

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

1.1. Penelitian Sebelumnya

Dalam penyusunan tugas akhir ini, melakukan penelitian terhadap pustaka yang ada berupa karya - karya skripsi maupun jurnal terdahulu yang memiliki relevansi terhadap topik yang diteliti oleh penulis. Berikut adalah karya ilmiah terdahulu yang berkaitan:

Penelitian yang dilakukan oleh Dedi Darwis, Nery Siskawati, Zaenal Abidin yang berjudul “Penerapan Algoritma Naive Bayes untuk *Analisis Sentimen Review Data* Twitter BMKG Nasional” pada tahun 2021. Hasil dari penelitian ini Klasifikasi ini menggunakan algoritma naive bayes. Klasifikasi dapat memberikan kemudahan bagi pengguna untuk melihat opini positif, negatif, dan netral. Hasil uji akurasi pada metode naive bayes untuk klasifikasi yaitu 69.97%.

Penelitian yang dilakukan oleh Tanthy Tawaqalia Widowati yang berjudul “Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma *Naïve Bayes* Dan *Support Vector Machine*” pada tahun 2020. Hasil dari hasil ujicoba diketahui bahwa untuk kasus yang diteliti, metode Naive Bayes menghasilkan kinerja yang lebih baik dengan *accuracy* 91.48%, *precision* 89.28% dan *recall* 91.58%[7].

Penelitian yang dilakukan oleh Kresentia Verena Septiana Toy , Yuita Arum Sari, Imam Cholissodin dengan judul “Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan *Relevance Frequency Feature Selection* (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)” pada tahun 2021. Hasil dari pengujian sebanyak 5 pengujian menggunakan

klasifikasi Naive Bayes, diperoleh rata-rata akurasi sebesar 62,6%, sementara hasil pengujian akurasi klasifikasi dengan penambahan RFFS diperoleh rata-rata akurasi sebesar 65,3% [8].

Penelitian yang dilakukan oleh Delima Ayu Wulandari, Rd. Rohmat Saedudin, Rachmadita Andreswari dengan judul “Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada RUU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma *Naïve Bayes*” pada tahun 2021. Hasil dari Penelitian ini yaitu menggunakan rasio *training* dan *testing* yaitu 75:25 dikarenakan memiliki akurasi yang tertinggi yaitu 88% dibandingkan dengan rasio lainnya. Hasil prediksi yang dihasilkan dari klasifikasi ternyata lebih dominan kepada label positif. Selain melakukan klasifikasi untuk analisis sentimen, penelitian ini juga mengukur performa dari model yang dibuat. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki nilai *precision* 92%, *recall* 84%, dan *f1-score* 86% [9].

Penelitian yang dilakukan oleh Diana Ikasari, Yuliana Fajarwati, Widiastuti dengan judul “Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum MRT Jakarta Menggunakan *Naïve Bayes Classifier*” pada tahun 2020. Hasil dari penelitian yaitu tahap klasifikasi dilakukan setelah melalui fase *preprocessing*, dimana hasil klasifikasi tweet berkecenderungan positif atau negatif, menggunakan metode *Naïve Bayes Classifier*. Akurasi sistem pada analisis sentimen terhadap tweet yang terdapat dalam twitter MRT Jakarta adalah 95.88% [10].

Penelitian yang dilakukan oleh Imam Fahrur Rozi, S.T., M.T, Annisa Taufika Firdausi, S.T., M.T, Khalimatul Islamiyah dengan judul “Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode *Naïve Bayes* Dengan Fitur *N-Gram*” pada tahun 2020. Hasil dari pengujian tersebut diperoleh nilai akurasi untuk *unigram* sebesar 76.67%, 84.44%, 90.00% dan 93.33%. Nilai akurasi untuk *bigram* sebesar 64.17%, 68.89 %, 75.00%, 86.67%. Dari empat pengujian didapatkan hasil akurasi tertinggi pada *unigram*

yaitu sebesar 93.33% dan *bigram* sebesar 86.67%. Jadi untuk nilai akurasi *unigram* lebih tinggi daripada *bigram* [11].

Penelitian yang dilakukan oleh M. Wino Adi Putra, Susanti, Erlin, Herwin dengan judul “Analisis Sentimen Dompot Elektronik Pada Media Sosial Twitter Menggunakan *Naïve Bayes Classifier*” Pada tahun 2020. Hasil penelitian ini menunjukkan bahwa dompet elektronik GoPay lebih banyak dinilai positif oleh pengguna twitter yaitu sebesar 46,67% diikuti oleh Dana sebesar 37,50% dan OVO sebesar 16,67%. Dompot elektronik OVO memiliki nilai negatif yang lebih tinggi yaitu sebesar 63,33% diikuti oleh GoPay sebesar 53,33% dan Dana sebesar 30,00% [12].

Berdasarkan penjelasan dari beberapa penelitian sebelumnya, maka ringkasan penelitian-penelitian tersebut dapat dimuat di dalam tabel 2.1 berikut ini:

Tabel 2.1 Penelitian Sebelumnya

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
1	Muhammad Riefky, Wara Pramesti[2]	2020	Analisis Sentimen terhadap Southeast Asian Games (SEA Games) di Filipina Tahun 2019 berdasarkan Opini Netizen dari Media Sosial Twitter dengan Metode K-nearest Neighbor dan Support Vector Machine	Tujuan dari penelitian ini adalah mendapatkan dan membandingkan hasil ketepatan klasifikasi sentimen pengguna twitter terhadap SEA Games di Filipina tahun 2019 menggunakan k-nearest neighbor dan support vector machine	Metode K-nearest Neighbor dan Support Vector Machine	hasil akurasi terbaik dalam klasifikasi k-nearest neighbor dan support vector machine adalah klasifikasi support vector machine dengan kernel polynomial sebesar 92.96% sehingga prediksi dari klasifikasi Support Vector Machine cenderung negatif.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
2	Tanthy Tawaqalia Widowati[7]	2020	Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naïve Bayes Dan Support Vector Machine	Untuk mengetahui opini masyarakat di twitter terhadap tokoh public yaitu Nadiem Makarim sebagai Menteri Kemendikbud yang baru	Naïve Bayes dan Support Vector Machine	hasil diketahui bahwa untuk kasus yang diteliti, metode Naive Bayes menghasilkan kinerja yang lebih baik dengan accuracy 91.48%, precision 89.28% dan recall 91.58%. Polaritas sentimen opini netizen terhadap Nadiem Makariem adalah dominan positif. Dengan persentase opini positive sebesar 69,72% dan negative sebesar 30,28%.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
3	Kresentia Verena Septiana Toy , Yuita Arum Sari, Imam Cholissodin[8]	2021	Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)	Agar pemerintah dapat mengetahui bagaimana pendapat masyarakat terkait kebijakan New Normal di Indonesia.	Metode Naive Bayes dengan Relevance Frequency Feature Selection	Hasil dari pengujian sebanyak 5 pengujian menggunakan klasifikasi Naive Bayes, diperoleh rata-rata akurasi sebesar 62,6%, sementara hasil pengujian akurasi klasifikasi dengan penambahan RFFS diperoleh rata-rata akurasi sebesar 65,3%.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
4	Delima Ayu Wulandari, Rd. Rohmat Saedudin, Rachmadita Andreswari[9]	2021	Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada RUU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naïve Bayes	Untuk mengetahui kecenderungan komentar atau tweets pengguna Twitter terhadap adanya RUU Cipta Kerja dengan melakukan analisis sentimen.	Metode Naive Bayes	Hasil dari Penelitian ini yaitu menggunakan rasio training dan testing yaitu 75:25 dikarenakan memiliki akurasi yang tertinggi yaitu 88% dibandingkan dengan rasio lainnya. Hasil prediksi yang dihasilkan dari klasifikasi ternyata lebih dominan kepada label positif. Selain melakukan klasifikasi untuk analisis sentimen, penelitian ini juga mengukur performa dari model yang dibuat. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memiliki nilai precision 92%, recall 84%, dan f1-score 86%

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
5	Diana Ikasari, Yuliana Fajarwati, Widiastuti[10]	2020	Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum MRT Jakarta Menggunakan Naïve Bayes Classifier	Untuk melihat kecenderungan respon pengguna MRT Jakarta apakah positif atau negatif berdasarkan hasil tweet dari Twitter MRT Jakarta. Analisis sentimen ini dapat membantu masyarakat Indonesia dalam menentukan pilihan transportasi umum yang nyaman dan aman berdasarkan ulasan transportasi umum dari Twitter oleh pengguna MRT Jakarta.	Naïve Bayes Classifier	Berdasarkan pengujian akurasi, didapatkan hasil akurasi klasifikasi tweet dari sistem analisis sentimen dengan menggunakan Naive Bayes Classifier sebesar 95.88% dengan precision positif sebesar 70% dan precision negatif sebesar 30%.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
6	Imam Fahrur Rozi, S.T., M.T, Annisa Taufika Firdausi, S.T., M.T, Khalimatul Islamiyah[11]	2020	Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram	Dapat mengetahui klasifikasi dari opini masyarakat di twitter mengenai bencana alam sehingga hasil klasifikasi tersebut dapat digunakan untuk memprioritaskan bantuan dengan mengkategorikan setiap hasil klasifikasi	Naïve Bayes Dengan Fitur N-Gram	Hasil dari pengujian tersebut diperoleh nilai akurasi untuk unigram sebesar 76.67%, 84.44%, 90.00% dan 93.33%. Nilai akurasi untuk bigram sebesar 64.17%, 68.89 %, 75.00%, 86.67%. Dari empat pengujian didapatkan hasil akurasi tertinggi pada unigram yaitu sebesar 93.33% dan bigram sebesar 86.67%. Jadi untuk nilai akurasi unigram lebih tinggi daripada bigram

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
7	M. Wino Adi Putra, Susanti, Erlin, Herwin[12]	2020	Analisis Sentimen Dompot Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier	Penelitian ini bertujuan untuk menganalisa tanggapan masyarakat atas hadirnya berbagai aplikasi dompet elektronik di Indonesia dengan mengklasifikasikannya kedalam tiga kelas yakni positif, negatif dan netral dengan menggunakan metode Naïve Bayes	Naïve Bayes	Hasil penelitian ini menunjukkan bahwa dompet elektronik GoPay lebih banyak dinilai positif oleh pengguna twitter yaitu sebesar 46,67% diikuti oleh Dana sebesar 37,50% dan OVO sebesar 16,67%. Dompot elektronik OVO memiliki nilai negatif yang lebih tinggi yaitu sebesar 63,33% diikuti oleh GoPay sebesar 53,33% dan Dana sebesar 30,00%.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
8	La Ode Mohamad Zulfiqar, Nurul Renaningtias and M. Yoka Fathoni[13]	2019	Educational Data Mining in Graduation Rate and Grade Predictions Utilizing Hybrid Decision Tree and Naïve Bayes Classifier	to predict graduation scores and student graduation rates	Utilizing Hybrid Decision Tree and Naïve Bayes Classifier	Objectives on this research achieved. First, we are able to implement the combination of naïve bayes with hybrid decision tree to make predictions of graduation grade and graduation rate. Which is the result, the accuracy of the prediction for graduation rate and graduation grade is 72.73% on the highest value partition, its 90%. secondly, secondly we are proved that, the higher value of the partition that is applied on the collaboration hybrid decision tree and naïve bayes classification the higher consistency value of its accuracy gained.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
9	Siti Khomsah[14]	2020	Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews	The objective of this study is to determine the performance of the Particle Swarm Optimization (PSO) to optimize the Naive Bayes Classifier	Naive Bayes Classifier	PSO is able to optimize the accuracy of Naive Bayes at a value of 89% to 91.76%. PSO performance is determined by the parameters used, especially the number of particles, the number of iterations, and inertial weight. The large number of particles accompanied by an increase in inertial weight can improve accuracy. The number of particles 20-30 has reached optimal accuracy.

No	Peneliti	Tahun	Judul penelitian	Tujuan penelitian	Metode Penelitian	Hasil Penelitian
10	Manjit Singh, Swati Gupta[15]	2020	Sentiment Analysis using Naive Bayes Classifier and Information Gain Feature Selection over Twitter	Tujuan dari penelitian ini adalah untuk meningkatkan akurasi algoritma Naïve Bayes dalam mengklasifikasikan dokumen beserta Metodologi Pengumpulan Informasi.	Naive Bayes Classifier and Information Gain	Akurasi dari mengklasifikasikan dokumen yang dicapai dalam penelitian ini sebesar 88,80% yang sudah sesuai untuk mengevaluasi sentimen.

Pada Tabel 2.1 diatas adalah beberapa penelitian-penelitian terdahulu yang sudah dilakukan berkaitan dengan analisis sentimen dengan menggunakan metode *naïve bayes*. Penelitian-penelitian diatas dijadikan sebagai acuan untuk penelitian ini dan digunakan sebagai landasan teori untuk menunjang penelitian ini.

1.2. Dasar Teori

1.2.1. Sea Games

Southeast Asian Games (SEA Games) merupakan salah satu event olahraga yang diselenggarakan di wilayah Asia Tenggara dan dilaksanakan setiap dua tahun sekali yang melibatkan sebelas negara anggota *Association of South East Asian Nation (ASEAN)*. Negara-negara anggota *Association of South East Asian Nation (ASEAN)* diantaranya adalah Indonesia, Malaysia, Brunei Darussalam, Singapura, Timor Leste, Filipina, Thailand, Laos, Myanmar, Kamboja, dan Vietnam. *Southeast Asian Games (SEA Games)* memiliki banyak tujuan, yaitu untuk mengeratkan kerja sama antar negara anggota *Association of South East Asian Nation (ASEAN)*, menyatukan pemahaman dan mempererat hubungan antar negara anggota *Association of South East Asian Nation (ASEAN)* di kawasan semenanjung Asia Tenggara.

1.2.2. Twitter

Media sosial seperti Twitter adalah salah satu media sosial viral di dunia *cyber*. Twitter didirikan pada 2006 tetapi per 18 Mei 2013 telah mencapai 554.750.000 pengguna, yang 22 kali lebih banyak dari MySpace, media sosial lainnya didirikan tiga tahun lebih awal dari Twitter. Twitter memiliki sifat dan karakteristiknya sendiri, yang mungkin lebih sederhana daripada alat media sosial lainnya. Banyak istilah yang hanya ada di situs simbol burung ini. Pertama, Twitter mengoperasikan layanan informasi digital yang memungkinkan pengguna untuk mengirim dan melacak pesan mikro yang dikenal sebagai (tweet) yang terdiri dari tidak lebih dari 140 karakter. Fasilitas tweet dirancang untuk digunakan pada perangkat seluler dan komputer. Twitter digunakan untuk memposting status apa pun, memposting ulang status pengguna lain (retweet), merespons posting pengguna lain (reply), dan berbagi tautan. *Blogger* juga dapat memiliki tautan ke posting blog mereka yang dikirim melalui Twitter, baik secara *post-by-post* atau secara otomatis, sebagai contoh kasus dengan semua media baru, media lama tidak dilenyapkan tetapi digolongkan dan dipromosikan di Twitter. Untuk memperbarui informasi yang dimiliki oleh pengguna atau pemilik akun mana pun, orang dapat mengikuti (berlangganan) akun Twitter seseorang. Ketika pengguna Twitter telah mengikuti akun seseorang, setiap informasi

yang disiarkan (*posting*) akan muncul di umpan yang diperbarui yang disebut dengan timeline. Jumlah following (jumlah total orang yang di ikuti) dan follower (jumlah total orang yang mengikuti) selalu diperbarui dalam kotak tertentu di halaman profil Twitter[16].

1.2.3. Analisis Sentimen

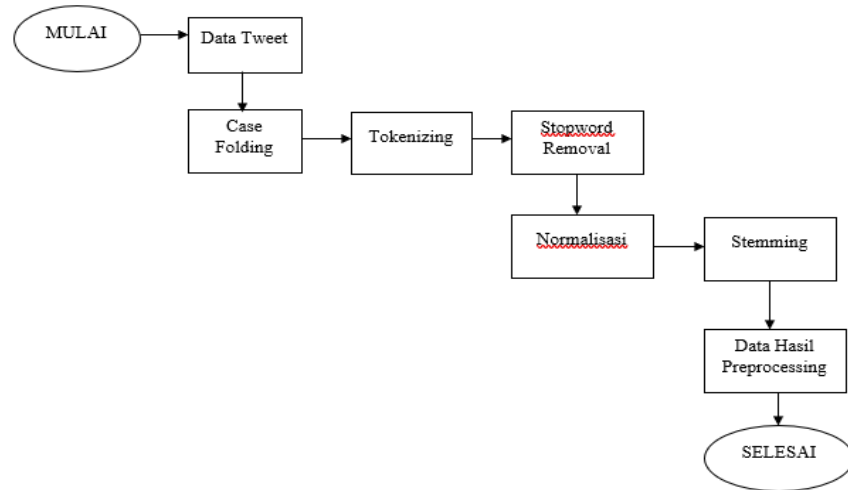
Analisis sentimen adalah sesuatu yang menganalisa pendapat, sentimen, evaluasi, dan emosi orang tentang suatu topik yang yang diekspresikan melalui teks. Tujuan analisis sentimen adalah menemukan pendapat atau ulasan, berdasarkan pendapat orang yang diklasifikasikan pada berbagai polaritas seperti positif, negatif, atau netral. Analisis sentimen juga dapat menyatakan perasaan seperti emosional sedih, gembira atau marah.

2.2.4. Natural Language Processing (NLP)

Natural Language Processing atau sering disingkat *NLP* adalah salah satu bidang ilmu komputer, kecerdasan buatan dan bahasa (linguistik) yang berkaitan dengan interaksi antara komputer dengan bahasa alami manusia. Bahasa alami adalah bahasa yang dapat dipahami manusia. Pada prinsipnya, bahasa alami adalah suatu bentuk informasi yang akan disampaikan dari satu pengguna ke pengguna lainnya. Bahasa alami dapat direpresentasikan dalam bentuk suara atau teks[17].

2.2.5. Pre-Processing

Pre-processing merupakan tahapan awal yang dilakukan dalam penerapan *text mining*. *Pre-processing* pada *Text Mining* bertujuan agar memperoleh informasi menarik berdasarkan data yang tidak terstruktur serta menghapus kata tidak penting pada dokumen. Terdapat beberapa tahapan dalam *Pre-processing* yaitu *tokenization*, *case folding*, *stopwords removal*, dan *stemming*.



Gambar 2.1 Proses *Text Preprocessing*

Tahap *Pre-Processing* yang dilakukan adalah seperti pada Gambar 2.1 :

1. *Case Folding*

Case Folding bertujuan untuk membersihkan data atau menghilangkan berbagai karakter tanda baca yang tidak perlu dalam sebuah text dokumen seperti tanda titik (.), tanda koma (,), tanda seru (!), tanda tanya (?) dan lain-lain.

2. *Tokenizing*

Dalam tahap *tokenizing* ini bertujuan untuk melakukan partisi atau pemisahan setiap kata yang terdapat dalam suatu kalimat. Hal ini dilakukan dengan tujuan mempermudah untuk proses analisis teks selanjutnya. Proses *tokenizing* untuk mempartisi atau pemisahan kata dalam setiap kalimat pada data dilakukan berdasarkan setiap spasi yang ada dalam sebuah kalimat.

3. *Stopword Removal*

Merupakan kumpulan kata kata yang sering muncul dalam satu dokumen. *Stopword* pada umumnya adalah sebuah kata penghubung yang tidak begitu penting, maka *stopword* dapat diabaikan dan tidak ikut dalam proses pengindeksan. *Stopword* adalah kata kata tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*.

Stopword Removal yaitu proses penghapusan kata-kata yang terdapat pada *stop list*. *Stop list* itu sendiri berisi kosakata-kosakata yang bukan merupakan

ciri dari suatu dokumen dan bertujuan agar menghilangkan kata yang tidak berguna.

4. *Stemming*

Stemming adalah tahapan yang juga diperlukan untuk memperkecil jumlah indeks yang berbeda dari satu data sehingga sebuah kata yang memiliki *suffix* maupun *prefix* akan kembali ke bentuk dasarnya. Selain itu juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapatkan imbuhan yang berbeda pula.

2.2.6. **Klasifikasi**

Teknik klasifikasi adalah salah satu dari teknik *data mining* yang termasuk *supervised learning*. *Supervised learning* artinya proses pembentukan sebuah korespondensi menggunakan sebuah *training dataset*. Tujuannya adalah untuk memprediksi target dari beberapa atribut.

1. *Naïve Bayes*

Naive Bayes Classification didasarkan pada teorema *Bayesian*. *Naive Bayes Classification* bekerja dengan cara membagi tiap atribut menjadi atribut independen. Atribut yang ada ditentukan nilai probabilitasnya sebagai penentu dalam keputusan akhir. Atribut yang independen berarti atribut tersebut tidak saling mempengaruhi satu sama lain. Dalam penentuan probabilitas akhir digunakanlah rumus teorema *bayesian*. Berikut merupakan rumus bayes[18]. Adapun, persamaan dari Teorema Bayes dapat dilihat pada persamaan (2.3)

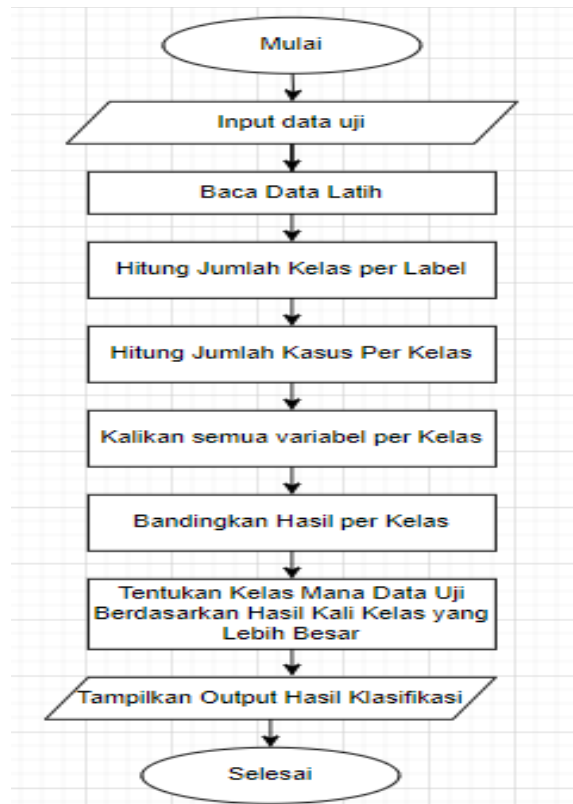
Keterangan:

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

X = Data dengan class (label) yang belum diketahui

H = Hipotesa bahwa X adalah data dengan class (label)

- $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (
- $P(H)$ = Probabilitas hipotesis H (prior prob.)
- $P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut
- $P(X)$ = Probabilitas X



Gambar 2.2 *Flowchart Naïve Bayes*

Setiap alur dari penelitian analisis sentimen yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 2.2. Disana terdapat bebrapa tahap yang dilakukan, dari mulai sampai selesai.

2.2.7. *Lexicon Based*

Lexicon Based adalah suatu proses pemilihan kata penting pada dokumen berdasarkan suatu kamus/leksikon yang sudah ada. Dalam pengaplikasiannya, terdapat dua buah kamus yang digunakan untuk menjadi *wordlist*. Kamus dengan kumpulan kata bersentimen positif dan kamus dengan kumpulan kata bersentimen negatif[19].

2.2.8. Evaluasi

Evaluasi performansi dilakukan untuk mengetahui keakuratan dari Pemodelan yang telah diterapkan pada data latih. Kemudian membandingkan hasil dari dua dataset yang berbeda dengan menerapkan *confusion matrix* untuk menghitung *precision*, *recall*, *f1-score*, dan *accuracy*.

1. Akurasi

Akurasi merupakan persentase dari total sentimen yang benar dikenali. Perhitungan akurasi dilakukan dengan cara membagi jumlah data sentimen yang benar dengan total data dan data uji. Untuk menghitung nilai akurasi nya dilakukan dengan menggunakan Persamaan:

$$Akurasi = \frac{Jumlah\ Sentimen\ Benar}{Jumlah\ Data\ Tes} \times 100 \quad (2.4)$$

2. Precision

Precision merupakan perbandingan jumlah data relevan yang ditemukan terhadap jumlah data yang ditemukan. Perhitungan *precision* dilakukan dengan cara membagi jumlah data benar yang bernilai positif dibagi dengan jumlah data benar yang bernilai positif dan data salah yang bernilai positif. Nilai dari data salah bernilai positif diambil dari jumlah nilai selain *true positive* kolom yang sesuai tiap kelas nya. Untuk menghitung nilai *precision* dapat dilakukan dengan menggunakan Persamaan:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2.5)$$

3. Recall

Recall merupakan perbandingan jumlah materi relevan yang ditemukan terhadap jumlah materi yang relevan. Perhitungan *recall* dilakukan dengan cara membagi data benar bernilai positif dengan hasil penjumlahan dari data benar yang bernilai positif dan data salah yang bernilai negatif. Nilai dari data salah yang bernilai negatif diambil dari jumlah nilai selain *true positive* baris yang sesuai tiap kelas nya. Perhitungan *recall* dapat menggunakan Persamaan:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2.6)$$

4. *F-1 Measure*

F-measure merupakan parameter tunggal ukuran keberhasilan *retrieval* yang menggabungkan *recall* dan *precision*. Nilai *F-measure* hasil perkalian *precision* dan *recall* dibagi *measure* didapat dari perhitungan hasil perkalian *precision* dan *recall* dibagi dengan hasil penjumlahan *precision* dan *recall* kemudian dikalikan dua dan perhitungan *f-measure* menggunakan Persamaan:

$$f - \text{Measure} = 2 * \frac{\text{Precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (2.7)$$

2.2.9. *Python*

Python merupakan bahasa pemrograman yang freeware atau perangkat bebas dalam arti sebenarnya, tidak ada batasan dalam penyalinan atau mendistribusikannya. Lengkap dengan source code, debugger dan profiler, antarmuka yang terkandung di dalamnya untuk pelayanan antarmuka, fungsi sistem, GUI (antarmuka pengguna grafis), dan basis datanya. *Python* adalah suatu dari bahasa pemrograman skrip dalam samaran dari bahasa perograman tradisional[20].