

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 di wajibkan 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 dikelompokan 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. Prediski persentase kanaikan 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 :

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

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

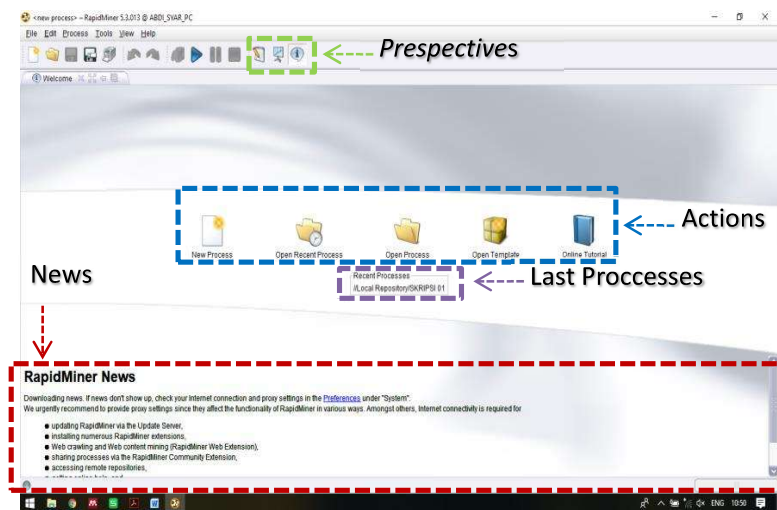
Rapid Miner adalah sebuah *software* yang digunakan dalam teknik yang berada di lingkungan *machine learning*, *data mining*, *text mining* dan *predictive analytics* (Rachman & Wasiyanti, 2019). *Rapid Miner* merupakan *software* atau perangkat lunak untuk pengolahan data dan perangkat lunak sering 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 Recent 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*

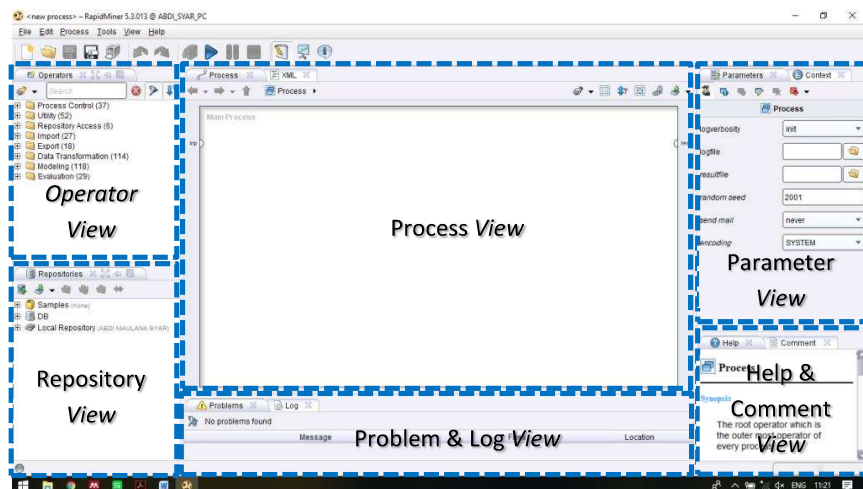
Menampilkan daftar proses analisis yang baru saja dilakukan. Hal ini akan memudahkan Anda jika ingin Melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut.

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 Prespective*

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*

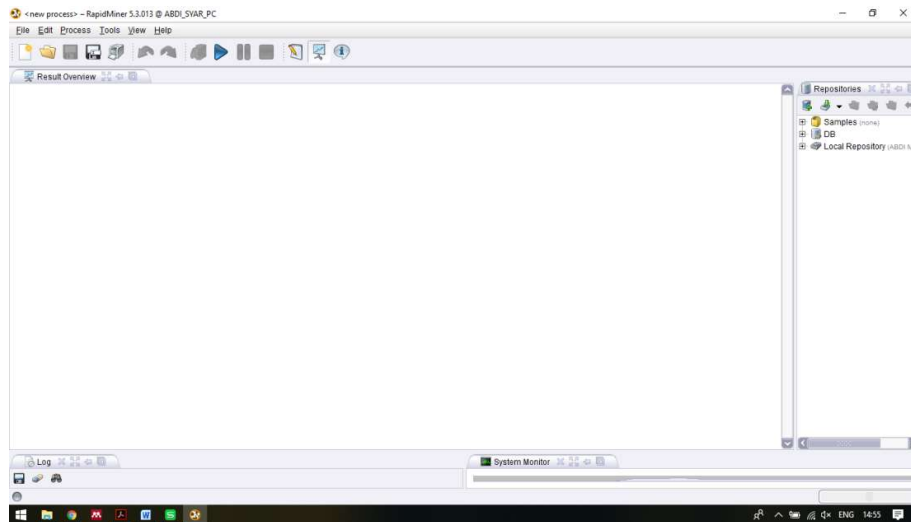
Setiap kali memilih *Operator* pada *Operator View* atau *Process View*, maka jendela bantuan dalam *Help View*.

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

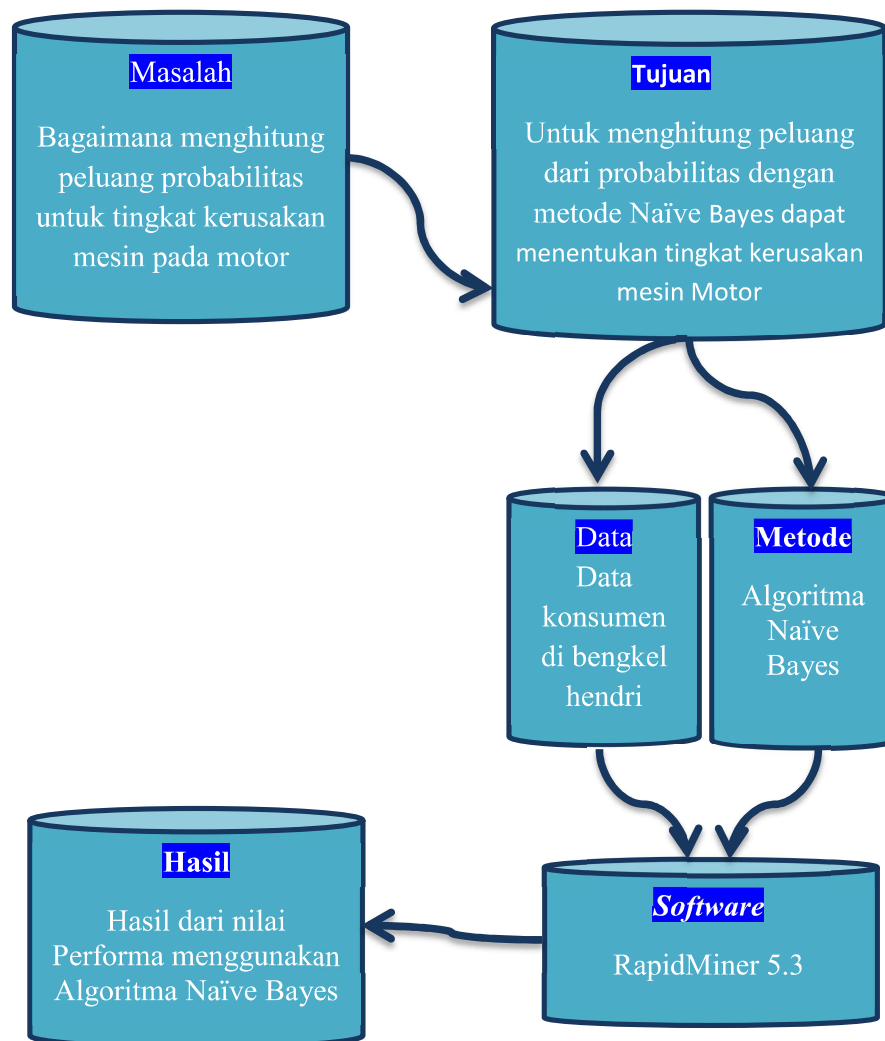
No	Judul	Peneliti	Hasil Penelitian
1	Pemodelan Permasalahan Sparepart Pada Sepeda Motor Yamaha 4 Stroke Injection Dengan Menggunakan Metode Naïve Bayes	(Studi et al., 2016) No ISSN: 2407 – 3911	Pengujian yang telah dilakukan maka hasil probabilitas yang tertinggi yang menjadi dasar rekomendasi keputusan 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>Naïve 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*.