

**ANALISA DATA MINING MENGGUNAKAN METODE
BAYES UNTUK MENGUKUR TINGKAT
KERUSAKAN MESIN MOTOR**

SKRIPSI



Oleh:
Sheryl Febry Irianti Simanjuntak
161510073

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2021**

**ANALISA DATA MINING MENGGUNAKAN METODE
BAYES UNTUK MENGUKUR TINGKAT
KERUSAKAN MESIN MOTOR**

SKRIPSI

**Untuk memenuhi syarat
memperoleh gelar Sarjana**



**Oleh:
Sheryl Febry Irianti Simanjuntak
161510073**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2021**

HALAMAN PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Sheryl Febry Irianti Simanjuntak

NPM : 161510073

Fakultas : Teknik dan Komputer

Program Studi : Sistem Informasi

Menyatakan bahwa “SKRIPSI” yang saya buat dengan judul :

Analisa Data Mining Menggunakan Metode Bayes untuk Mengukur Tingkat Kerusakan Mesin Motor

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Seperngetahuan saya, di dalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah di tulis atau di terbitkan oleh orang lain, kecuali yang secara tertulis di kutip di dalam naskah ini dan di sebutkan dalam sumber kutipan dan daftar pustaka.

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

Demikian pernyataan ini saya buat dengan sebenarnya tanpa paksaan dari siapapun

Batam, 27 Januari 2021



Sheryl Febry Irianti Simanjuntak

161510073

**ANALISA DATA MINING MENGGUNAKAN METODE
BAYES UNTUK MENGUKUR TINGKAT
KERUSAKAN MESIN MOTOR**

SKRIPSI

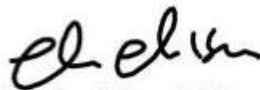
Untuk memenuhi syarat
memperoleh gelar Sarjana

Oleh:

**Sheryl Febry Irianti Simanjuntak
161510073**

Telah di setujui Pembimbing pada tanggal
seperti tertera di bawah ini

Batam, 27 Januari 2021



Erlin Elisa, S.Kom., M.Kom.

Pembimbing

ABSTRAK

Bengkel merupakan sebuah usaha yang bergerak di bidang usaha otomotif seperti usaha menengah kebawah bengkel menyediakan *spare part* motor yang lengkap tidak hanya itu bengkel secara umum juga melayani *service* motor dan menerima *service* ringan maupun *service* berat untuk segala jenis merek motor. Hendri Motor merupakan salah satu bengkel yang berada di piayu yang sudah banyak memiliki konsumen yang pada setiap bulannya menerima *service* motor di Kota Batam, lokasinya strategis karena dekat dengan pemukiman yang banyak memiliki motor. Tingkat kerusakan motor yang bisa di kategorikan kerusakan motor yaitu rusak ringan, rusak sedang, maupun rusak berat karena kebanyakan pemakai lebih memilih ganti oli tiap sebulan sekali ternyata ada juga yang lewat dari jangka waktu yang di tentukan untuk menggantikan oli misalkan bisa sampai 2 bulan sampai 5 bulan tidak menggantikan oli yang baru. Analisa diperlukan untuk melihat pola dari data konsumen sehingga dapat menghasilkan probabilitas *service* motor yang nantinya akan berguna untuk di kategorikan dalam *service* yang ringan ataupun *service* berat. Dari data konsumen yang begitu banyak, maka dilakukan *Data Mining* dengan menggunakan *Algoritma Naïve Bayes* . Hasil dari kegiatan *mining* ini diharapkan dapat memberikan sebuah keputusan untuk melihat pola prediksi perilaku konsumen *service* motor.

Kata kunci: *algoritma naïve bayes, data mining, klasifikasi, probabilitas.*

ABSTRACT

Workshop is a business that is engaged in the automotive business such as a medium-sized business down the workshop providing complete motorcycle spare parts not only that the workshop in general also serves motorcycle service and accepts light and heavy service for all types of motorcycle brands. Hendri Motor is one of the workshops in Piayu which already has many customers who receive motorcycle service every month in Batam City, the location is strategic because it is close to settlements that have many motorbikes. The level of motor damage that can be categorized as motor damage is mild damage, moderate damage, or severe damage because most users prefer to change oil once a month, in fact there are also those who have passed the specified time period to replace oil, for example it can be up to 2 months to 5 month does not replace new oil. Analysis is needed to see the pattern of consumer data so that it can produce motor service probabilities that will later be useful for categorizing in light or heavy service. From so much consumer data, Data Mining is performed using the Naïve Bayes Algorithm. The results of this mining activity are expected to provide a decision to see the prediction patterns of motor service consumer behavior.

Keywords: naïve bayes algorithm, data mining, classification, probability.

KATA PENGANTAR

Segala puji dan syukur hanya bagi 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 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, pembimbing, dan dorongan dari berbagai pihak. Untuk itu, 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 Welly Sugiyanto, S.T., M.M. selaku Dekan Fakultas Teknik dan Komputer Universitas Putera Batam
3. Bapak Muhammat Rasid Ridho, S.Kom., M.SI. selaku Ketua Program Studi Sistem Informasi Universitas Putera Batam
4. Ibu Erlin Elisa, S.Kom., M.Kom. selaku pembimbingan Skripsi
5. Bapak Nopriadi, S.Kom., M.Kom. selaku pembimbing Akademik
6. Seluruh Dosen dan Staff Universitas Putera Batam yang telah memberikan pengetahuan dan ilmunya untuk penulis
7. Mama dan Bapa tercinta, saudara kandung (Stephen, Mitha, Putri, Riko, dan Osta) yang telah mendo'akan, dan memberikan semangat, serta inspirasi kepada penulis dalam penyusunan skripsi ini.
8. Rekan-rekan seperjuangan yaitu sahabatku Sistem Informasi Angkatan Tahun 2016 dan teman-teman tongkrongan di kantin Universitas Putera Batam serta penulis yang meneliti diangkatan tahun 2021 saat ini.

9. Bagi pencipta Tuhan Yesus Kristus yang telah memberikan kesehatan dan memberikan kesempatan untuk mengerjakan Skripsi.

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

Batam, 27 Januari 2021



Sheryl Febry Irianti Simanjuntak

DAFTAR ISI

ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar belakang masalah	1
1.2 Identifikasi Masalah	3
1.3 Batasan Masalah	3
1.4 Perumusan Masalah	3
1.5 Tujuan Penelitian	3
1.6 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Teori Umum	5
2.1.1 <i>Service</i> Kendaraan	5
2.2 Teori Khusus	6
2.2.1 <i>Knowledge Discovery in Database (KDD)</i>	6
2.2.2 <i>Data Mining</i>	6
2.2.3 Klasifikasi	8
2.2.4 <i>Clustering</i>	8
2.2.5 <i>Naïve Bayes</i>	9
2.2.6 <i>Software Rapidminer</i>	11
2.3 Penelitian Terdahulu	15
2.4 Kerangka Pemikiran	19
2.5 Hipotesis Penelitian	20
BAB III METODE PENELITIAN	21
3.1 Desain Penelitian	21

3.2 Lokasi Penelitian	23
3.3 Populasi dan Sampel	23
3.3.1 Populasi.....	23
3.3.2 Sampel	23
3.4 Variabel Penelitian	23
3.5 Teknik Pengumpulan Data	24
3.6 Operasional Variabel.....	24
3.7 Model Penelitian.....	25
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	26
4.1 Hasil.....	26
4.1.1 Pengumpulan Data Training	26
4.1.2 Menghitung Class Probalilitas	30
4.1.3 Pengujian Metode <i>Naïve Bayes</i>	33
4.2 Pembahasan	37
BAB V SIMPULAN DAN SARAN	44
5.1 Simpulan.....	44
5.2 Saran	45
DAFTAR PUSTAKA	46
LAMPIRAN 1	48
LAMPIRAN 2	56
LAMPIRAN 3	57

DAFTAR GAMBAR

Gambar 2.1 <i>View Welcome Perspective</i>	12
Gambar 2.2 <i>View Design Prespective</i>	13
Gambar 2.3 <i>View Result Perspective</i>	15
Gambar 2.4 Kerangka Pemikiran.....	19
Gambar 3.1 <i>Design Penelitian</i>	21
Gambar 3.2 Model Penelitian	25
Gambar 4.1 Halaman Utama <i>Rapid Miner</i>	38
Gambar 4.2 Area Kerja <i>Rapid Miner</i>	38
Gambar 4.3 Operator <i>Rapid Miner</i>	39
Gambar 4.4 <i>Data Import Wizard Step 1 of 4</i>	39
Gambar 4.5 <i>Data Import Wizard - Step 2 of 4</i>	40
Gambar 4.6 <i>Data Import Wizard - Step 4 of 4</i>	41
Gambar 4.7 Proses Pengujian Data.....	41
Gambar 4.8 Hasil Pengujian <i>Correct Prediction</i>	42
Gambar 4.9 Hasil Pengujian <i>Wrong Prediction</i>	42
Gambar 4.10 Hasil Pengujian <i>Performance Vektor</i>	43

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	16
Tabel 4.1 <i>Data Training</i>	26
Tabel 4.2 <i>Class Probabilities</i>	30
Tabel 4.3 Atribut Merk Motor	31
Tabel 4.4 Atribut Masa <i>Service</i>	31
Tabel 4.5 Atribut Masalah Ringan	32
Tabel 4.6 Atribut Masalah Sedang	32
Tabel 4.7 Atribut Masalah Berat	33
Tabel 4.8 Hasil Pengujian	35

BAB I

PENDAHULUAN

1.1 Latar belakang masalah

Dengan berjalannya waktu, teknologi Informasipun mulai berkembang sangat pesat di dalam negeri maupun di luar negeri. Kemajuan Teknologi pun terus-menerus berkembang di bidang industri dan jasa dalam hal ini sudah memasuki otomasi di perkiraan mempengaruhi kecepatan perkembangan industri otomotif di tingkat persaingan antar perusahaan di bidang perekonomian Indonesia. Dalam Pendataan dan ketersediaan data pun semakin cepat adanya sehingga banyak data dan kebutuhan informasi sebagai pendukung pengambilan keputusan untuk membuat bisnis dan dukurngan infrastruktur di bidang teknik informatika (Azwanti, 2018).

Terlihat berdasarkan catatan Asosiasi Industri Sepeda Motor Indonesia (AISII) dari data penjualan motor di dalam negeri hingga Oktober 2019 sukses mencatatkan 5.517.630 unit. Peroleh tersebut mengungguli capaian tahun 2018 dalam bulan 10 pertama dengan 5.332.564 unit. Dengan mendekati target penjualan motor dalam negeri untuk tahun 2019 ini sebesar 6.4 juta unit. Menurut Data dari Badan Pusat Statistik di Kota Batam jumlah kendaraan roda dua mencapai 216.248 unit dan kendaraan roda empat mencapai 413.986 unit.

Bengkel merupakan usaha yang bergerak di bidang otomotif sehingga banyak persaingan bisnis yang membuat pebisnis tertarik menjalankan bisnis otomotif ini. Meskipun usaha menengah kebawah, bengkel menyediakan *spare part* motor yang lengkap tidak hanya itu bengkel secara umum juga melayani *service* motor dan menerima *service* ringan maupun *service* berat untuk segala jenis merek motor. Hendri Motor adalah salah satu bengkel yang tepat berada di

piayu yang sudah banyak memiliki konsumen yang pada setiap bulannya menerima *service* motor di Kota Batam, lokasinya strategis karena dekat dengan pemukiman yang banyak memiliki motor. Salah satunya dekat dengan rumah kos yang jarang berjarak di piayu pada hal kebanyakan karyawan-karyawan PT yang setiap hari senin sampai hari sabtu bekerja menggunakan motor. Tidak hanya untuk bekerja memakai motor biasanya mau ke warung terdekat pun harus bermotor karena tidak membuang waktu berlama-lama untuk berjalan, Sehingga mempercepat aktivitas. Seringnya pemakaian kendaraan membuat hal-hal yang tidak di inginkan perlu peralatan dan perawatan khusus yang di lakukan salah satu solusi utama perbaikan motor yaitu di bengkel yang di tangani dengan cepat oleh orang yang benar-benar ahli sehingga cepat di perbaiki.

Permasalahan yang di alami pihak bengkel yaitu belum dapat menentukan tingkat kerusakan motor yang bisa di kategorikan kerusakan motor yaitu rusak ringan, rusak sedang, maupun rusak berat karena kebanyakan pemakai lebih memilih ganti oli tiap sebulan sekali ternyata ada juga yang lewat dari jangka waktu yang di tentukan untuk menggantikan oli misalkan bisa sampai 2 sampai 5 bulan tidak menggantikan oli yang baru. Tidak hanya oli, *spare part* yang lain pun jarang untuk di ganti, jika masih bisa di pakai kenapa harus ganti yang baru sehingga membuat motor sampai rusak berat dulu baru pemakai membawa motornya ke bengkel untuk di *service* karena pemakai memiliki alasan tersendiri, ada yang ekonominya belum cukup untuk menggantikan *spare part* dengan yang baru atau sering di sebut *Original* karena *spare part* terlalu mahal dan ada juga yang tidak memiliki waktu untuk memperbaiki motornya ke bengkel.

Berdasarkan permasalahan diatas, perlu adanya metode *Naïve Bayes Classifier* pengklasifikasi probabilitas yang sederhana berdasarkan teorema Bayes (Hikmah & Utammimah, 2017). Ini memungkinkan pemakai dapat mengetahui secara cepat tingkat kerusakan yang terjadi, *spare part* apa yang harus diganti atau *spare part* yang tidak harus di ganti dan bagaimana perawatan yang harus di *service*. Berdasarkan hasil pengamatan yang diamati penulis untuk melakukan penelitian dengan membuat Tugas Akhir ke dalam bentuk skripsi

dengan judul ”ANALISA *DATA MINING* MENGGUNAKAN METODE *BAYES* UNTUK MENGUKUR TINGKAT KERUSAKAN MESIN MOTOR”

1.2 Identifikasi Masalah

Berdasarkan dari penulis yang meneliti dapat disimpulkan identifikasi masalahnya adalah:

1. Kesulitan dari pemilik bengkel untuk menentukan jenis *service* yang akan di berikan kepada pelanggan.
2. Sering terjadinya kesalahan prediksi yang mengakibatkan kendaraan pelanggan lebih buruk.

1.3 Batasan Masalah

Berdasarkan masalah yang ada penelitian ini fokus pada *service* motor ringan, sedang dan berat. Objek penelitian ini beralamat di piayu Bengkel Perum. Bida Ayu, Jl. S. Parman, Mangsang, Kec. Sei Beduk, Kota Batam, Kepulauan Riau 29444. Menggunakan data konsumen yang memiliki kerusakan pada mesin motor dan di olah menggunakan metode *Naïve Bayes Classifier* adalah pengklasifikasi probabilitas

1.4 Perumusan Masalah

Berdasarkan dari penulis yang meneliti latar belakang permasalahan dapat diambil :

1. Bagaimana menentukan kerusakan ringan, kerusakan sedang, dan kerusakan berat untuk tingkat kerusakan mesin pada motor ?
2. Bagaimana menentukan *service* motor yang tepat sasaran kepada konsumen hendri motor?

1.5 Tujuan Penelitian

Tujuan dari penulis untuk penelitian ini adalah sebagai berikut:

1. Untuk menentukan oleh pihak bengkel tentang mesin motor yang rusak ringan maupun rusak berat.

2. Sering terjadinya kesalahan prediksi yang mengakibatkan kendaraan pelanggan lebih buruk.

1.6 Manfaat Penelitian

Manfaat penelitian ini yang dapat diterima peneliti dan pihak lainnya :

- a. Bagi penulis
Memberikan wawasan yang luas dengan menggunakan *Big Data Mining Bayes*.
- b. Bagi Universitas Putera Batam
Memberikan referensi dan pedoman penelitian bagi mahasiswa yang akan melakukan penelitian khususnya yang menggunakan Teknik perhitungan *Data Mining*.
- c. Bagi peneliti lain
Dapat memberi pengetahuan mengenai probabilitas penggunaan Metode *Naïve Bayes* dengan cara yang berbeda dalam menambah kajian baru.
- d. Objek penelitian
Memberikan informasi penggunaan metode *Naïve Bayes* sebagai solusi untuk tingkat kerusakan pada mesin motor sehingga *service* ringan, *service* sedang, dan *service* berat.

BAB II

TINJAUAN PUSTAKA

2.1 Teori Umum

2.1.1 Service Kendaraan

Service adalah perbaikan atau jasa, penjelasan dari perbaikan yaitu merupakan usaha mengembalikan kondisi dan fungsi dari suatu benda misalnya alat yang rusak atau tidak layak pakai mengakibatkan penggunaan alat ini pada kondisi awal. Proses perbaikan tidak mengukur kenyamanan sesuai kondisi pertama, yang diutamakan adalah alat yang di gunakan bisa berfungsi sebagaimana penggunaannya. Perbaikan akan terjadinya pergantian bagian alat (*sparepart*) (Damayanti & Ambarita, 2018).

Sedangkan kendaraan yaitu alat transportasi yang beranake ragam bentuk nya dari motor, mobil, dan lain sebagainya. Salah satu kendaraan *favorite* tersebut yaitu motor merupakan alat kendaraan beroda dua yang dihidupkan oleh sebuah mesin. Motor memerlukan perawatan secara berkala, ini merupakan salah satu kegiatan perawatan berkala maka akan memberikan banyak manfaat bagi pemilik kendaraan sehingga diwajibkan untuk service motor. Untuk perawatan berkala sebaiknya di lakukan di bengkel resmi kendaraan karena pengerjaannya sesuai dengan prosedur.

Kompetensi perawatan secara berkala motor yaitu: (1) periksa setiap sistem motor untuk memastikan sistem tersebut masih berfungsi; (2) membersihkan sistem yang kotor; (3) mengatur kembali system sehingga sesuai dengan standar; (4) Memperbaiki dan mengganti komponen sepeda motor yang rusak (Hidayat et al., 2018).

2.2 Teori Khusus

2.2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery In Database merupakan proses menentukan sebuah informasi yang berguna serta menentukan pola-pola yang berada didalam data. Informasi ini terkandung didalam database yang berukuran besar sebelumnya tidak diketahui (Hartini & Informasi, n.d.)

Hasil dari penelitian yaitu tahapan - tahapan proses didalam KDD (*Knowledge Data Discovery*) sebagai berikut (Utami, 2020) :

1. *Data Selection*, data seleksi ini dilakukan untuk memilih data yang sesuai variabel sehingga dibutuhkan dalam penelitian. Caranya dengan menentukan atribut-atribut data mana yang akan digunakan dalam penelitian dari sekelompok data operasional.
2. *Pre- processing/Cleaning*, proses cleaning tersebut dilakukan terhadap keseluruhan data yang jumlah yang di teliti 50 data. Setelah di lakukan proses cleaning sebesar 50 data, dihasilkan dari data bersih 50 *record* data yang digunakan untuk proses analisis.
3. *Transformation*, proses yang satu ini menghasilkan satu *record set* data.
4. Analisis data.

2.2.2 *Data Mining*

Data mining adalah suatu proses yang paling pertama yang akan dilakukan yaitu mencari informasi baru (Elisa et al., 2019). *Data mining* di gunakan untuk mengambil intisari pengetahuan dari sekumpulan data yang dapat dengan mudah dimengerti oleh manusia meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, *visualisasi*, dan *online updating* (Laia et al., 2018) .

Gartner Group mengatakan *Data mining* adalah proses menemukan hubungan baru yang memiliki arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan

menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Idris, 2019).

Data mining dikelompokkan berdasarkan tugas yang dapat dilakukan, yaitu :

1. *Description* (Deskripsi)

Peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. *Estimation* (Estimasi)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

3. *Prediction* (Prediksi)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah :

- a. Prediksi harga minyak dalam tiga bulan yang akan datang.
- b. Prediksi tingkat pengangguran lima tahun akan datang.
- c. Predisksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan. Beberapa metode dan teknik yang

digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

2.2.3 Klasifikasi

Classification (Klasifikasi) data merupakan sebuah cara untuk menemukan properti-properti yang sama pada himpunan objek di dalam sebuah data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang di tetapkan. Bertujuan untuk menemukan model dari *training set* yang membedakan kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum di ketahui. Beberapa metode klasifikasi di gunakan dalam *Data mining*.

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu :

- 1) Pendapatan tinggi
- 2) Pendapatan sedang
- 3) Pendapatan rendah

Contoh lain klasifikasi dalam bisnis dan penelitian adalah :

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan penyakit apa yang di alaminya.

2.2.4 Clustering

Clustering (pengklusteran) merupakan pengelompokan record, memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma

pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal. Contoh pengklusteran dalam bisnis dan penelitian adalah :

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

2.2.5 Naïve Bayes

Teorema Bayes dikombinasikan dengan “*Naïve*” yang berarti setiap atribut atau variable bersifat bebas (*independent*). *Naïve Bayes Classifier* dapat dilatih dengan efisien dalam pembelajaran terawasi (*supervised learning*). Klasifikasi memiliki keuntungan hanya butuh sejumlah kecil data pelatihan untuk dapat memperkirakan parameter (varians dan sarana dari variabel) yang diperlukan untuk klasifikasi. Karena variabel yang independen diasumsikan, hanya saja variasi dari variabel pada masing-masing kelas harus ditentukan, bukan semua matriks kovarians (Hikmah & Utammimah, 2017).

Metode *Bayesian Learning* yang paling mudah, cepat, dan sederhana. oleh karena itu berasal dari teorema *Bayes* dan hipotesis kebebasan, menghasilkan *classifier statistic* berdasarkan peluang. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistic yang di kemukakan oleh ilmuwan inggris Thomas Bayes (Idris, 2019). Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Langkah-langkah yang di gunakan untuk perhitungan metode *Naïve Bayes* dibawah ini :

1. Menentukan kriteria-kriteria
2. Membuat tabel aturan

3. Menghitung probabilitas yang sering muncul setiap nilai untuk atribut dari setiap kriteria
4. Menghitung nilai dari setiap kriteria
5. Menghitung nilai probabilitas.

Probabilitas yang didalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

Teorema Bayes berasal dari rumus :

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \dots \dots \dots (1)$$

Dimana $P(A|B)$ artinya peluang A keadaan dan B persamaan (1)

$$P(A \cap B) = P(B|A) P(A) \dots \dots \dots (2)$$

Sehingga di dapatkan teorema Bayes :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots \dots \dots (3)$$

Keterangan :

P : Probabilitas

A : Kriteria

B : Keadaan

Teorema Bayes memiliki rumus :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Keterangan :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

2.2.6 Software Rapidminer

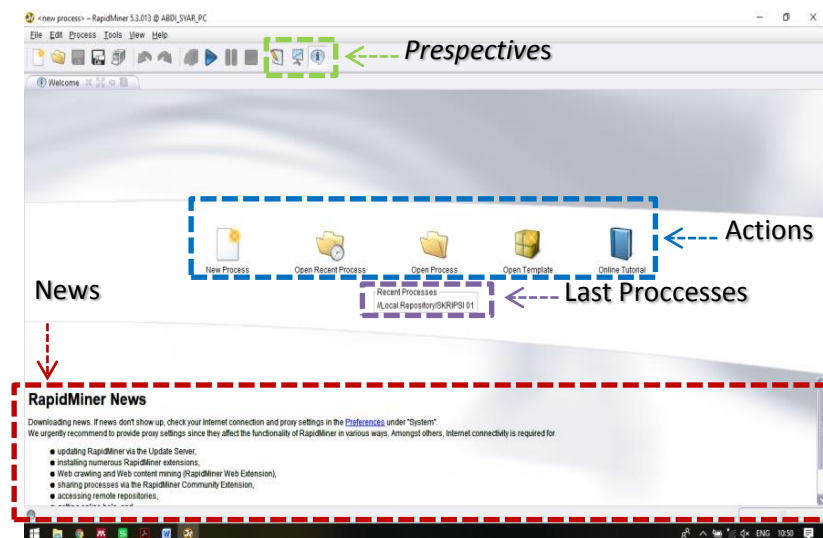
Rapid Miner adalah sebuah *software* yang digunakan dalam teknik yang berada di lingkungan *machine learning*, *data mining*, *text mining* dan *predictive analytics* (Rachman & Wasiyanti, 2019). *Rapid Miner* merupakan *software* atau perangkat lunak untuk pengolahan data dan perangkat lunak sering di sebut *Open Source* yang bisa di gunakan untuk analisa *data mining*. Dengan menggunakan prinsip dan algoritma *data mining*, *RapidMiner* memperoleh pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. *Rapid Miner* memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan *Rapid Miner* pun dapat ditampilkan secara visual dengan grafik. Menjadikan *Rapid Miner* adalah software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining Brilian Rahmat C.T.I. Fitur-fitur yang terdapat dari *Rapid Miner*, sebagai berikut :

- a. Algoritma *Data mining*, seperti *Decision tree* dan *self-organization map*.
- b. Bentuk grafis yang canggih, seperti tumpang tindih diagram *histogram*, *tree chart* dan *3D Scatter plots*.
- c. Variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
- d. Menyediakan prosedur *Data mining* dan *machine learning* termasuk: ETL (*extraction, transformation, Loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi.
- e. Proses *Data mining* tersusun atas *Operator-Operator* yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI.

Untuk pengenalan *interface* RapidMiner sebagai berikut. Tampilan pada RapidMiner dikenal dengan istilah *Perspective*, yang mana RapidMiner terdapat *Perspective Welcome Perspective, Design Perspective* dan *Result Perspective*.

1. *Welcome Perspective*

merupakan tampilan awal ketika baru membuka *software Rapid Miner*, berikut ini:



Gambar 2.1 *View Welcome Perspective*

Berikut ini keterangan bagian dari *Welcome Perspective* sebagai berikut :

I. *Perspectives*

Terdiri dari *ikon-ikon* untuk menampilkan *Perspective* dari *Rapid Miner*

II. *Actions*

Menunjukkan daftar aksi yang dapat anda lakukan pada *Rapid Miner*, yaitu rincian lengkap dari aksi tersebut:

- a. *New*: Aksi tersebut berguna untuk memulai proses analisis baru.
- b. *Open Recyent Process*: Aksi tersebut berguna untuk membuka proses yang baru saja ditutup.
- c. *Open Process*: Aksi tersebut untuk membuka *Repository Browser* yang berisi daftar proses.
- d. *Open Template*: Aksi tersebut menunjukkan pilihan lain yang sudah ditentukan oleh proses analisis.

- e. *Online Tutorial*: Aksi digunakan untuk memulai *Tutorial* secara online (terhubung internet). *Tutorial* yang dapat secara langsung digunakan dengan RapidMiner ini, memberikan pengenalan dan beberapa konsep *Data mining*.

III. *Last Processes*

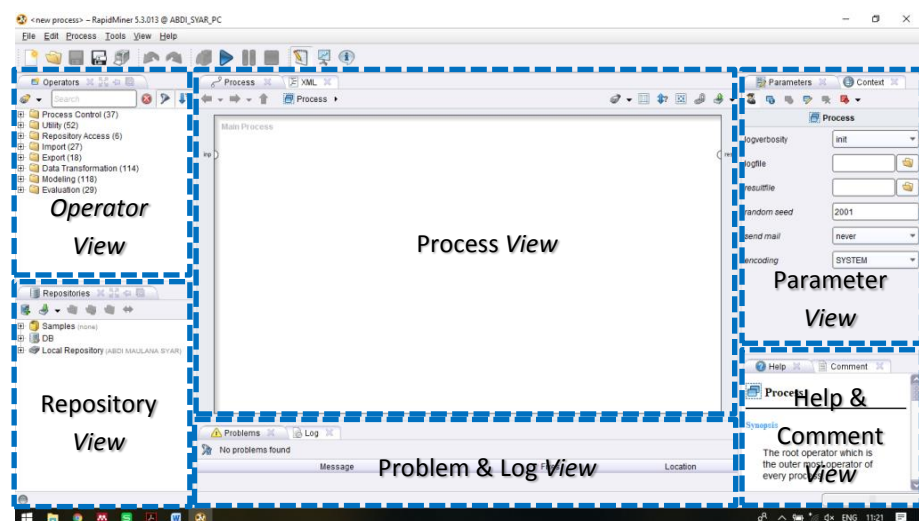
Menampilkan daftar proses analisis yang baru saja dilakukan. Hal ini akan memudahkan Anda jika ingin Melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut.

IV. *News*

Bagian yang akan menampilkan berita terbaru mengenai *RapidMiner*.

2. *Design Perspective*

Merupakan lingkungan kerja *Rapid Miner*. Dimana *Design Perspective* ini merupakan *Perspective* utama dari RapidMiner yang digunakan sebagai area kerja untuk membuat dan mengelola proses analisis. Berikut tampilan dari *Design Perspective*:



Gambar 2.2 *View Design Perspective*

Sebagai lingkungan kerja, *Design Perspective* memiliki beberapa *view*. Berikut ini beberapa *view* yang ditampilkan pada *Design Perspective*:

a) *Operator View*

View yang paling penting pada *Perspective* ini. Semua *Operator* atau langkah kerja dari *Rapid Miner* disajikan dalam bentuk kelompok hierarki di *Operator View* ini sehingga *Operator-Operator* tersebut dapat digunakan pada proses analisis. Hal ini akan memudahkan dalam mencari dan menggunakan *Operator* yang sesuai dengan kebutuhan.

b) *Repository View*

Merupakan komponen utama dalam *Design Perspective* selain *Operator View*. *View* ini dapat Anda gunakan untuk mengelola dan menata proses Analisis Anda menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.

c) *Process View*

Menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Anda dapat menambahkan langkah baru dengan beberapa cara. hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali.

d) *Parameter view*

Beberapa *Operator* dalam *Rapid Miner* membutuhkan satu atau lebih parameter agar dapat difungsi dengan baik.

e) *Help & Comment View*

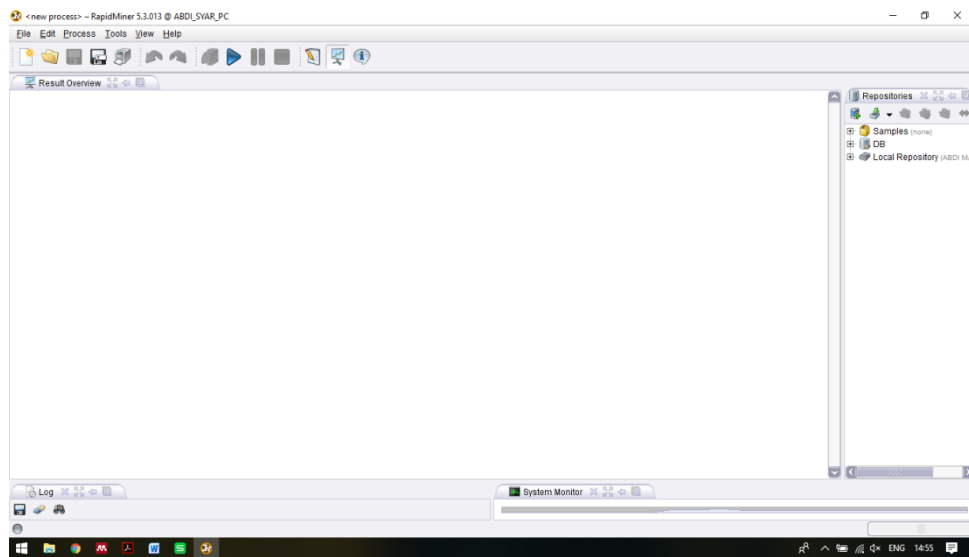
Setiap kali memilih *Operator* pada *Operator View* atau *Process View*, maka jendela bantuan dalam *Help View*.

f) *Problem & Log View*

Komponen yang sangat berharga dan merupakan sumber bantuan selama merancang proses analisis. Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam *Problem View*.

3. *Result perspective*

Merupakan *Perspective* yang menampilkan hasil dari analisa data, adapun tampilan dari *Result Perspective* sebagai berikut:



Gambar 2.3 *View Result Perspective*

2.3 Penelitian Terdahulu

Hasil dari penelitian terdahulu sehingga berkaitan dengan *Data mining* menggunakan *Naïve Bayes* :

Tabel 2.1 Penelitian Terdahulu

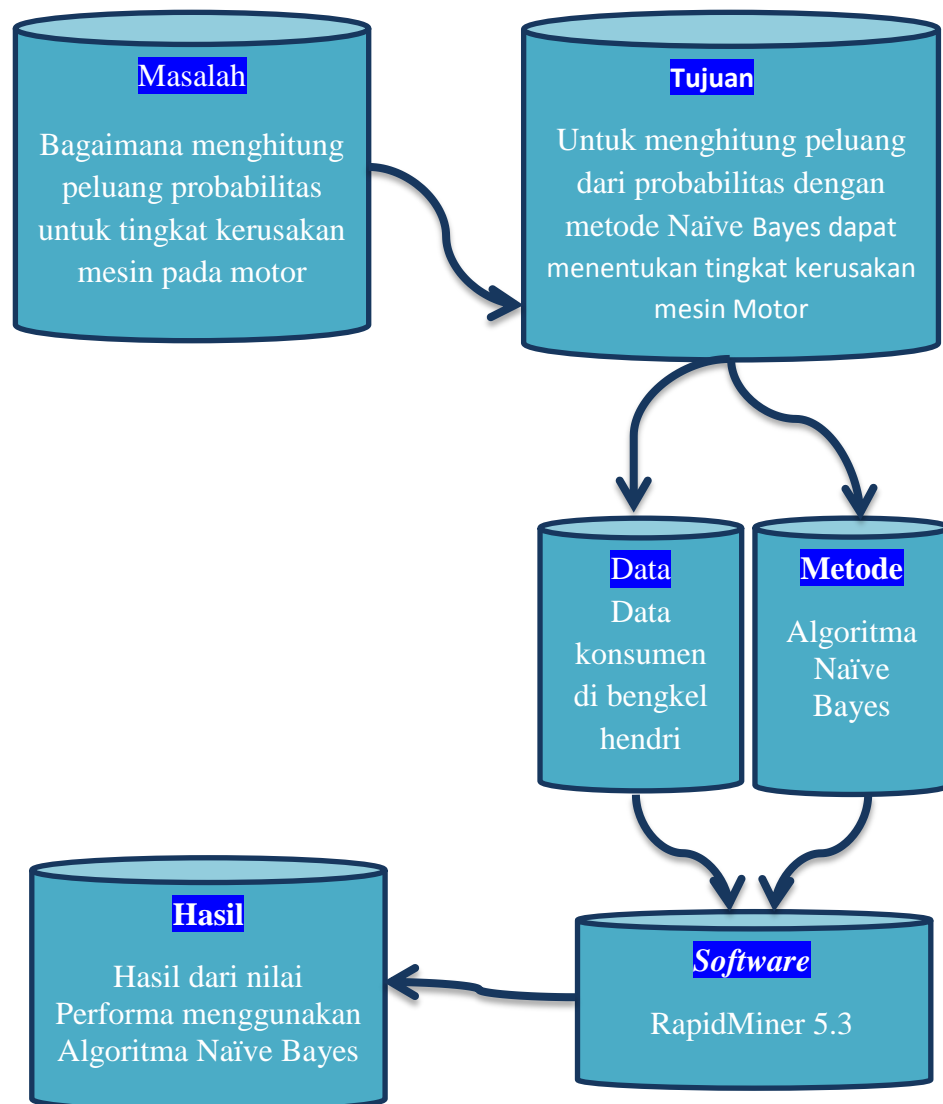
No	Judul	Peneliti	Hasil Penelitian
1	Pemodelan Permasalahan Sparepart Pada Sepeda Motor Yamaha 4 Stroke Injection Dengan Menggunakan Metode Naïve Bayes	(Studi et al., 2016) No ISSN: 2407 – 3911	Pengujian yang telah dilakukan maka hasil probabilitas yang tertinggi yang menjadi dasar rekomendasi keputusan mengenai bermasalah atau tidak bermasalahnya sparepart sepeda motor berdasarkan gejala yang dialami oleh pengguna sepeda motor. Pengujian juga dilakukan terhadap 5 (lima) data pengujian dan diperoleh 80% data sesuai dan 20% data yang tidak sesuai dengan data permasalahan sparepart sebenarnya.
2	Pemanfaatan Metode Naïve Bayes Classifier dalam Pembuatan Sistem Pakar untuk Diagnosa Penyakit Kelamin	(Hikmah & Utammimah, 2017) No ISSN: 2088-4591	Hasil dari penelitian diambil kesimpulan sebagai berikut Sistem mampu mendiagnosa dengan tepat sesuai pendapatan pakar sebenarnya sebesar 95% dari percobaan 10 sampel penyakit kelamin laki-laki dan 10 sampel penyakit kelamin perempuan.

3	Implementasi Data Mining Dengan Algoritma Naïve Bayes Untuk Memprediksi Angka Kelahiran	(Idris, 2019) No. ISSN: 2301-9425	Penentuan data <i>training</i> dapat mempengaruhi hasil pengujian, karena pola dari data <i>training</i> tersebut akan dijadikan sebagai rule untuk menentukan kelas pada data kependudukan. Sehingga tinggi atau rendah tingkat akurasi yang di dapat dipengaruhi oleh penentuan data <i>training</i> .
4	Sistem Pakar Untuk Mendiagnosa Penyakit Pada Tanaman Jagung Dengan Metode Bayes	(SIHOTANG, 2019) No ISSN: 2541-3724	Hasil diagnosa menjadi sangat <i>valid</i> , maka data gejala yang di masukkan oleh seorang administrator / pakar kedalam suatu data penyakit, harus lengkap artinya gejala-gejala yang bisa mengarah ke suatu penyakit tersebut harus di masukkan secara lengkap
5	Sistem Pengecek Kelayakan Pakai Oli Motor Matic Berdasarkan Parameter Warna dan Viskositas Menggunakan Metode Bayes	(Haqiqi et al., 2019) No ISSN : 2548-964X	Akurasi yang diperoleh sistem pengecek kelayakan oli yang diuji dengan jumlah data latih sebanyak 35 data dan data uji sebanyak 18 data adalah senilai 94.44% sehingga dapat disimpulkan sistem ini memiliki akurasi yang cukup tinggi dalam mengklasifikasitingkat kelayakan pada oli.

6	Implementasi Algoritma <i>Naive Bayes</i> Pada Data Set Hepatitis Menggunakan <i>Rapid Miner</i>	(Rachman & Wasiyanti, 2019) No. ISSN : 2579-3500	Metode <i>Naive Bayes</i> termasuk algoritma yang akurat untuk memprediksi karena hasil akurasi menggunakan <i>Rapid Miner</i> menunjukkan lebih dari 50% yaitu sebesar 76.77%. Dengan hasil <i>Class Precision</i> yang tertinggi sebesar 98.88% untuk prediksi " <i>Life</i> ", dan <i>Class Recall</i> sebesar 96.88% untuk Prediksi " <i>Die</i> ".
---	--	--	---

2.4 Kerangka Pemikiran

Kerangka pemikiran adalah gambaran diagram yang di jelaskan menggunakan gambar dan memiliki garis besar bagaimana logika yang berjalan disebuah penelitian. Kerangka pemikiran pada gambar 2.5 sebagai berikut ini.



Gambar 2.4 Kerangka Pemikiran

2.5 Hipotesis Penelitian

Hipotesis menjelaskan jawaban sementara terhadap rumusan masalah dari penelitian, karena itu rumusan masalah penelitian telah disusun dalam bentuk sebuah pertanyaan-pertanyaan yang tertuju pada kerangka pemikiran. Dikatakan sementara karena jawaban yang diberikan baru didasarkan pada teori yang relevan, belum didasarkan pada fakta-fakta empiris yang diperoleh melalui pengumpulan data.

Hipotesis penelitian didasarkan pada kerangka pemikiran di atas sebagai berikut:

1. Di duga *Data mining* cara algoritma Naive Bayes dapat mengeluarkan informasi dari data konsumen *service*.
2. Diduga Peluang yang terbentuk dari algoritma Naive Bayes dapat membantu pemilik bengkel dalam penentuan jenis *service*.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Pada desain penelitian ini telah di uraikan, desain penelitian adalah langkah-langka untuk dilakukannya penyelesaian masalah yang akan di bahas. Adapun desain penelitian ini dapat di lihat pada gambar 3.1



Gambar 3.1 *Design Penelitian*

Berdasarkan desain penelitian yang tertuju pada gambar 3.1 maka langkahnya di uraikan yaitu :

1. Mendeskripsikan Masalah

Dengan menentukan dan mendefinisikan batasan masalah yang akan di teliti, sehingga membantu dalam mendapatkan suatu solusi yang terbaik dari masalah tersebut. Jadi, langkah pertama ini adalah langkah awal yang terpenting dalam penelitian ini.

2. Analisa Masalah

Langkah analisis untuk dapat memahami masalah yang telah di tentukan ruang lingkup atau batasannya. Dengan menganalisis masalah yang telah di tentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik.

3. Mempelajari Literatur

Untuk mencapai tujuan, maka di pelajari beberapa literature-literatur yang di perkirakan dapat digunakan. Kemudian literature-literatur di pelajari tersebut di seleksi untuk dapat ditentukan literature-literatur mana yang akan digunakan dalam penelitian.

4. Mengumpulkan Data

Melakukan observasi yaitu pengamatan secara langsung di bengkel, kemudian di lakukan wawancara untuk mendapatkan informasi atau data yang di butuhkan. Selain itu di lakukan studi kepustakaan yaitu dengan membaca jurnal atau buku yang menunjang dalam melakukan analisis terhadap data dan informasi yang di dapat. Analisis Teknik Pengolahan Data menggunakan Naive Bayes data yang di peroleh dari tempat penelitian selanjutnya melakukan perancangan Algoritma *Naive Bayes*.

5. Perancangan Algoritma *Naive Bayes*

Pada proses dilakukan perancangan dari pola sistem dengan algoritma Naive Bayes sehingga membentuk Peluang probabilitas di bengkel Hendri Motor.

6. Implementasi Algoritma *Naive Bayes*

Membandingkan dokumen data latih dan dokumen data uji. Keduanya dibandingkan melalui beberapa tahap persamaan, yang akhirnya diperoleh hasil probabilitas tertinggi yang ditetapkan sebagai kategori dokumen baru.

7. Pengujian hasil

Pada pengujian hasil, penulis melakukan pengujian dan hasil perancangan menggunakan *software Data mining Rapid Miner*. Sistem di ujinya Data akan membuat pemodelan yang tersembunyi.

3.2 Lokasi Penelitian

Lokasi Penelitian ini beralamat di piayu Bengkel Perum. Bida Ayu, Jl. S. Parman, Mangsang, Kec. Sei Beduk, Kota Batam, Kepulauan Riau 29444.

3.3 Populasi dan Sampel

3.3.1 Populasi

Populasi yang peneliti gunakan pada penelitian ini yaitu data konsumen yang *service* di bengkel Hendri Motor dalam periode *service* 01 Mei 2020 hingga 31 Juni 2020.

3.3.2 Sampel

Sampel yang akan digunakan dalam penelitian sebanyak 150 data yang diambil dari data *service* dalam waktu periode 01 Mei 2020 hingga 31 Juni 2020 yang sudah dipilih-pilih dengan variabel yang akan disesuaikan.

3.4 Variabel Penelitian

Variabel penelitian ini sebagai berikut :

a. Variabel Dependen (Y)

Yaitu merupakan jenis variabel terikat yang digunakan oleh peneliti adalah data konsumen untuk melakukan *Service*.

b. Variabel Independen (X)

Yaitu merupakan variabel bebas yang tidak tergantung pada nilai variabel lainnya yaitu:

- Merk Motor
- Masa *Service*
- Masalah Ringan
- Masalah Sedang
- Masalah Berat

Variabel yang menjadi keputusan adalah Ringan, Sedang, dan Berat.

3.5 Teknik Pengumpulan Data

Pengumpulan data dapat dilakukan dengan berbagai cara terlihat di bagian bawah ini :

1. Observasi

Observasi adalah seseorang yang mengamati dan mendokumentasikan hal yang terjadi selama tindakan berlangsung saat itu juga.

2. Wawancara

Peneliti menanyakan secara langsung atau bertatap muka kepada pihak pengelola bengkel Hendri Motor permasalahan yang sering terjadi.

3. Kuesioner Pustaka

Pengumpulan informasi-informasi yang penulis pelajari dari seseorang mulai dari keyakinan, perilaku, sikap dan karakteristik beberapa orang-orang didalam ruang lingkup organisasi yang mengacu dari permasalahan di penelitian.

4. Studi Pustaka

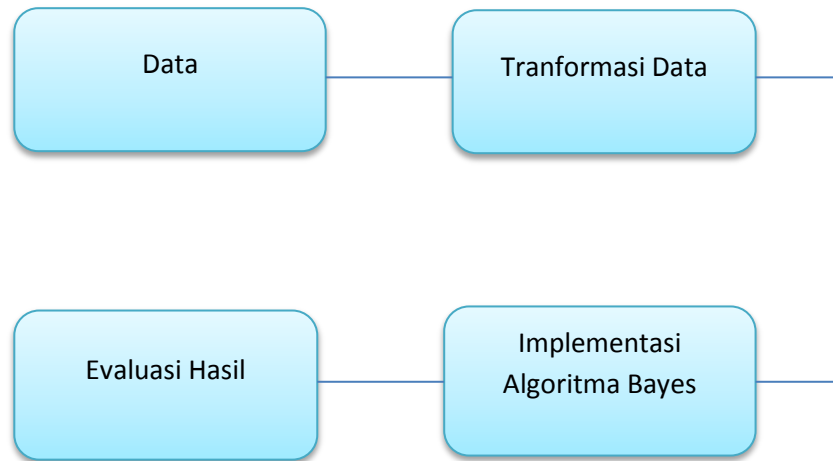
Cara mencari berbagai referensi pendukung penelitian berupa sebuah buku-buka di perpustakaan dan jurnal-jurnal yang terdapat pada *google scholar* yang berkaitan berdasarkan permasalahan yang di teliti.

3.6 Operasional Variabel

Adapun variabel diperlukan oleh penelitian agar bertujuan untuk menemukan peluang untuk melihat tingkat kerusakan mesin motor pola *service* ringan maupun berat ke dalam klasifikasi *service* dalam kategori *service* ringan, *service* sedang atau *service* berat.

3.7 Model Penelitian

Gambar di bawah ini menunjukkan model penelitian yang dilaksanakan peneliti



Gambar 3.2 Model Penelitian