

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terdahulu ini dilakukan guna melihat penelitian yang telah dilakukan sebelumnya. Penelitian terdahulu dapat dilihat pada tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

Peneliti	Judul	Pembahasan
Pikir Claudia Septiani Gulo, (Program Studi Sistem Informasi Jurusan Teknik Elektro dan Informatika Fakultas Sains dan Teknologi, 2021)	Analisi Sentimen Kuliah Online Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes	Twitter adalah salah satu media sosial dengan pengguna lebih dari 29,5 juta di Indonesia dan 383 juta tweet per hari. Peneliti melakukan analisis sentimen menggunakan algoritma Naive Bayes terhadap hasil pencarian tweets dengan kata kunci #kuliahonline selama pandemi Covid-19. Didapatkan sebanyak 2000 data tweet yang akan diklasifikasi dengan menggunakan algoritma Naïve Bayes. Naive Bayes adalah teknik text mining untuk membangun model pengklasifikasi yang sederhana tetapi memiliki keakuratan yang tinggi dalam mengklasifikasi. Algoritma Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat dipublikasikan kedalam basis data dengan jumlah yang besar. Didapatkan hasil nilai akurasi sebesar 82,75% dengan nilai Precision sebesar 82%, nilai Recall sebesar 83% dan nilai F1-Score sebesar 82%.
Ami Natuzzuhriyyah, dkk (Program studi Teknik Informatika, Fakultas Ilmu Komputer,	Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara daring Menggunakan	Semenjak penyebaran Covid-19 di Indonesia, pada awal Maret 2020 menyebabkan aktivitas lembaga pendidikan terganggu. Sebagai gantinya kegiatan pembelajaran konvensional menjadi daring. Pembelajaran daring di Universitas

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

Peneliti	Judul	Pembahasan
Universitas Singaperbangsa, 2021)	Algoritma Naïve Bayes	Singaperbangsa dimulai sejak adanya peraturan dari Kemdikbud RI, dari pembelajaran secara daring tersebut mempengaruhi konsentrasi, kendala seperti signal, suasana pembelajaran. Penelitian ini bertujuan untuk mengetahui tingkat kepuasan mahasiswa terhadap pembelajaran daring menggunakan algoritma Naïve Bayes dengan menggunakan tools RapidMiner dengan hasil yang didapatkan yaitu tingkat akurasi sebesar 76,92%, class precission 100.00%, class recall 57.14% serta nilai AUC 0.881 atau mendekati angka 1, jadi model yang dihasilkan baik. Dengan kata lain, dari hasil yang didapatkan menggunakan algoritma Naïve Bayes dapat digunakan sebagai bahan untuk pengambilan keputusan tingkat kepuasan pembelajaran secara daring.

Berdasarkan tabel 2.1, diketahui pada penelitian Pikir Claudia Septiani Gulo, dkk (2021) Algoritma Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat dipublikasikan kedalam basis data dengan jumlah yang besar. Didapatkan hasil nilai akurasi sebesar 82,75% dengan nilai Precision sebesar 82%, nilai Recall sebesar 83% dan nilai F1-Score sebesar 82%. Ami Natuzzuhriyyah, dkk (2021) Metode Naïve Bayes dengan menggunakan tools rapidminer menghasilkan tingkat akurasi sebesar 76,92%, %, class precission

100.00%, class recall 57.14% serta nilai AUC 0.881 atau mendekati angka 1, jadi model yang dihasilkan baik.

2.2 Landasan Teori

Kajian Teoritis Membahas mengenai definisi analisis, klasifikasi, pembelajaran daring, media sosial, Instagram, *export comments*, *crawler*, komentar, *text mining*, *data mining*, *naive bayes*, , RapidMiner, *Word Cloud*, *flowchart* dan kerangka berpikir yang digunakan untuk penelitian ini.

2.2.1 Analisis

Pengertian analisis yaitu penjabaran dari suatu sistem informasi yang utuh ke dalam berbagai macam bagian komponennya dengan maksud agar kita dapat mengidentifikasi atau mengevaluasi berbagai macam masalah yang akan timbul pada sistem, sehingga masalah tersebut dapat ditanggulangi, diperbaiki atau juga dilakukan pengembangan. Kata Analisis sendiri berasal dari kata analisa, dimana penggunaan pada kata ini mempunyai arti kata yang berbeda tergantung bagaimana kita meletakkan kata ini.

Jadi secara umum, pengertian analisis adalah aktivitas yang terdiri dari serangkaian kegiatan seperti; mengurai, membedakan, dan memilah sesuatu untuk dikelompokkan kembali menurut kriteria tertentu dan kemudian dicari kaitannya lalu ditafsirkan maknanya.

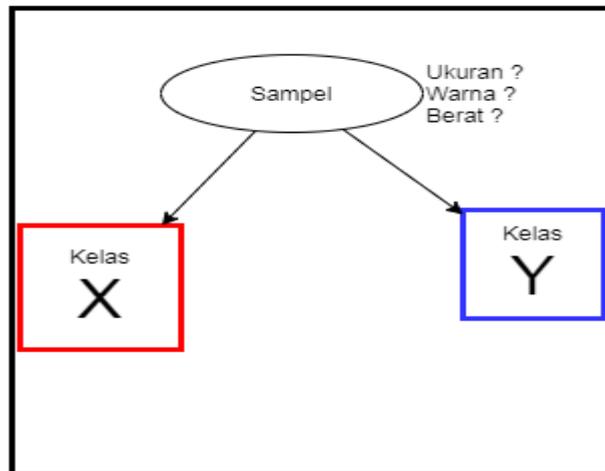
Menurut Sugiono (2015), Analisis adalah kegiatan untuk mencari pola, atau cara berpikir yang berkaitan dengan pengujian secara sistematis terhadap sesuatu untuk menentukan bagian, hubungan antar bagian, serta hubungannya

dengan keseluruhan. Menurut Satori dan Komariyah (2014), Definisi Analisis adalah usaha untuk mengurai suatu masalah menjadi bagian-bagian. Sehingga, susunan tersebut tampak jelas dan kemudian bisa ditangkap maknanya atau dimengerti duduk perkaranya.

Pola standarisasi pada dasarnya subjektif, setiap orang menafsirkan sikap teks yang sama secara berbeda. Itu tergantung pada moral, nilai dan kepercayaan pribadi seseorang. Kata/kalimat positif memiliki elemen positif yang menyertainya (Misalnya, ketika beberapa teks menunjukkan kebahagiaan, antusiasme, kebaikan, dll., Umumnya diklasifikasikan sebagai memiliki kata/kalimat positif). Demikian pula, kalimat negatif memiliki elemen negatif yang menyertainya (Misalnya, ketika beberapa teks menunjukkan kesedihan, kebencian, kekerasan, diskriminasi, dll., Kalimat tersebut umumnya diklasifikasikan sebagai memiliki kata/kalimat negatif). Jika tidak ada emosi yang tersirat, mereka diklasifikasikan sebagai netral.

2.2.2 Klasifikasi

Menurut Mulyanto (2011) Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu obyek. Pada gambar 2.1 dapat dilihat ilustrasi dalam melakukan klasifikasi. Kelas yang ada tentulah lebih dari satu. Penentuan kelas dari suatu dokumen dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di kelas yang satu dengan nilai probabilitas suatu sampel berada di kelas yang lain.



Gambar 2.1 Ilustrasi proses klasifikasi
Sumber : Agus Mulyanto (2011)

2.2.3 Pembelajaran Daring

Pembelajaran daring atau yang dikenal dengan istilah *E-learning* merupakan pembelajaran yang memanfaatkan teknologi. Menurut Dimiyati (2017) menjelaskan bahwa pembelajaran daring atau yang dikenal dengan istilah elearning merupakan bentuk pemanfaatan teknologi dalam mendukung proses belajar mengajar jarak jauh. Hal ini meningkat perubahan gaya belajar yang semakin pesat. Pengertian *e-learning* menurut (Mutia, 2013) dalam jurnalnya mengatakan bahwa *e-learning* berasal dari dua kata yakni “e” dan “learning”. “e” merupakan singkatan dari electronic dan learning adalah pembelajaran. Jadi *e-learning* merupakan pembelajaran yang memanfaatkan teknologi dengan menggunakan media elektronik berupa komputer, laptop maupun handphone selama pembelajaran berlangsung.

Menurut Rosenberg dalam jurnal (Ucu dkk., 2018) *e-learning* merujuk pada penggunaan teknologi internet dalam mengirimkan serangkaian solusi yang mampu meningkatkan pengetahuan dan keterampilannya. Sedangkan menurut

Sutabri dalam jurnal (Agusli & Azianah, 2014) menyatakan bahwa *e-learning* merupakan cara terbaru dalam proses belajar mengajar, *e-learning* lebih merujuk kepada pembelajaran yang di dukung melalui web sehingga dapat dilakukan didalam kelas sebagai pendukung pengajaran tradisional, dalam mengakses *e-learning* dapat dilakukan baik itu di rumah atau di dalam ruang kelas, juga dapat dilakukan dalam ruang kelas virtual, dimana semua kegiatan dilakukan online dan pelaksanaan kelas tidak melakukannya secara fisik langsung.

2.2.4 Media Sosial

Menurut Kaplan & Haenlein (2011) Media sosial adalah sekumpulan aplikasi berbasis internet, beralaskan pada ideologi dan teknologi Web 2.0 sehingga memungkinkan penciptaan dan pertukaran konten oleh penggunanya. Menurut Perrin (2015) waktu rerata yang dihabiskan setiap individu untuk menggunakan media sosial semakin meningkat dari tahun ke tahun. Hampir dua pertiga masyarakat dewasa di Amerika menggunakan lebih dari satu media sosial.

2.2.5 Instagram

Instagram adalah sebuah aplikasi yang digunakan untuk membagi-bagikan foto dan vidio. Instagram sendiri masih merupakan bagian dari Instagram yang memungkinkan teman Instagram itu mengikuti kita dalam akun sosial media instagram. Makin populernya instagram sebagai aplikasi yang digunakan untuk membagi foto mengakibatkan banyak pengguna yang terjun ke ranas bisnis seperti akun sosial bisnis yang turut mempromosikan produk-produknya lewat Instagram(M Nisrina, 2015).

Menurut Atmoko (2012) instagram adalah layanan jejaring sosial berbasis fotografi. Jejaring sosial ini diresmikan pada tanggal 6 Oktober 2010 oleh Kevin Systrom dan Mike Krieger yang mampu menjaring 25 ribu pengguna di hari pertama. nama instagram merupakan kependekan dari kata “instan-telegram”. Jadi bila dilihat dari perpaduan dua kata “insta” dan “gram”, instagram berarti kemudahan dalam mengambil serta melihat foto yang kemudian dapat dikirimkan atau dibagikan kepada orang lain.

2.2.6 Komentar

Komentar menurut Kamus Besar Bahasa Indonesia (KBBI) berarti ulasan atau tanggapan atas berita, pidato, dan sebagainya (untuk menerangkan atau menjelaskan).

2.2.7 Google Chrome

Menurut McFedries (2013) Google Chrome merupakan mesin pencarian mampu melakukan penelusuran dalam waktu kurang dari beberapa detik dengan perangkat lunak yang telah diinstal ke dalam Sistem Operasi windows untuk memberikan pengguna aksesoris pendukung seperti mediator layanan *browser*, *file manager*, *downloader* dan lain-lain. Kombinasi tampilan teknologi canggih membuat Google Chrome menjadi sangat diminati pengguna sebagai *browser* canggih yang dimanfaatkan saat ini. Sebagai salah satu layanan *software* yang memungkinkan pengguna *website* menelusuri informasi, media video dan audio, serta data teknis Google Chrome tersedia dan sangat mendukung untuk semua Operasi Sistem Desktop hingga pengguna *smartphone* seperti Android dan Apple

agar *browser* menjadi terkendali untuk diterima, ditelusuri, disimpan hingga digunakan sebaikbaiknya dalam dunia maya.

2.2.8 *Export Comments*

Export Comments adalah salah satu *website* yang berguna untuk mengambil data pada sosial media seperti Instagram, youtube, Instagram, twitter dll. Data diambil dari komentar pada sosial media tersebut, yang bisa di download dalam format excel dan CSV.

2.2.9 *Crawler*

Menurut Kobayashi (2000), *Web Crawler* merupakan sebuah program/*script* yang bisa memproses halaman web secara otomatis. *Web crawler* sering juga disebut dengan *web robot* atau *web spider*, ide dasarnya sebenarnya sederhana dan mirip dengan saat kita menjelajahi halaman *web* secara manual dengan memakai *browser*. Dimulai dengan sebuah *link* alamat *website* yang dibuka pada *browser*, lalu *browser* melakukan *request* dan mengunduh data dari *web server* melalui protokol HTTP.

Setiap *Hyperlink* yang ditemui pada konten yang tampil akan dibuka lagi pada *windows/tab browser* yang baru, demikian proses terus berulang. Sebuah *crawler* akan mengotomatisasikan pekerjaan itu.

Kesimpulannya, dua fungsi utama *web crawler* merupakan:

1. Mengidentifikasi *Hyperlink*

Hyperlink yang ditemui pada konten web akan ditambahkan pada daftar kunjungan, disebut juga dengan istilah *frontier*.

2. Melakukan proses kunjungan secara rekursif.

Dari berupa *hyperlink*, *crawler* akan menjelajahnya dan melakukan proses berulang, dengan ketentuan yang disesuaikan dengan keperluan aplikasi.

Khusus untuk proses *hyperlink* pengunjung, dapat terjadi *spider trap*, yaitu proses berulang tanpa akhir sebab *crawler* terperangkap untuk terus melakukan pencarian dalam jumlah tak terbatas. Hal ini bisa terjadi bagus secara sengaja maupun tak sengaja.

Kesengajaan dapat saja terjadi apabila *websie* memang dirancang untuk melumpuhkan *crawler*, misalnya dengan memnbuat halaman dinamis dengan angka tak terhingga.

Sedangkan ketidaksengajaan dapat terjadi sebab terbiasa kesalahan pada desain program *crawler* sehingga membaca ulang *hyperlink* yang telah diakses, atau sebuah *website* secara tak sengaja mempunyai halaman dinamis yang tak terbatas, contohnya halaman dinamis yang tercipta berdasarkan tanggal kalender.

2.2.10 Polarity

Menurut Kamus Besar Bahasa Indonesia (KBBI) polaritas (*polarity*) adalah hal atau benda yang memperhatikan dua sifat yang berlawanan atau hal yang bertentangan langsung. *Polarity* digunakan untuk melihat seberapa positif atau negatif sebuah teks. Semakin tinggi *polarity*, itu menandakan *feel/emotion* yang semakin positif juga dari sebuah text.

2.2.11 Tahapan Text Mining

Seperti pada jurnal yang ditulis oleh Kumar & Bhatia (2013) *text mining* memiliki area diantaranya yaitu *data mining*, *information extraction*, *informatiion*

retrieval, dan *data mining*. Beberapa tahapan yang dilakukan pada proses ini adalah:

2.2.11.1 Text Preprocessing

Menurut Ratniasih, Sudarma, & Gunantara (2017) tahap ini melakukan analisis semantik (kebenaran arti) dan sintaktik (kebenaran susunan) terhadap teks. Tujuan dari pemrosesan awal adalah untuk mempersiapkan teks menjadi data yang akan mengalami pengelolaan lebih lanjut. Beberapa tahapan pada proses ini yaitu *filtering*, *stemming*, *tagging*, dan *analyzing*. Tahapan tersebut dijabarkan oleh. Proses *part-of-speech* melakukan *parsing* terhadap seluruh kalimat teks kemudian memberikan peran kepada setiap kata, misalnya : petani (subyek) pergi (predikat) ke (kata hubung) sawah (keterangan). Hasil dari *part-of speech tagging* bisa dipakai untuk *parse tree*, di mana masing-masing kalimat berdiri sebagai sebuah pohon mandiri.

Untuk proses *parsing* sederhana tak dibangun *parse tree* seperti cara sebelumnya. Pada proses *parsing* sederhana sistem akan memecah teks menjadi sekumpulan kata-kata, yang kemudian akan dibawa sebagai input untuk tahap berikutnya pada proses *text mining*.

2.2.11.2 Text Transformation (Attribute Generation)

Transformasi teks atau pembentukan atribut mengacu pada proses untuk mendapatkan representasi dokumen yang diharapkan. Pendekatan representasi dokumen yang lazim digunakan adalah model *bag of words* dan model ruang vektor (*vector space model*). Transformasi teks sekaligus juga melakukan pengubahan kata-kata ke bentuk dasarnya dan pengurangan dimensi kata di dalam

dokumen. Tindakan ini diwujudkan dengan menerapkan *stemming* dan menghapus *stopword*.

Stopword adalah ciri (kata unik) dari suatu dokumen seperti kata sambung, dan kata kepunyaan. Memperhitungkan *stopword* pada transformasi teks akan membuat keseluruhan sistem *text mining* bergantung kepada faktor bahasa. Hal ini menjadi kelemahan dari proses penghilang *stopword*, namun proses penghilang *stopword* tetap dipakai sebab proses ini akan sangat mengurangi beban kerja sistem. Dengan menghilangkan *stopword* dari suatu teks maka sistem hanya akan memperhitungkan kata-kata yang dianggap penting.

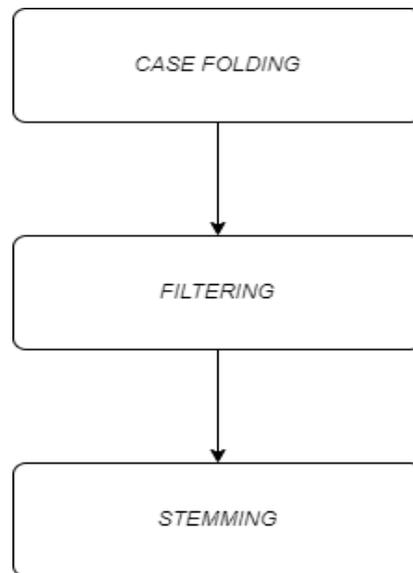
Proses *stemming* dan penghilangan *stopword* bisa dipakai secara terpisah atau bersamaan, di mana dilakukan proses penghilangan *stopword* terlebih dulu yang diikuti dengan proses *stemming*. Hal ini dilakukan untuk menemukan pola dari teks dalam data tersebut.

Gambar 2.2 adalah tahapan yang dilakukan dari pra-proses teks sampai *text transformation* dengan melakukakukan *case folding*, *filtering*, hingga *stemming* yaitu pengubahan karakter huruf menjadi huruf rendah.

1. *Case Folding*

Case Folding dibutuhkan dalam mengkonversi keseluruhan teks di dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau *lowercase*). Sebagai contoh, *user* yang ingin mendapatkan informasi “KOMPUTER” dan mengetik “KOMPUTER”, “KomPUter”, atau “komputer”, tetap diberikan hasil *retrieval* yang sama yakni “komputer”. *Case folding* adalah mengubah semua huruf dalam dokumen menjadi

huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.



Gambar 2.2 Tahap *Text Preprocessing*
 Sumber : Ratniasih, Sudarma, &Gunantara (2017)

2. *Filtering*

Menurut Merina (2016) *Filtering* adalah tahap mengambil kata dari hasil proses token. Bisa menggunakan algoritma *stoplist* atau *word list* yang didapatkan pada jurnal tentang *stop word* bahasa Indonesia dan penambahan istilah yang terdapat pada UMKM sebagai *bag of words*. *Filtering* dapat juga diartikan sebagai proses mengambil kata-kata penting dari hasil proses token atau penghapusan *stopwords*. *Stopwords* merupakan kosa kata yang bukan merupakan cir (kata unik) dari suatu dokumen berbahasa Indonesia secara umum yang dijadikan sebagai *bag of words*. Beberapa kata yang dianggap tidak penting disimpan pada *database* yang nantinya dicocokkan kepada kata yang ada. Untuk contoh tahap *filtering* terlihat pada tabel 2.2.

Tabel 2.2 Tahap *Filtering*

Hasil Case Folding	Hasil Filtering
Terdapat	Kekurangan
Kekurangan	sisi
Pada	pemasaran
Sisi	sertifikat
Pemasaran	halal
Saya	mui
Dikarenakan	
Belum	
Memiliki	
Sertifikat	
Halal	
Dari	
MUI	

Sumber : Merina (2016)

3. *Stemming*

Teknik *stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan

hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia semua kata imbuhan baik itu sufiks dan prefiks juga dihilangkan.

Tabel 2.3 Tahap *Stemming*

Hasil <i>Filtering</i>	Hasil <i>Stemming</i>
Membela	Bela
menguatnya	menguat
dikatakan	kata
dibandingkan	dibanding

Sumber : Merina (2016)

2.2.12 *Data Mining*

Menurut Kusrini dan emha Taufiq (2011) *Data mining* adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Secara sederhana *data mining* adalah penambangan atau informasi baru dengan mencari suatu pola atau aturan tertentu dari sejumlah data yang sangat besar (Agus Nursikuwagus, 2017).

Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada *data mining*. Dalam konteks ini *data mining* merupakan satu langkah dari proses KDD.

Definisi umum dari *data mining* itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam *database*, *data warehouse*, atau media penyimpanan informasi yang lain. Hal penting yang terkait di dalam data mining adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan induksi yang bermanfaat.

2.2.13 Naive Bayes

Menurut Saleh (2015) *Naive Bayes* merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naive bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh iluman inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan di masa sebelumnya.

Dikutip dari jurnal Aida Indriani (2014) yang berjudul Klasifikasi Data Forum dengan menggunakan Metode *Naive Bayes Classifier*, klasifikasi terdapat 2 (dua) proses yaitu model pelatihan dan model pengujian. Pada model pelatihan

terdapat suatu kumpulan data yang memiliki kelas yang telah ditetapkan (data latih). Kumpulan data latih digunakan untuk membentuk model klasifikasi. Pada model pengujian, model klasifikasi yang sudah terbentuk diuji dengan data yang lainnya (data uji) untuk mengetahui akurasi dari model klasifikasi tersebut. Model klasifikasi yang dihasilkan dapat digunakan untuk menentukan kelas dari data yang baru.

Data *training* digunakan untuk melatih algoritma, sedangkan data testing dipakai untuk mengetahui performa algoritma yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat sebelumnya. Ini biasanya disebut dengan generalisasi. Hasil dari pelatihan tersebut bisa disebut dengan model.

Menurut Saleh (2015) keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Persamaan dari teorema Bayes dapat dilihat di persamaan 2.1 :

$$P(H/X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Dimana :

X : data dengan *class* yang belum diketahui

H : hipotesis data menggunakan suatu *class* spesifik

$P(H/X)$: Probabilitas hipotesis H berdasarkan kondisi X (parteriori probabilitas)

$P(H)$: probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: probabilitas H

2.2.14 RapidMiner

RapidMiner berkembang sejak tahun 2001, sebelumnya disebut dengan nama YALE (*Yet Another Learning Environment*). *Software* ini dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, serta Simon Fischer pada Unit *Artificial Intelligence* dari *Technical University of Dortmund*.

RapidMiner adalah platform analisis modern yang meliputi *data mining*, *machine learning*, analisis prediktif, *text mining* dan analisis bisnis. *Software* ini digunakan untuk mengukur kinerja algoritma dan untuk menemukan algoritma terbaik yang akan berguna untuk klasifikasi, prediksi dan teknik lainnya di *data mining*. RapidMiner merupakan *software* yang *user friendly* dan memiliki GUI (*Graphic User Interface*) yang efektif yang digunakan untuk bekerja dengan mudah. RapidMiner memberikan *mechine learning* dan data prosedur termasuk *loading data* dan transformasi (*Extract, Transform, Load (ELT)*), data *preprocessing* dan pemodelan statistik, visualisasi dan analisis prediktif, penyebaran dan evaluasi. Terdapat sifat-sifat yang dimiliki oleh RapidMiner, yakni:

1. Penulisan menggunakan bahasa Java. Hal ini memungkinkan RapidMiner bisa berjalan pada sistem operasi yang berbeda-beda.
2. Proses menemukan pengetahuan dituangkan dalam model operator trees.

3. Merepresentasikan XML internal guna memungkinkan format standar pertukaran data.
4. Penggunaan bahasa scripting yang memungkinkan untuk eksperimen dalam skala besar dan pengotomatisasian eksperimen.
5. Konsep *multi-layer* yang menjadikan tampilan data menjadi efisien serta memastikan penanganan data.
6. Mempunyai GUI (*Graphic User Interface*), command line mode, serta Java API yang bisa dipanggil melalui program lain

2.2.14.1 Fitur RapidMiner

Adapun fitur-fitur yang terdapat pada RapidMiner adalah sebagai berikut :

1. Terdapat banyak algoritma *data mining* seperti *Support Vector Machine*, *Naive Bayes* dan *self-organization map*.
2. Bentuk grafis yang handal seperti tumpang tindih diagram histogram, *tree chart* dan *3d Scatter plots*.
3. Memiliki banyak variasi plugin, seperti text plugin yang dapat digunakan untuk melakukan analisis teks.
4. Tersedia prosedur data mining dan *machine learning* termasuk ELT, data *preprocessing*, visualisasi, *modelling* dan evaluasi.
5. Proses data mining disusun berdasarkan operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibangun dengan GUI.
6. Mengintegrasikan proyek *data mining* Weka dan statistik R.

2.2.14.2 Interface RapidMiner

RapidMiner memiliki tampilan antarmuka yang *user friendly* yang memberikan kemudahan kepada pengguna dalam memakainya dan dikenal dengan sebutan *Perspective*. adapun *perspective* yang terdapat pada RapidMiner yakni:

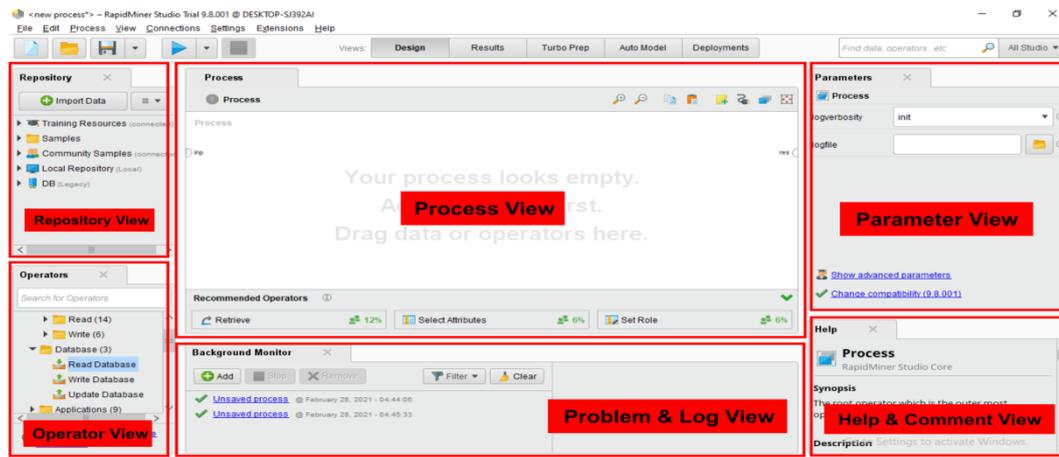
1. *Design Perspective*

Design perspective (Gambar 2.3) merupakan area kerja pada RapidMiner yang berguna untuk membuat serta melakukan pengolahan terhadap proses analisis.

- a. *Operator View*: Menyajikan seluruh operator atau langkah kerja dalam bentuk hierarki agar bisa bekerja pada proses analisis. Operator ini memiliki beberapa kelompok operator, yaitu :

- 1) *Process Control*: Berguna untuk mengatur aliran proses melalui operator perulangan dan percabangan.
- 2) *Utility*: Operator bantuan seperti operator macros, *log in*, subproses, dan sebagainya.
- 3) *Repository Access*: Berguna untuk membaca atau menulis akses pada repository melalui operator-operator yang terdapat didalamnya.
- 4) *Import*: Terdiri dari berbagai operator yang berguna untuk membaca data dan objek dari format tertentu seperti *file*, *database*, dan lain-lain.
- 5) *Export*: Berisi operator yang berguna untuk menulis data dan objek menjadi format tertentu.
- 6) *Data Transformation*: Kelompok ini berisi seluruh operator yang dapat digunakan dalam transformasi data dan meta data.

- 7) *Modelling*: Terdiri dari proses data mining guna menerapkan model yang dihasilkan menjadi set data yang baru.
 - 8) *Evaluation*: Digunakan untuk menghitung kualitas pemodelan dan untuk data baru.
- b. *Repository View*: Merupakan komponen utama pada *Design Perspective* selain *Operator View*. *Repository view* bisa digunakan untuk mengelola sekaligus menata proses analisis menjadi proyek dan sebagai sumber data maupun meta data.
 - c. *Process View*: Menunjukkan langkah-langkah tertentu pada proses analisis serta sebagai penghubung langkah-langkah tersebut yang mana penghubung tersebut dapat dilepas kembali.
 - d. *Parameter View*: Beberapa operator pada RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar.
 - e. *Help & Comment View*: *Help View* menunjukkan penjelasan mengenai operator yang dipilih melalui *Operator View* maupun *Process View*, sedangkan *Comment View* digunakan untuk menuliskan komentar pada langkah-langkah proses tertentu.
 - f. *Problem & Log View*: Setiap peringatan dan pesan kesalahan akan ditampilkan pada *Problem View*.



Gambar 2.3 Tampilan *Design Perspective*

2.2.15 *Word Cloud*

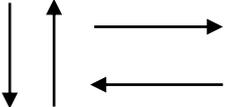
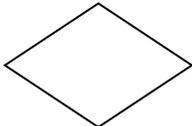
Barret (2011) menyebutkan bahwa visualisasi data mengacu pada penggunaan alat untuk mempresentasikan data dalam bentuk bagan, peta, *tag clouds*, animasi, atau sarana grafis apa pun yang membuat konten lebih mudah dipahami. Jadi, wordle atau *word cloud* adalah untuk merepresentasikan data yang dibentuk oleh kolase kata berwarna yang membuat konten lebih mudah dipahami.

2.2.16 *Flowchart*

Menurut Wibawanto (2017) *Flowchart* adalah suatu bagan dengan simbol-simbol tertentu yang menggambarkan urutan proses secara mendetail dan hubungan antara suatu proses (intruksi) dengan proses lainnya dalam suatu program. Digram alur dapat menunjukkan secara jelas, arus pengendalian suatu algoritma yakni bagaimana melaksanakan suatu kegiatan secara logis dan sistematis. Bagian alir program dibuat menggunakan simbol-simbol seperti pada tabel 2.4.

Tabel 2.4 Simbol-Simbol *Flowchart*

SIMBOL	NAMA	KETERANGAN
--------	------	------------

	<i>Terminator</i>	Digunakan untuk menunjukkan awal dan akhir dari suatu proses.
	<i>Input/Output</i>	Digunakan untuk mewakili data <i>Input/output</i>
	Garis Alir	Digunakan untuk menunjukkan alur dari proses
	Proses	Digunakan untuk mewakili suatu proses
	Persiapan	Digunakan untuk memberi nilai awal suatu besaran
	Penghubung	Digunakan untuk menunjukkan sambungan dari bagan alir yang terputus.
	<i>Decision</i>	Perbandingan pernyataan dan penyeleksian data yang memberikan pulihan untuk langkah selanjutnya

Sumber : Wibawanto (2017)