

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

Landasan Teori yang digunakan bersumber dari penelitian terdahulu yang menjadi bahan rujukan penulisan skripsi dengan menganalisis penelitian terdahulu yang relevan dengan penelitian yang sudah dilakukan, ini dapat dilihat sebagai berikut :

Tabel 2.1 *Literatur review*

Penulis	Judul	Metode	Hasil
(Duei Putri et al., 2022)	Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes Classifier</i>	<i>Naive Bayes Classifier</i>	Berdasarkan penelitian menggunakan algoritma <i>Naive Bayes</i> didapatkan accuracy score sebesar 0.8 atau 80% hal ini berarti sistem mampu memprediksi 80% secara akurat dari total data testing sebesar 20%.
(Buntoro, n.d.)	Analisis Sentimen Hatespeech Pada <i>Twitter</i> Dengan Metode <i>Naive Bayes Classifier</i>	<i>Naive Bayes Classifier</i> Dan <i>Support Vector Machine</i>	Nilai akurasi tertinggi didapat dengan metode klasifikasi <i>Support Vector Machine (SVM)</i> dengan

Tabel 2.1 *Literatur review*

(Mas Pintoko & Muslim, n.d.)	<p>Dan <i>Support Vector Machine</i></p> <p>Analisis Sentimen Jasa Transportasi <i>Online</i> pada</p>	<p><i>Naïve Bayes Classifier</i></p>	<p>tokenisasi <i>unigram</i>, <i>stopword list</i> Bahasa Indonesia dan <i>emoticons</i>, dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai <i>recall</i> 66,7% nilai <i>TP rate</i> 66,7% dan nilai <i>TN rate</i> 75,8%. Dalam penelitian ini juga dapat diketahui metode klasifikasi <i>Support Vector Machine (SVM)</i> lebih tinggi akurasinya untuk klasifikasi sentiment <i>tweet</i> HateSpeech Bahasa Indonesia dibandingkan metode klasifikasi <i>Naïve Bayes Classifier (NBC)</i>. Berdasarkan hasil analisis yang dilakukan terkait berdasarkan data dari <i>Twitter</i>.</p>
------------------------------	--	--------------------------------------	---

Tabel 2.1 *Literatur review*

(Mas Pintoko & Muslim, n.d.)	Analisis Sentimen Jasa Transportasi <i>Online</i> pada	<i>Naïve Bayes Classifier</i>	Berdasarkan hasil analisis yang dilakukan terkait berdasarkan data dari <i>Twitter</i> .
	<i>Twitter</i> Menggunakan Metode <i>Naïve Bayes Classifier</i>		sentimen terhadap jasa transportasi <i>online</i> dengan menggunakan metode <i>Naïve Bayes</i> untuk menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil menunjukkan tingkat sentimen positif lebih besar dibandingkan dengan tingkat sentimen negatif. Dengan hasil penelitian ini bisa dilihat bahwa banyak masyarakat yang merasa puas dengan layanan jasa transportasi <i>online</i> di Indonesia

Tabel 2.1 *Literatur review*

(Fajar et al., n.d.)	Implementasi Algoritma <i>Naive Bayes</i> Terhadap Analisis Sentimen Opini Film Pada <i>Twitter</i>	<i>Naïve Bayes Classifier</i>	Dari hasil pengujian dan analisis yang telah dilakukan dapat disimpulkan beberapa hal yaitu pengklasifikasian data opini film berbahasa Indonesia berdasarkan sentimennya dapat dilakukan dengan algoritme <i>Naive Bayes Classifier</i> dengan pembagian <i>datasetnya</i> menggunakan <i>5-fold cross validation</i> . Semakin banyak data <i>training</i> yang digunakan maka akan mempengaruhi kinerja dari sistem. Hasil akurasi akan semakin tinggi dan itu menandakan sistem berhasil melakukan klasifikasi dengan baik. Akurasi tertinggi didapat pada fold kedua yaitu 90%, <i>precision</i> 92%, <i>Recall</i> 90% dan <i>f-measure</i> 90%.
----------------------	---	-------------------------------	--

Tabel 2.1 *Literatur review*

(Al Khadafi et al., 2022)	Penerapan Metode <i>Naïve Bayes Classifier</i> Dan <i>Lexicon Based</i> Untuk Analisis Sentimen <i>Cyberbullying</i> Pada BPJS	<i>Naïve Bayes Classifier</i> Dan <i>Lexicon Based</i>	Untuk pengujian menggunakan metode <i>Confusion matrix</i> dan hasil pada klasifikasi memperlihatkan bahwa metode <i>Naïve Bayes Classifier</i> dan <i>Lexicon Based</i> memberikan nilai presentase <i>accuracy</i> 80%, <i>precision</i> 80%, <i>recall</i> 90% dan <i>F1_Score</i> 86%. Sedangkan pada <i>Lexicon Based</i> memberikan nilai presentase akurasi sebanyak 22%.
(Darwis et al., 2020)	Penerapan algoritma SVM Untuk Analisis Sentimen Pada Data <i>Twitter</i> Komisi Pemberantasan Korupsi Republik Indonesia	<i>Support Vector Machine</i>	Hasil dari klasifikasi menggunakan metode SVM dibagi menjadi tiga kelas, yaitu kelas positif sebanyak 8%, kelas negatif sebanyak 77%, dan kelas netral sebanyak 15%. Berdasarkan hasil pengujian dan evaluasi yang dilakukan pada nilai akurasi, <i>precession</i> , <i>recall</i> , dan

Tabel 2.1 *Literatur review*

			<p><i>F1-score</i>, dapat disimpulkan bahwa sentimen masyarakat pengguna <i>twitter</i> mengenai kinerja KPK dengan presentase sangat kurang baik yaitu ditunjukkan dengan munculnya kecondongan sentimen negatif sebesar 77% dengan keakuratan hasil pengujian akurasi sebesar 82% dan pengujian <i>precision</i> sebesar 90%, serta <i>recall</i> sebesar 88% dan <i>f1-score</i> sebesar 89%.</p>
(Deviyanto & Wahyudi, 2018)	<p>Penerapan Analisis Sentimen Pada Pengguna <i>Twitter</i> Menggunakan Metode <i>K-Nearest Neighbor</i></p>	<i>K-Nearest Neighbor</i>	<p>Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa penelitian tentang analisis sentimen pengguna <i>Twitter</i> terhadap topik Pilkada DKI 2017 dengan menggunakan metode <i>K Nearest Neighbor</i></p>

Tabel 2.1 *Literatur review*

			telah berhasil dilakukan. Hasil akurasi terbesar adalah 67,2% dengan nilai $k=5$. Sedangkan nilai presisi tertinggi sebesar 56,94% saat $k=5$ dan <i>recall</i> terbesar 78,24 % ketika $k=15$
(Anjas Ramadhan & Budi Setiawan SSi, n.d.)	Analisis Sentimen Program Acara Di SCTV Pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes</i> Dan <i>Support Vector Machine</i>	<i>Naive Bayes</i> Dan <i>Support Vector Machine</i>	Berdasarkan hasil penelitian untuk Analisis Sentimen Program Acara di SCTV pada <i>Twitter</i> menggunakan Metode <i>Naive Bayes</i> dan <i>Support Vector Machine</i> yang telah dilakukan mendapatkan Labelling terbaik didapatkan dengan cara Labelling Otomatis sedangkan akurasi terbaik dari Metode <i>Naive Bayes</i> dan <i>Support Vector Machine</i> adalah Metode <i>Support Vector Machine</i> dengan Seluruh

Tabel 2.1 *Literatur review*

			<p>Program Acara didapatkan hasil akurasi 88,57% berikut perkategori yang diperoleh Berita didapatkan hasil akurasi 79,81%, Entertainment didapatkan hasil akurasi 89,80%, Sinetron didapatkan hasil akurasi 73,68% dan FTV didapatkan hasil akurasi 87,74%.</p>
(Mahardhika & Zuliarso, 2018)	<p>Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial <i>Twitter</i> Menggunakan Algoritma <i>Naive Bayes Classifier</i></p>	<i>Naive Bayes Classifier</i>	<p>Metode <i>Naive Bayes Classifier</i> dalam melakukan klasifikasi <i>tweet</i> sentimen negatif dan positif dengan 300 data latih dan 100 data uji mendapat hasil akurasi sebesar 97%. Dengan hasil akurasi yang cukup tinggi yaitu 97% maka metode <i>Naive Bayes Classifier</i> dapat digunakan untuk melakukan</p>

Tabel 2.1 *Literatur review*

			klasifikasi <i>tweet</i> dengan sentimen
(Arsi & Waluyo, 2021)	Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	<i>Support Vector Machine</i>	Berdasarkan hasil dari proses pengujian yang telah dilakukan terhadap <i>tweets</i> sentimen pemindahan ibu kota dari media sosial <i>twitter</i> sebanyak 1.236 <i>tweets</i> (404 positif dan 832 negatif) menggunakan SVM maka hasil hipotesa awal sesuai hasil akhir yakni SVM lebih baik dari metode sebelumnya (BM25 + KNN dan <i>Naive Bayes</i>). Dengan akurasi =96,68%, <i>precision</i> =95.82%, <i>recall</i> =94.04% dan AUC = 0,979.

2.1.1. Tinjauan Terhadap Literatur 1

Studi ini menganalisis opini publik terhadap kegiatan Dewan Perwakilan Rakyat (DPR) yang diungkapkan di media sosial *Twitter*. Analisis sentimen

melibatkan beberapa langkah yaitu pengumpulan data (*crawling*), preprocessing data yang terdiri dari proses cleaning data, *tokenization*, stop removal dan case folding, splitting data dan klasifikasi data menggunakan metode pengklasifikasi *Naive Bayes*. Penelitian ini menggunakan 1546 permintaan data. Hasil penelitian ini menunjukkan bahwa DPR mengidentifikasi 95 *tweet* positif dengan polaritas 0,75 atau 75% positif, 693 *tweet* netral dengan polaritas 0,79 atau 79% netral dan 758 *tweet* negatif dengan polaritas 0,82 atau 82% menerima negatif. opini dengan tingkat akurasi 0,8 atau 80% berdasarkan data uji sampai dengan 20%.

2.1.2. Tinjauan Terhadap Literatur 2

Sekarang ini, media sosial khususnya *Twitter* mempunyai pengaruh yang sangat besar bagi kesuksesan ataupun kehancuran citra seseorang. Banyak gerakan-gerakan yang dilakukan dalam media sosial khususnya *Twitter* yang semuanya sukses mampu mempengaruhi penggunanya. Ada gerakan yang bertujuan baik ada juga gerakan dengan tujuan jahat yaitu menebar kebencian kepada orang lain. Biasanya gerakan di *Twitter* itu dilakukan dengan menggunakan tagar (#), Gerakan terbaru ada tagar Hatespeech (#HateSpeech), dilihat dari namanya sudah jelas yaitu ucapan kebencian. Penelitian ini menganalisis tagar tersebut, benarkan semua yang diberi tagar itu sentimennya kebencian. Proses klasifikasi pada penelitian ini menggunakan metode klasifikasi *Naive Bayes Classifier* (NBC) dan Support Vector Machine (SVM) dengan *preprocessing* data menggunakan tokenisasi, *cleansing* dan filtering. Data yang digunakan adalah *tweet* dalam bahasa Indonesia dengan tagar HateSpeech (#HateSpeech), dengan jumlah *dataset* sebanyak 522 *tweet* yang didistribusikan secara merata menjadi dua sentimen HateSpeech dan GoodSpeech. Hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi Support

Vector Machine (SVM) dengan tokenisasi unigram, *stopword* list Bahasa Indonesia dan emoticons, dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai *recall* 66,7% nilai TP rate 66,7% dan nilai TN rate 75,8%.

2.1.3. Tinjauan Terhadap Literatur 3

Dengan berkembangnya teknologi informasi di dunia, sarana transportasi juga ikut berkembang dengan berkembangnya layanan transportasi online. Saat ini penggunaan jasa transportasi online sudah seperti kebutuhan hidup, perlu dilakukan analisis sentimen terhadap jasa transportasi online untuk memahami bagaimana respon masyarakat terhadap layanan transportasi online. Data yang digunakan haruslah data yang valid. Media yang digunakan penulis untuk mengambil data adalah dari salah satu platform media sosial yaitu *Twitter*. Tugas akhir ini dibuat untuk menganalisis reaksi masyarakat dengan menganalisis data berupa *tweet* kemudian mengklasifikasikannya ke dalam kategori positif dan negatif dengan menggunakan pendekatan *Naive Bayesian classifier*. Berdasarkan sistem yang dibangun diperoleh hasil sebesar 88,60% untuk sentimen positif dan 11,40% untuk sentimen negatif, dengan tingkat akurasi sebesar 86,80%. Hasilnya menunjukkan bahwa *tweet* publik memiliki tingkat sentimen positif yang lebih tinggi daripada sentimen negatif.

2.1.4. Tinjauan Terhadap Literatur 4

Film adalah topik yang sangat menarik. Ketika seseorang memberikan pendapat tentang sebuah film, semua elemen film tersebut dituliskan. Data opini film dalam penelitian ini diambil dari review film di *Twitter*. Banyaknya opini yang di-*tweet* perlu dipilah berdasarkan sentimen yang dimiliki sehingga mudah untuk memahami bagaimana opini tersebut condong ke arah film, baik positif maupun

negatif. Algoritma yang akan digunakan dalam penelitian ini adalah algoritma *Naive Bayes*. Berdasarkan hasil eksperimen, sistem dapat melakukan analisis sentimen dengan tingkat akurasi 90%, tingkat akurasi 92%, tingkat penarikan kembali 90%, dan ukuran-f 90%.

2.1.5. Tinjauan Terhadap Literatur 5

Presiden Joko Widodo menerbitkan Instruksi Presiden Republik Indonesia Nomor 1 Tahun 2022 tentang Optimalisasi Pelaksanaan Program Jaminan Kesehatan Nasional. Instruksi tersebut menetapkan bahwa peserta BPJS Kesehatan harus terdaftar untuk mengakses berbagai layanan publik seperti jual beli tanah, mengurus SIM, SKCK, serta kegiatan ibadah haji dan umrah. Aturan tersebut mulai berlaku pada 1 Maret 2022. Masalah muncul ketika pendapat diutarakan tanpa dasar etika, sehingga mengakibatkan kritikan atau penghinaan pada pihak yang bersangkutan, yang dapat mengarah pada tindakan cyberbullying. Oleh karena itu, diperlukan analisis sentimen pada komentar *Twitter* untuk mengidentifikasi *tweet* yang mengandung cyberbullying atau bukan. Sebagai hasilnya, analisis sentimen pada komentar *Twitter* diperlukan untuk mengategorikan *tweet* sebagai cyberbullying atau non-cyberbullying. Metodologi yang digunakan dalam penelitian ini adalah deskriptif kualitatif, dan prosedur pengolahan data cyberbullying menggunakan metode Lexicon Based dan *Naive Bayes Classifier*. Pengambilan data *tweet*, pre-processing, dan performa *Naive Bayes Classifier* merupakan langkah pertama dalam proses ini. Pendekatan Lexicon Based kemudian digunakan untuk melakukan proses klasifikasi, dengan keluaran sistem menunjukkan apakah *tweet* termasuk kategori cyberbullying atau tidak. *Naive Bayes Classifier* pada penelitian ini memiliki akurasi sebesar 80%, dibandingkan

dengan akurasi klasifikasi Lexicon Based sebesar 22%. Dari perbandingan kedua pendekatan tersebut, jelas bahwa *Naive Bayes Classifier* lebih baik dan lebih akurat daripada Lexicon Based.

2.1.6. Tinjauan Terhadap Literatur 6

Komisi Pemberantasan Korupsi (KPK RI) merupakan organisasi terdepan yang memiliki kewenangan penuh dan diharapkan melakukan pekerjaan yang baik dalam memberantas korupsi. Namun, kemajuan teknologi membuat masyarakat semakin mudah mengakses dan berbagi informasi atau pemikiran di media sosial tanpa terbatas oleh waktu atau ruang. Salah satu situs media sosial yang digunakan untuk menampung perspektif ini adalah *Twitter*. Algoritma ekstraksi fitur SVM dan TF-IDF digunakan dalam penelitian ini sebagai pendekatan klasifikasi. Studi ini menghasilkan 1890 data dan 3846 kata/istilah dari 2000 data yang diperoleh melalui *crawling Twitter*. Untuk pelabelan, nilai kemunculan kata dihitung, menghasilkan sikap yang positif, negatif, atau netral. Berdasarkan temuan pengujian, penerapan metode SVM menghasilkan skor akurasi sebesar 82% dan persentase yang lebih tinggi pada sikap yang dikategorikan sebagai negatif (77%), positif (8%), dan netral (25%).

2.1.7. Tinjauan Terhadap Literatur 7

Dalam penelitian ini, pendapat yang diungkapkan oleh pengguna *Twitter* tentang Pemilihan Gubernur Jakarta 2017 dianalisis menggunakan algoritma K-Nearest Neighbor (KNN). Program *Python* yang disebut *Twitterscraper* digunakan untuk mengumpulkan 2000 *tweet* berbahasa Indonesia yang menjadi data *twitter* penelitian selama bulan Januari 2017. Skor sentimen positif dan negatif akan dibagi menjadi dua kategori menggunakan algoritma KNN, pembobotan kata TF-IDF, dan fungsi kesamaan kosinus. Berdasarkan hasil pengujian, nilai *recall* adalah 78,24%

ketika $k=15$, presisi terbaik adalah 56,94%, dan nilai akurasi tertinggi adalah 67,2% ketika $k=5$.

2.1.8. Tinjauan Terhadap Literatur 8

Keberhasilan sebuah program televisi ditentukan sebagian oleh kepuasan penontonnya. Sebuah perusahaan dapat membuat jadwal tayang yang baik dan menentukan durasi dan jumlah episode suatu acara berdasarkan kepuasan penonton. Selain itu, perusahaan juga dapat meminimalkan kerugian akibat jumlah penonton yang tidak sesuai dengan harapan dengan melakukan analisis sentimen pada *tweet* *Twitter* SCTV untuk mengetahui sentimen dari setiap *tweet*. Dalam penelitian ini, digunakan metode klasifikasi *Naive Bayes* dan Support Vector Machine. Hasil penelitian menunjukkan bahwa pendekatan Support Vector Machine menghasilkan akurasi tertinggi, mencapai 88,57% pada seluruh program acara.

2.1.9. Tinjauan Terhadap Literatur 9

Tujuan dari penelitian ini adalah untuk menentukan apakah *tweet* yang disampaikan dalam bahasa Indonesia di *Twitter* bersentimen positif atau negatif. *Tweet* di klasifikasikan menggunakan algoritma *Naive Bayes* classifier. Hasil klasifikasi pada data uji menunjukkan bahwa algoritma *Naive Bayes* Classifier memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentimen negatif dan 98% untuk sentimen positif. *Twitter* merupakan situs jaringan sosial yang populer di mana pengguna dapat mengirimkan komentar dan konten lainnya tanpa batasan. Banyak pengguna yang menggunakan bahasa kasar bahkan ujaran kebencian. Banyak komentar yang diberikan tentang pemerintahan Joko Widodo, termasuk pujian, kritik, dan penghinaan. Dalam penelitian ini, analisis sentimen digunakan untuk mengkategorikan dokumen teks

menjadi dua kategori yaitu sentimen negatif dan sentimen positif. Data opini diperoleh dari *Twitter* berdasarkan pencarian dalam bahasa Indonesia. Penelitian ini menggunakan total 400 *tweet*, 300 untuk pelatihan dan 100 untuk pengujian. 150 data sentimen negatif dan 150 data sentimen positif digunakan untuk data latih. Penelitian ini bertujuan untuk menentukan apakah sebuah *tweet* dalam Bahasa Indonesia di *Twitter* memiliki sentimen positif atau negatif. Teknik klasifikasi *Naive Bayes classifier* digunakan untuk mengategorikan data *tweet*. Berdasarkan hasil klasifikasi pada data uji, metode klasifikasi *Naive Bayes classifier* memiliki nilai akurasi sebesar 97%. Nilai akurasi untuk sentimen positif sebesar 98% dan untuk sentimen negatif sebesar 96%.

2.1.10. Tinjauan Terhadap Literatur 10

Twitter merupakan salah satu platform media sosial yang paling digemari saat ini karena menjadi salah satu industri online yang berkembang dengan cepat. Pembahasan di *Twitter* meliputi berbagai topik, termasuk politik, isu sosial, budaya, hukum, dan lain-lain. Salah satu topik yang paling banyak diperdebatkan di *Twitter* adalah pemindahan ibu kota Indonesia. Namun, terdapat perbedaan pendapat di antara individu yang mendukung dan menentang topik tersebut, masing-masing dengan pandangan yang berbeda. Hal ini menyebabkan munculnya fenomena diskusi, terutama di *Twitter*, yang sebenarnya menunjukkan minat umum dalam percakapan publik. Tahapan dalam proses analisis sentimen adalah ekstraksi, pemahaman, dan pengolahan data teks yang tidak terstruktur secara otomatis, yang bertujuan untuk mengambil informasi sentimen dari pendapat atau pernyataan. Ada banyak cara yang sering digunakan dalam penerapan analisis sentimen menggunakan teknik pembelajaran mesin. Untuk tujuan klasifikasi sentimen pada *Twitter*, penelitian ini menyarankan menggunakan metode Support Vector Machine

(SVM) pada *tweet* mengenai pemindahan ibu kota Indonesia. Klasifikasi positif dan negatif dilakukan dengan menggunakan teknik klasifikasi. Berdasarkan hasil pengujian yang menggunakan 1.236 *tweet* sentimen (404 positif dan 832 negatif) tentang pemindahan ibu kota dari *Twitter*, ditemukan akurasi sebesar 96,68%, presisi sebesar 95,82%, *recall* sebesar 94,04%, dan AUC sebesar 0,979.

2.2. BPJS Kesehatan

BPJS Kesehatan sudah terbentuk sejak tahun 1968 dengan nama yang berbeda. Jika dulu banyak orang yang mengenalnya sebagai nama PT Askes. Namun sekarang ini berganti nama menjadi BPJS sejak tahun 2014. Melalui Program Jaminan Kesehatan Nasional-Kartu Indonesia Sehat (JKN-KIS) yang diselenggarakan oleh BPJS Kesehatan, negara hadir di tengah kita untuk memastikan seluruh penduduk Indonesia terlindungi oleh jaminan kesehatan yang komprehensif, adil, dan merata.

Adapun dari kebijakan – kebijakan BPJS Kesehatan sekarang banyak yang mengkritisi bahkan menolak, hal itu salah satunya dikarenakan adanya kenaikan iuran BPJS Kesehatan. Presiden Konfederasi Serikat Pekerja Indonesia (KSPI) Said Iqbal menyatakan bahwa kalangan buruh akan tetap menggelar aksi unjuk rasa untuk menyampaikan aspirasi pada hari ini, Rabu, 2 Oktober 2019. "Besok kaum buruh akan tetap melakukan aksi besar-besaran di 10 provinsi. Khusus di Jabodetabek, aksi akan di DPR RI," kata Said dalam siaran pers konfederasi yang diterima di Jakarta, Selasa, 1 Oktober 2019. Dalam unjuk rasa tersebut, serikat buruh akan menyampaikan tiga tuntutan utama, yaitu menolak revisi Undang-Undang Nomor 13 Tahun 2003 tentang Ketenagakerjaan, menolak kenaikan iuran BPJS Kesehatan, dan menuntut revisi Peraturan Pemerintah Nomor 78 tahun 2015

tentang Pengupahan. Maupun pengakuan dari pihak BPJS Kesehatan sendiri, hal ini disampaikan Kepala BPJS Kesehatan Cabang Ambon, Afliana Latumakulita, dalam rapat koordinasi bersama Pemerintah Provinsi Maluku di Kantor Gubernur Maluku, Kota Ambon, Rabu (18/9/2019). "Dalam rapat dengan pemprov Maluku tadi, saya sampaikan dari sisi kualitas pelayanan belum merata baik, dari sisi sumber daya manusia maupun sarana-prasarana," kata Alfiana kepada wartawan di kantor Gubernur Maluku, Kota Ambon. Menurutnya, dari 205 unit puskesmas, hanya 160-unit yang ditempatkan dokter gigi. Bukan saja itu, terdapat 30 puskesmas yang belum memiliki dokter umum. Dalam praktiknya BPJS Kesehatan masih mengalami kerugian. Untuk mengurangi defisit BPJS Kesehatan, Ikatan Dokter Indonesia (IDI) sempat mengusulkan skema sistem terpisah mirip di Thailand. Masyarakat mampu menggunakan asuransi swasta dan BPJS Kesehatan untuk masyarakat miskin. Peningkatan iuran untuk BPJS Kesehatan yang baru baru ini santer terdengar menuai pro-kontra. Terlepas dari itu, pemerintah perlu meninjau kemampuan bayar kelompok masyarakat miskin, kualitas pelayanan kesehatan yang diberikan selama ini mengingat selalu defisit serta efektivitas pelayanan yang ada seperti fasilitas dan plafon yang diberikan (Bima et al., 2020).

2.3. *Twitter*

Twitter adalah salah satu media komunikasi dan media sosial. Fungsi yang dibuat adalah agar pengguna dapat dengan mudah mengungkapkan dan berbagi informasi. *Twitter* sangat diminati oleh masyarakat di seluruh dunia. Media sosial ini dapat berbagi informasi tentang berbagai hal seperti topik, ekspresi, keinginan,

dll. berita terbaru. Dan opini yang ditulis oleh pengguna *Twitter* (Sari & Haranto, 2019).

Twitter merupakan media terbuka (publik), sehingga setiap orang dapat mengakses dan melihatnya, sehingga pengguna *Twitter* akan lebih mudah untuk terhubung dengan pengguna *Twitter* lainnya (termasuk pengguna *Twitter* terkini) dan lebih memperhatikan tren, cerita, Informasi dan berita terbaru dari seluruh dunia, kapanpun dan dimanapun. Pengguna *Twitter* dapat menulis artikel yang disebut *tweet*, yang berisi informasi atau data dalam bentuk teks dengan aturan hanya 140 karakter yang dapat dimasukkan. Oleh karena itu, sangat layak untuk menggunakan *twitter* sebagai wadah sumber data atau informasi dalam bentuk teks, yang dapat digali dan digunakan untuk berbagai keperluan di berbagai bidang penelitian.

2.4. Text Mining

Text mining dijelaskan sebagai proses ekstraksi data berbasis teks, umumnya dari dokumen, untuk menemukan istilah yang dapat mencerminkan informasi yang terkandung di dalamnya dan menganalisis hubungan antara dokumen. Penanganan data teks dalam jumlah besar seperti pemrosesan, pengorganisasian, atau pengelompokan dapat diatasi dengan bantuan *text mining* (Al Khadafi et al., 2022). Beberapa aplikasi dalam *text mining* diantaranya (Lukmana et al., 2019) :

- a. *Text Categorization*, merujuk pada proses penentuan satu atau lebih kategori yang belum terdefinisi pada suatu teks.

- b. *Spell and Grammar Checking*, mengacu pada proses pengecekan ejaan kata dan memberikan saran kata alternatif.
- c. *Topic Modeling* digunakan untuk menemukan topik-topik yang terkandung dalam sekelompok dokumen.
- d. *Information Extraction* berguna dalam menemukan informasi relevan terhadap pertanyaan yang diajukan oleh pengguna.
- e. *Question Answering*, berkaitan dengan kecerdasan buatan, yang biasanya terdapat dalam aplikasi seperti *auto-chat* dan *Google Assistant*.
- f. Analisis Sentimen merupakan proses identifikasi sentimen dan pendapat pada sebuah teks.

2.5. Klasifikasi

Proses menentukan sebuah objek sebagai kategori, kelas, atau grup berdasarkan proses, definisi, dan atribut yang telah ditentukan disebut dengan klasifikasi. Tujuan dari klasifikasi adalah untuk menempatkan sebuah objek yang hanya memiliki definisi ke dalam salah satu kelas. Meskipun pengetahuan ahli dapat digunakan, namun teknik pembelajaran umumnya digunakan untuk membuat model klasifikasi dalam machine learning karena *dataset* yang relatif besar. Metode pembelajaran mandiri dari *dataset* ini dapat memberikan model klasifikasi yang memetakan objek data x dalam bentuk *labeled training dataset* dan menghasilkan model klasifikasi (Handoko & Suyanto, 2019).

2.6. Analisis Sentimen

Analisis sentimen merupakan sebuah proses untuk menganalisis atau mengidentifikasi sebuah opini seseorang yang menunjukkan sikap terhadap suatu topik atau produk tertentu masuk ke dalam kategori positif, negatif, atau netral.

Tidak seperti informasi faktual, opini dan sentimen memiliki karakteristik yang menunjukkan bahwa hal tersebut bersifat subjektif. Pemeriksaan sekumpulan opini dari banyak pihak diperlukan karena hal tersebut merupakan pandangan subjektif yang asalnya lebih dari satu orang sehingga diperlukan sebuah ringkasan untuk mewakili suatu pendapat (Akhmad et al., 2019). Terdapat 3 tahap dalam melakukan analisa sentimen yaitu *preprocessing*, pembobotan, dan klasifikasi.

1. *Preprocessing*

merupakan tahap persiapan data yang bertujuan agar mempermudah proses pengolahan data. *Preprocessing* memfokuskan pada data *cleaning & cleansing*, termasuk menghilangkan *noise* di data, mengatasi struktur data yang tidak baik, dan informasi yang hilang. Berikut tahapan yang ada dalam *preprocessing* data:

a. *Casefolding*

Saat menulis *tweet*, biasanya terdapat variasi dalam penggunaan huruf besar dan kecil. Tahap ini melibatkan mengubah huruf-huruf tersebut menjadi huruf kecil (*lower case*), atau disebut juga proses standarisasi huruf.

b. *Cleansing*

Tujuan dari pembersihan data adalah untuk mengurangi *noise* pada data *tweet*. Oleh karena itu, beberapa kata atau tanda harus dihilangkan, seperti URL, hashtag (#), username (@username), email, emotikon (: @, :*, :D), serta tanda baca seperti koma (,) dan titik (.

c. *Tokenizing*

Pada dasarnya, *tokenizing* atau *parsing* adalah tahap pemisahan string input menjadi kata-kata yang membentuknya. Proses *tokenizing* ini dapat

diibaratkan sebagai pemotongan kalimat menjadi kata-kata yang terdapat di dalamnya. Sebagai contoh, jika kita memiliki sebuah kalimat, maka proses tokenizing akan memisahkan kalimat tersebut menjadi kumpulan kata-kata.

d. Normalisasi

Normalisasi mengubah kata-kata tidak baku menjadi kata baku, yaitu “yg” menjadi “yang” atau “apotik” menjadi “apotek”. Perbedaan dengan tahap satu adalah pada tahap satu lebih berfokus pada bahasa slang, sedangkan tahap dua jauh lebih luas lagi. Selain itu juga terdapat kamus “*kamusalay*” yang dapat membantu dalam mengubah akronim dari bahasa gaul (*slang words*) menjadi bentuk kata baku. Ini membantu dalam memastikan bahwa teks lebih mudah dipahami dan tidak ambigu. Bahasa Indonesia memiliki bentuk-bentuk kolokialisme seperti yang ditemukan dalam bahasa lain. Beberapa kata kolokial memiliki etimologi yang serupa dengan yang ada dalam bahasa Inggris, misalnya karena perubahan fonem atau suara, kasus morfologis seperti afiksasi, atau bahkan kasus seperti “bahasa gey.” Dari sudut pandang linguistik, kita juga bisa menggunakan leksikon untuk mengamati tren terkini dari bahasa slang di media sosial. Secara khusus, kita dapat membandingkan kemunculan dan kategori kata-kata slang dalam leksikon dengan penelitian sebelumnya untuk menganalisis bagaimana kolokialisme dalam bahasa Indonesia bervariasi dari waktu ke waktu (Aliyah Salsabila et al., 2018).

e. *Stopword removal*

Stopword removal adalah sebuah proses di mana kita menghapus kata-kata yang tidak memberikan makna penting terhadap sesuatu yang sedang

dideskripsikan. Beberapa contoh kata-kata *stopword* yang biasanya dihapus antara lain "yang", "di", "ke", "itu", dan sebagainya. Sebagai contoh, ketika kita melakukan proses *stopword removal* pada sebuah kalimat, maka kata-kata *stopword* tersebut akan dihapus sehingga tersisa kata-kata yang lebih relevan dan memberikan makna yang lebih jelas terhadap kalimat tersebut.

f. *Stemming*

Stemming merupakan suatu proses yang dilakukan untuk mengubah kata-kata yang berimbuhan menjadi kata dasar yang sesuai dengan aturan bahasa Indonesia. Proses *stemming* ini dilakukan dengan tujuan untuk menghasilkan kata dasar yang memiliki makna yang sama dengan kata yang berimbuhan tersebut. Sebagai contohnya, ketika kita melakukan *stemming* pada sebuah kalimat, kata-kata yang memiliki awalan atau akhiran akan diubah menjadi kata dasar sehingga kalimat tersebut lebih mudah dipahami dan memiliki makna yang lebih jelas.

Pada penelitian ini modul yang digunakan adalah Sastrawi. *Stemmer* Sastrawi menerapkan algoritma yang didasarkan pada Nazief dan Adriani, lalu diperkuat dengan algoritma CS (*Confix Stripping*), Algoritma ECS (*Enhanced Confix Stripping*), dan selanjutnya ditingkatkan dengan ECS yang Dimodifikasi. Diharapkan bahwa dengan menggunakan perpustakaan Sastrawi ini dalam proses pra-pemrosesan dokumen keluhan mahasiswa, terutama pada tahap *stemming* dan penghapusan *stopword*, dapat menghilangkan kata-kata yang kurang penting sehingga meningkatkan akurasi (Rosid et al., 2020).

2.7. Naive Bayes Classifier

Naive Bayes Classifier adalah klasifikasi paling dasar dan sering digunakan. *Naive Bayes* adalah metode pembelajaran mesin yang memanfaatkan probabilitas dengan pendekatan Bayesian sebagai strategi komputasi dan klasifikasi, dengan asumsi independensi prediktor. Teorema Bayes menyatakan bahwa *Naive Bayes Classifier* memprediksi probabilitas masa depan berdasarkan pengalaman sebelumnya. Ketika diterapkan pada set data besar, *Naive Bayes* merupakan pendekatan klasifikasi probabilitas paling sederhana dan memiliki tingkat akurasi yang sangat tinggi. Selain itu, jika dibandingkan dengan metode klasifikasi lainnya, model *Naive Bayes* memiliki tingkat kesalahan yang relatif rendah. Untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian, algoritma *Naive Bayes* hanya membutuhkan sejumlah kecil data pelatihan. Karena variabel yang diasumsikan independen, menentukan klasifikasi hanya memerlukan variansi dari suatu variabel dalam sebuah kelas, bukan keseluruhan matriks kovarians. Dalam *library Python*, terdapat tiga jenis model *Naive Bayes* yaitu *Gaussian Naive Bayes*, *Multinomial Naive Bayes*, *Bernoulli Naive Bayes*.

Multinomial Naive Bayes memiliki bentuk umum sebagai berikut :

$$P(c|d) \propto P(c) \prod P(tk|c) \quad (1)$$

Di mana:

1. $P(c|d)$: Probabilitas dokumen d berada di kelas c
2. $P(c)$: Prior probability suatu dokumen berada di kelas c
3. $\{t_1, t_1, t_1, \dots, t_n\}$: Token dalam dokumen d yang merupakan bagian dari vocabulary dengan jumlah n
4. $P(tk|c)$: Probabilitas bersyarat term tk berada di dokumen pada kelas c

2.8. Confusion matrix

Confusion matrix adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek benar atau salah (Mustakim & Oktaviani, 2016).

Tabel 2.2 *Confusion matrix*

	<i>Predicted class</i>		
		<i>P</i>	<i>N</i>
<i>Actual Class</i>	<i>P</i>	<i>True Positives (TP)</i>	<i>False Negatives (FN)</i>
	<i>N</i>	<i>False Positives (FP)</i>	<i>True Negatives (TN)</i>

Nilai TP (*true positive*) dan TN (*true negative*) menunjukkan tingkat ketepatan klasifikasi. Jika label prediksi keluaran bernilai benardan nilai sebenarnya bernilai salah disebut sebagai FP (*false positive*). Sedangkan jika prediksi label keluaran bernilai salah dan nilai sebenarnya bernilai benar maka hal ini disebut sebagai FN (*false negative*) (Sasongko, 2016).

2.9. Performa

Confusion matrix akan menguji hasil performa sebuah metodeklasifikasi berupa akurasi, presisi, *recall*, dan *f-measure*. Akurasi adalah ketetapan sistem dalam melakukan prosesklasifikasi dengan benar. Persamaan akurasi ditunjukkan pada persamaan 3.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

Presisi adalah rasio jumlah data yang relevan dengan total jumlahdokumen yang ditemukan pada sistem klasifikasi. Persamaan presisi ditunjukkan pada persamaan 4.

$$Presisi = \frac{TP}{TP+FP} \quad (3)$$