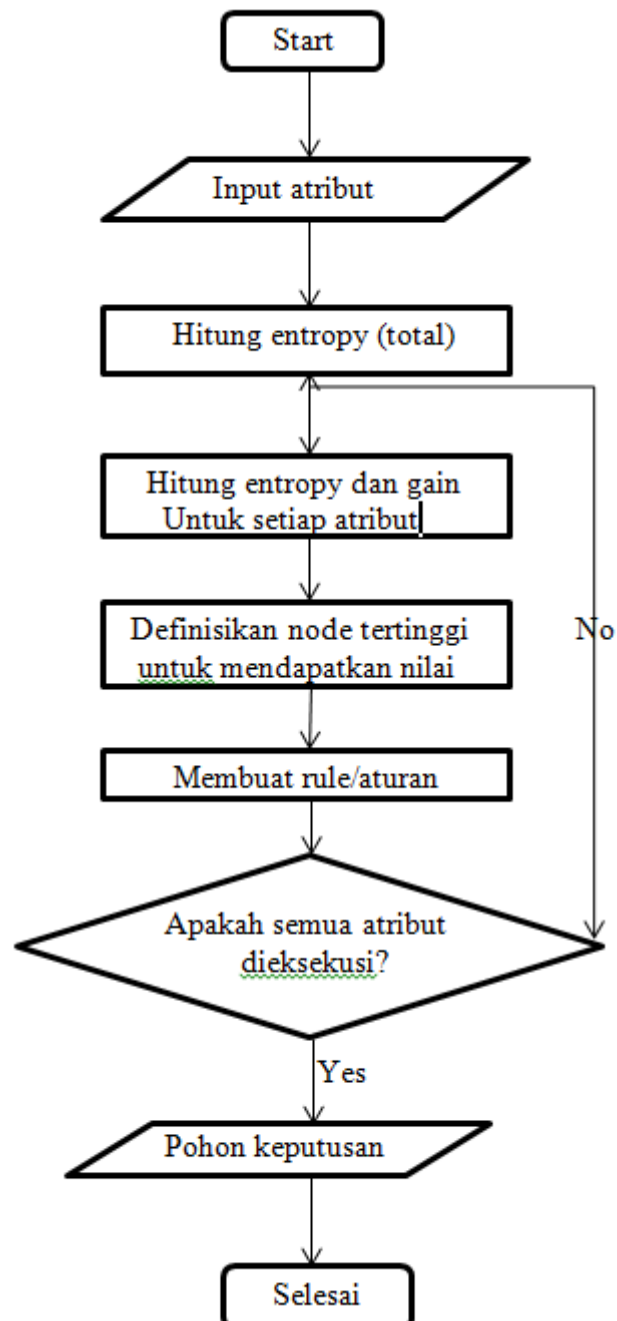
penjualan perbulannya.dari keseluruhan data penjualan perbulannya akan dibagi berapa unit sepeda motor yang terjual via showroom, sp plaza, panbil, dan kepri mall.

3. Tahun produksi

Data-data yang digunakan dalam operasional variabel ini berdasarkan tahun produksi yang memiliki 2 variabel yaitu tahun produksi 2017 dan 2018. Pada variabel ini akan dibagi jumlah keseluruhan penjualan perbulannya, berapa unit sepeda motor yang terjual pada tahun produksi 2017 dan 2018.

3.4 Metode Analisis dan Rancangan sistem

Adapun metode analisis dan perancangan sistem pada penelitian ini adalah sebagai berikut.



Gambar 3.2 Rancangan Sistem
Sumber : Data Penelitian 2019

Pada gambar 3.2 pertama-tama yang dilakukan yaitu mengumpulkan data-data yang dibutuhkan untuk penelitian, selanjutnya data yang sudah didapatkan akan dilakukan seleksi atribut, kemudian melakukan proses *cleaning* untuk membuang data yang tidak diperlukan, setelah sampel sudah didapat akan diolah dengan menggunakan algoritma *C4.5 classification*. Langkah-langkah dari metode *C4.5 classification* adalah

1. Tentukan atribut-atribut yang digunakan sebagai sampel untuk melanjutkan penelitian.
2. Menghitung node-node (root) pada atribut untuk mengetahui nilai entropy total dari masing-masing atribut.
3. Menghitung gain untuk tiap-tiap atribut dan menentukan gain tertinggi dari masing-masing atribut.
4. Setelah nilai entropy total, entropy masing-masing atribut, nilai gain dan menentukan gain tertinggi sudah didapatkan selanjutnya membuat pohon keputusan.
5. Membuat aturan atau rule berdasarkan pohon keputusan.
6. Menguji hasil perhitungan yang dilakukan secara manual menggunakan algoritma *C4.5* dengan *software RapidMiner*.
7. Kembali ke langkah 2 jika hasilnya belum sesuai, setelah proses pengolahan data menggunakan algoritma *C4.5* telah selesai maka akan mendapatkan hasil yang sesuai dengan perhitungan manual.

3.5 Lokasi dan Jadwal Penelitian

Berikut ini merupakan lokasi dan jadwal pelaksanaan penelitian yang telah dilakukan oleh peneliti yaitu sebagai berikut :

3.5.1 Lokasi Penelitian

Lokasi penelitian dilakukan pada showroom Honda PT Capella Dinamik Nusantara cabang muka kuning yang beralamat di Jl. LetJend.Suprpto Komp.Taman Merapi Subur Blok A3 No. 1 - 5 Tembesi - Muka Kuning, Tembesi, Kec. Sagulung, Kota Batam, Kepulauan Riau 29424.

3.5.2 Jadwal Penelitian

Penelitian ini dilaksanakan mulai dari september 2019 hingga bulan februari 2020. Jadwal penelitian dalam penelitian ini dapat dilihat ditabel berikut:

Tabel 3.1 Jadwal Penelitian

Kegiatan	Waktu Kegiatan																						
	Sept 2019			Okt 2019				Nov 2019				Des 2019				Jan 2020				Feb 2020			
	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pengajuan Judul	■	■	■																				
Penyusunan BAB I				■	■	■	■																
Penyusunan BAB II								■	■	■	■												
Penyusunan BAB III												■	■	■	■								
Penyusunan BAB IV																■	■	■	■				
Penyusunan BAB V																				■	■		
Pengumpulan Skripsi																						■	■

Sumber: Data Penelitian, 2019