

BAB II

TINJAUAN PUSTAKA

2.1 Twitter

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan (*tweet*). Twitter merupakan sebuah situs *micro-blogging* yang sangat populer di Indonesia. Hal ini terlihat dari jumlah pengguna Twitter yang mencapai 19,5 juta pengguna dari total 330 juta pengguna di dunia. Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan "pesan singkat dari Internet."

Twitter mengalami pertumbuhan yang pesat dan dengan cepat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif. Tingginya popularitas Twitter menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai aspek, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat. Kebiasaan masyarakat mengutarakan pendapatnya melalui media sosial terutama Twitter dalam menanggapi kejadian atau hal-hal yang terjadi di lingkungannya dapat menjadi salah satu acuan untuk mengetahui sentimen masyarakat terhadap lingkungan atau kota tempat tinggal masyarakat tersebut

berupa kritik atau saran (Arifiyanti, 2014). Twitter juga dihadapkan pada berbagai masalah dan kontroversi seperti masalah keamanan dan privasi pengguna, gugatan hukum, dan penyensoran.

2.2 Covid-19

Covid-19 adalah penyakit yang disebabkan oleh virus *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2). *Covid-19* dapat menyebabkan gangguan sistem pernapasan, mulai dari gejala yang ringan seperti flu, hingga infeksi paru-paru, seperti pneumonia (Alodokter, 2021).

Covid-19 (coronavirus disease 2019) adalah jenis penyakit baru yang disebabkan oleh virus dari golongan *coronavirus*, yaitu SARS-CoV-2 yang juga sering disebut virus *Corona*.

Kasus pertama penyakit ini terjadi di kota Wuhan, Cina, pada akhir Desember 2019. Setelah itu, *covid-19* menular antarmanusia dengan sangat cepat dan menyebar ke puluhan negara, termasuk Indonesia, hanya dalam beberapa bulan.

Untuk percepatan penanganan *covid-19* Kementerian Kesehatan (Kemenkes) menerbitkan peraturan Pembatasan Sosial Berskala Besar (PSBB) agar bisa segera dilaksanakan di berbagai daerah. Aturan PSBB tercatat dalam Peraturan Menteri Kesehatan Nomor 9 Tahun 2020.

2.3 PSBB

Pembatasan Sosial Berskala Besar (PSBB) adalah pembatasan kegiatan tertentu penduduk dalam suatu wilayah yang diduga terinfeksi *Corona Virus Disease 2019 (Covid-19)* (Kemenko PKM, 2019). Sekretaris Jenderal

Kementerian Kesehatan Oscar Primadi dalam keterangan tertulisnya mengatakan PSBB melingkupi pembatasan sejumlah kegiatan penduduk tertentu dalam suatu wilayah yang diduga terinfeksi *Covid-19*. “Pembatasan tersebut meliputi peliburan sekolah dan tempat kerja, pembatasan kegiatan keagamaan, pembatasan kegiatan di tempat atau fasilitas umum, pembatasan kegiatan sosial budaya, pembatasan moda transportasi, dan pembatasan kegiatan lainnya khusus terkait aspek pertahanan dan keamanan,” kata dia.

Kriteria wilayah yang menerapkan PSBB adalah memiliki peningkatan jumlah kasus dan kematian akibat penyakit *Covid-19* secara signifikan dan cepat serta memiliki kaitan epidemiologis dengan kejadian serupa di wilayah atau negara lain.

2.4 Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. *Data Mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Definisi lain mengatakan *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar.

Maka dari beberapa definisi di atas penulis dapat menyimpulkan bahwa *data mining* merupakan proses atau kegiatan untuk mengumpulkan data yang

berukuran besar, kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan.

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*. Tahap-tahap *Data Mining* adalah sebagai berikut:

1. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan-kan *noise* dan data yang tidak konsisten atau data tidak relevan.

2. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

3. Seleksi data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*.

5. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining*.

6. Evaluasi pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

7. Presentasi pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.5 Text Mining

Text mining juga dikenal sebagai data mining teks atau penemuan pengetahuan dari *database* tekstual. Sesuai dengan buku *The Text Mining Handbook*, *text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam *text mining* adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks dan pengelompokkan teks.

Text mining dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian/pengelompokkan dan menganalisa *unstructured* data dalam jumlah besar, dalam hal ini data yang akan digunakan adalah data yang diambil dari Twitter. Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *Data Mining*, *Information Retrieval*, Statistik dan Matematik, *Machine Learning*, *Linguistic*,

Natural Language Processing dan *Visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan penyimpanan teks, preprocessing akan konten teks, pengumpulan data statistik serta *indexing* dan analisis sentimen.

2.6 Analisis Sentimen

Analisis sentimen adalah tugas menemukan opini dari penulis tentang suatu entitas tertentu (Ruhyana, 2019). Analisis Sentimen mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang yang berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu.

Tugas dasar analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral. Analisis sentimen juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

Kita dapat mencari pendapat tentang produk-produk, merek atau orang-orang dan menentukan apakah mereka dilihat positif atau negatif di web. Hal ini memungkinkan kita untuk mencari informasi tentang:

- a. Persepsi produk baru.
- b. Persepsi Merek.
- c. Manajemen reputasi.

Ekspresi mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada *subject* yang

berbeda. Tujuan dari analisis sentimen adalah untuk menentukan perilaku atau opini dari seorang peneliti dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan (bagaimana si penulis ingin mempengaruhi pembaca).

2.7 Klasifikasi

Klasifikasi merupakan salah satu topik utama dalam data mining atau machine learning. Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam supervised learning atau pembelajaran yang diawasi (Wibowo, 2017). Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction* yang memanfaatkan kumpulan pengujian dari data set yang terklasifikasi. Dalam definisi yang lain klasifikasi dalam *data mining* dapat didefinisikan sebagai cara menganalisis data dengan menggunakan *training data* untuk membentuk sebuah model yang digunakan untuk memprediksi kelas label dari sebuah data.

Pengukuran kinerja klasifikasi dilakukan dengan mengevaluasi hasil pengujian menggunakan konfusi atau *confusion matrix*. *Confusion Matrix* merupakan metode untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah.

2.8 Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan *teorema Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya.

Naïve Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Persamaan dari teorema Bayes adalah:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots (1)$$

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

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

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

$P(X)$: Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut:

$$P_C F_1 \dots F_n = \frac{P_C P(F_1 \dots F_n | C)}{P(F_1 \dots F_n)} \dots (2)$$

Di mana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \dots (3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss*:

$$P(X_i = x_i | Y = y_j) = \frac{1}{2\pi\sigma_{ij}} e^{-\frac{x_i^2}{2\sigma_{ij}^2}} \dots (4)$$

Dimana :

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

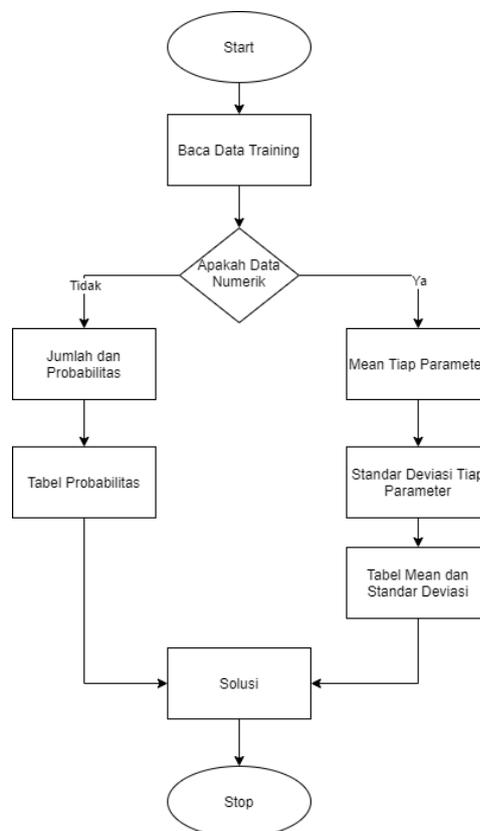
Y : Kelas yang dicari

y_i : Sub kelas Y yang dicari

μ : *mean*, menyatakan rata – rata dari seluruh atribut

σ : *Deviasi standar*, menyatakan varian dari seluruh atribut.

Alur dari metode *Naive Bayes* dapat dilihat pada gambar 1 sebagai berikut:



Gambar 2.1 Alur Metode *Naïve Bayes*

- a. Baca data training
- b. Hitunglah jumlah dan probabilitas, namun apabila data numerik maka:
 - Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

Adapun persamaan yang digunakan untuk menghitung nilai rata-rata (*mean*) dapat dilihat sebagai berikut;

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \dots (5)$$

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \dots (6)$$

di mana :

μ : rata – rata hitung (*mean*)

x_i : nilai sample ke - i

n : jumlah sampel

dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut:

$$\sigma = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} \dots (7)$$

di mana:

σ : standar deviasi

x_i : nilai x ke - i

μ : rata-rata hitung

n : jumlah sampel

- Cari nilai *probabilistic* dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
- c. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas.
- d. Solusi kemudian dihasilkan.

2.9 Tools

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, data *preprocessing* dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa *java* sehingga dapat bekerja di semua sistem operasi.

RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. RapidMiner didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai *software open source* untuk *data mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai *Software data mining* pada polling oleh

KDnuggets, sebuah portal *data-mining* pada 2010-2011. RapidMiner yang digunakan pada penelitian ini memakai *RapidMiner Studio Version 9.5*.

Microsoft Excel adalah sebuah program atau aplikasi yang merupakan bagian dari paket instalasi Microsoft Office, berfungsi untuk mengolah angka menggunakan *spreadsheet* yang terdiri dari baris dan kolom untuk mengeksekusi perintah. Microsoft Excel telah menjadi *software* pengolah data/angka terbaik di dunia, selain itu Microsoft Excel telah didistribusikan secara multi-platform. Microsoft Excel tidak hanya tersedia dalam platform Windows, Microsoft Excel juga tersedia di MacOS, Android dan Apple.

2.10 Matriks Penelitian

Tabel 2.1. Matriks Penelitian

No	Peneliti	Metode					Algoritma								Tools			Objek Penelitian						
		Klasifikasi	Regresi	Clustering	Asosiasi	AHP	C.45	K-Means	Support Vector Machine	Bayesian Network	Expectation Maximisation	Iterative Dichotomiser 3	K-Nearest Neighbors	Naïve Bayes	Naïve Bayes Classifier	Apk Antarmuka	RapidMiner	Weka	Python	Politik	Hiburan	Teknologi	Ekonomi	Lain-lain
1.	Faishol Nurhuda, dkk (2013)	✓												✓		✓			✓					
2.	Eza Putra Nuansa (2017)	✓						✓						✓				✓	✓					
3.	Rizal Setya perdana, dkk (2017)	✓											✓					✓		✓				

Penelitian ini menjadi salah satu acuan dalam melakukan penelitian sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Ulasan penelitian terkait, dilakukan dengan maksud untuk menganalisis penelitian yang telah dilakukan sebelumnya. Penelitian terdahulu dapat dilihat sebagai judul berikut:

1. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode *Naive Bayes Classifier*

Nama penulis dari jurnal tersebut adalah Nurhuda et al dengan tahun terbitan jurnal yaitu JURNAL ITSMART Vol 2. No 2. Desember 2013 ISSN : 2301–7201.

Penelitian ini membahas tentang pengamatan sentimen masyarakat melalui jejaring sosial Twitter menunjukkan jumlah percakapan terhadap capres dan cawapres di jejaring sosial Twitter menjelang mendekati pemilu 2014 semakin meningkat. Dalam penelitian ini telah dilakukan analisis sentimen masyarakat terhadap calon presiden dan wakil presiden Indonesia 2014 yang diungkapkan melalui jejaring sosial Twitter. Ada beberapa tahap untuk melakukan analisis sentimen, diantaranya adalah tahap pengumpulan data, *praprosesing data*, *POS Tagging*, ekstrasi opini menggunakan *rule based* dan klasifikasi opini menggunakan metode *Naive Bayes Classifier*.

Hasil dari penelitian ini didapatkan bahwa pasangan capres dan cawapres Prabowo Subianto dan Hatta Rajasa mendapatkan jumlah percakapan 53% dan pasangan Joko Widodo – Jusuf Kalla mendapatkan 47%. Sedangkan untuk hasil

polaritas sentimen, Prabowo Subianto – Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 26,4% sentimen negatif dan 25,9% sentimen netral. Sedangkan pasangan Joko Widodo – Jusuf Kalla mendapatkan total 37,6% sentimen positif, 34,4% sentimen negatif, dan 27,9 sentimen netral.

2. Analisis Sentimen Pengguna *Twitter* Terhadap Pemilihan Gubernur DKI Jakarta Dengan Metode *Naive Bayes Classification* dan *Support Vector Machine*

Nama penulis dari skripsi tersebut Putra Nuansa dengan tahun terbitan Skripsi 2017.

Penelitian ini membahas tentang analisis sentimen untuk mencari kata kunci dan hubungan pola antar tiap calon Gubernur DKI Jakarta. Metode *Naive Bayes Classifier* pada masing-masing pengukuran performa akurasi, *precision*, *recall*, dan *F-Measure* sebesar 85.77%; 85.90%; 85.77%; 85.67%. Metode *Support Vector Machine* kernel RBF tiap pengukuran performa akurasi, *precision*, *recall*, dan *F-Measure* adalah 87.80%; 98.48%; 87.80%; 92.64%. Untuk hasil SNA didapatkan hasil yang tinggi untuk *degree centrality* sebesar 0.865 yang menunjukkan pengaruh antar node kata kunci dengan kata kunci yang lain.

3. Analisis Sentimen Tentang Opini Film pada Dokumen *Twitter* Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku

Nama penulis dari jurnal tersebut adalah Prananda Antinasari, Rizal Setya Perdana dengan tahun terbit Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 1, No. 12, Desember 2017, hlm. 1733-1741 e-ISSN: 2548-964X.

Penelitian ini membahas tentang analisis sentimen untuk mengklasifikasikan kedalam sentimen negatif atau positif. Didalam Tweet terkandung banyak ragam bahasa yang digunakan, yaitu diantaranya bahasa dalam bentuk tidak baku seperti bahasa slang, penyingkatan kata, dan salah eja. Oleh sebab itu dibutuhkan penanganan khusus pada *tweet*. Pada penelitian ini digunakan kamus kata tidak baku dan normalisasi *Levenshtein Distance* untuk memperbaiki kata yang tidak baku menjadi kata baku dengan pengklasifikasian *Naive Bayes*. Berdasarkan hasil pengujian yang telah dilakukan didapatkan akurasi tertinggi dengan nilai *accuracy*, *precision*, *recall*, dan *f-measure* sebesar 98.33%, 96.77%, 100%, dan 98.36%.

4. Analisis Sentimen Pasar Otomotif Mobil: *Tweet* Twitter Menggunakan *Naive Bayes*

Nama penulis dari jurnal tersebut adalah Rustiana & Rahayu dengan tahun terbit Jurnal SIMETRIS, Vol 8 No 1 April 2017 ISSN: 2252-4983.

Penelitian ini membahas tentang Media sosial seperti twitter banyak dipakai untuk mencurahkan isi hati pemakai nya, sehingga data tersebut dapat menggambarkan sentimen. *Naive bayes* merupakan metode klasifikasi analisis sentiment yang banyak digunakan karena ketangguhan, sederhana dan akurasi yang tinggi dalam mengklasifikasi data. Pada penelitian ini menganalisis data dari Twitter dengan pencarian merk mobil yang terlaris berdasarkan data dari GAIKINDO

(Gabungan Industri Kendaraan Bermotor Indonesia). Dengan menggunakan tiga kelas maka dapat menghasilkan nilai akurasi sebesar 93%, dan tingkat *precision* positif sebesar 90%, *precision* negatif sebesar 90% dan sentimen yang netral 100% sehingga *naïve bayes* mampu menganalisis dengan akurasi yang tinggi.

5. Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers

Nama dari jurnal tersebut adalah A.Jabbar Alkubaisi et al dengan tahun terbit *Computer and Information Science*; Vol. 11, No. 1; 2018 ISSN 1913-8989 E-ISSN 1913-8997.

Penelitian ini membahas tentang Rendah akurasi dalam klasifikasi memiliki efek langsung pada keandalan indikator pasar saham. Penelitian utamanya berfokus pada analisis dataset Twitter. Selain itu, model yang lebih baik diusulkan dalam penelitian ini; ini dirancang untuk meningkatkan akurasi klasifikasi. Fase pertama dari model ini adalah pengumpulan data, dan yang kedua melibatkan penyaringan dan transformasi, yang dilakukan untuk mendapatkan hanya data yang relevan. Fase yang paling penting adalah pelabelan, di mana polaritas data ditentukan dan nilai negatif, positif atau netral ditetapkan pendapat orang. Fase keempat adalah fase klasifikasi di mana pola yang sesuai dari pasar saham diidentifikasi dengan hibridisasi *Naïve Bayes Classifiers* (NBCs), dan fase terakhir adalah kinerja dan evaluasi. Penelitian ini mengusulkan *Hybrid Naïve Bayes Classifiers* (HNBCs) sebagai metode pembelajaran mesin untuk pasar saham klasifikasi. Hasilnya sangat penting bagi investor, perusahaan, dan peneliti yang memungkinkan mereka untuk merumuskan rencana

mereka sesuai dengan sentimen orang. Metode yang diusulkan telah menghasilkan yang signifikan hasil itu telah mencapai akurasi sama dengan 90.38%.

6. Sentimen Analisis *Tweet* Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan *Naïve Bayes*

Nama Penulis dari jurnal tersebut adalah Pudjajana & Manongga dengan tahun terbit Jurnal SIMETRIS, Vol. 9 No. 1 April 2018 ISSN: 2252-4983.

Penelitian ini membahas tentang Penolakan yang dialami membuat kaum homoseksual membuat kaum tersebut menggunakan media sosial seperti Twitter untuk membuka diri di khalayak umum, mencari pasangan sejenis, dan mencari penghasilan. Penelitian ini dilakukan sentimen analisis pada *tweet* Twitter sebagai *text mining* menggunakan metode *Naïve Bayes*. Tujuan penelitian ini adalah mengetahui hasil sentimen positif dan negatif terhadap data uji *tweet* dan berdasarkan hasil pengujian tersebut dapat disampaikan kepada pengguna Twitter secara luas untuk menggunakan Twitter secara tepat. Selain itu juga, perhitungan *Naive Bayes* dibandingkan dengan *k-Nearest Neighbor* (k-NN) untuk mengetahui tingkat akurasi. Hasil sentimen analisis terhadap 500 data uji menunjukkan bahwa nilai sentimen negatif sangat tinggi yaitu 68.4%. Sedangkan hasil perbandingan akurasi kedua metode adalah metode *Naïve Bayes* sebesar 87.48% dan k-NN 85.40% dimana metode *Naive Bayes* lebih akurasi dibanding metode k-NN.

7. *Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm*

Nama penulis dari jurnal tersebut adalah Malik & Kumar, n.d dengan tahun terbit *International Journal on Recent and Innovation Trends in Computing and Communication Volume: 6 Issue: 4 April 2018* ISSN: 2321-169 120 - 125.

Penelitian ini membahas tentang keingintahuan seberapa baik kinerja NLTK dan algoritma *Naïve Bayes Machine Learning* untuk Analisis Sentimen. Dalam penelitian tersebut, ini bekerja dengan baik untuk komentar negatif. Masalah muncul ketika *tweet* itu ironis, sarkastik memiliki referensi atau konteks yang sulit. Pertimbangkan *tweet* berikut: "Muhaha, betapa sedihnya bahwa Liberal tidak bisa menghancurkan Trump. Berbaris maju. " Seperti yang mungkin sudah Anda pikirkan, kata-kata sedih dan hancur sangat memengaruhi evaluasi, meskipun *tweet* ini harus positif ketika mengamati makna dan konteksnya. Untuk meningkatkan akurasi evaluasi, perlu adanya sesuatu untuk mempertimbangkan konteks dan referensi. Setelah itu, akan mencoba membangun jaringan LSTM, dan membandingkan hasilnya dengan implementasi pembelajaran mesin NLTK ini.

8. Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan *Naïve Bayes*

Nama penulis dari jurnal tersebut adalah Lorosae & Prakoso dengan tahun terbit Seminar Nasional Teknologi Informasi dan Multimedia 2018 ISSN : 2302-3805.

Jurnal ini membahas tentang permasalahan yang terjadi dalam penelitian ini terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian ini bertujuan untuk mengembangkan aplikasi yang mampu

melakukan analisis sentimen konsumen, dimana *Naive Bayer Classifier* digunakan sebagai metode klasifikasi.

Terdapat penelitian terdahulu menghasilkan informasi sentimen masyarakat yang mengarah ke sentimen positif dan negatif mengenai asuransi dengan menggunakan metode *Naive Bayes Classifier* dengan menghasilkan akurasi hingga 95%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen pada twitter kedalam sentiment positif, netral atau negatif. Hasil yang didapatkan dari akurasi *naive bayes* pada data uji positif memperoleh ketepatan 84%.

9. Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma *Naive Bayes*

Nama penulis jurnal tersebut adalah Muljono, Dian Putri Artanti, Abdul Syukur, Adi Prihandono dengan tahun terbit Konferensi Nasional Sistem Informasi 2018.

Jurnal ini membahas tentang analisis sentimen terhadap postingan opini pelanggan *online marketplace* di Indonesia pada Twitter yang nantinya bisa digunakan untuk menentukan *rating online marketplace* agar masyarakat tidak salah pilih situs *marketplace* untuk berbelanja di masa depan. Langkah pertama melakukan koleksi data opini masyarakat di twitter dari situs belanja *online*. Kemudian dilakukan *pre-processing* pada data yang meliputi *cleansing data, case folding, tokenizing, case normalization, stop word, convert negation* dan *stemming*. Selanjutnya dilakukan proses pemberian label (kelas) pada data tersebut yang dilakukan oleh ahli bahasa.

Hasil dari proses *clustering* tercipta dua kelas data yaitu kelas data positif dan kelas data negatif dengan jumlah total 1200 data. Data yang sudah memiliki kelas data ini, digunakan sebagai data *training* untuk mesin pengklasifikasi, dalam riset ini menggunakan algoritma mesin pengklasifikasi *Naïve Bayes*. Terakhir, kami mengukur kinerja dari mesin pengklasifikasi menggunakan *10-fold cross validation*. Hasil evaluasi menunjukkan rata-rata akurasi sebesar 93.33%.

10. Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier

Nama penulis jurnal tersebut adalah Hermanto et al. dengan tahun terbit *Journal of Physics: Conf. Series* 1140 (2018) 012037 IOP Publishing doi:10.1088/1742-6596/1140/1/012037 .

Jurnal ini membahas tentang tujuan wisata saat ini di Indonesia telah berkembang pesat yang terpengaruh oleh peran teknologi yang memiliki pengaruh besar terhadap kebebasan akses dalam mencari informasi. *Tweet* juga memiliki kemungkinan berisi informasi atau kondisi tentang tujuan wisata yang akan atau telah mereka kunjungi, seperti pengalaman pengunjung dalam tur, pendapat pengunjung tentang tempat wisata, dan tempat wisata lainnya. Metode *Naive Bayes* menggunakan data pelatihan untuk menciptakan probabilitas setiap kriteria untuk kelas yang berbeda, namun demikian, nilai probabilitas kriteria dapat ditingkatkan untuk menghasilkan sentimen analisis terhadap pendapat di twitter. Penelitian ini dilakukan dengan mengklasifikasikan pendapat pada PT bentuk komentar menjadi dua kelas, yaitu positif dan negatif dengan tingkat akurasi yang dipengaruhi oleh proses

pelatihan. Berdasarkan hal itu, dapat disimpulkan dengan jelas bahwa informasi sentimen publik ke tempat wisata termasuk dalam positif sentimen.

11. Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode *Naïve Bayes*

Nama penulis tersebut adalah Haqqizar & Larasyanti dengan tahun terbit Prosiding TAU SNAR-TEK 2019 Seminar Nasional Rekayasa dan Teknologi 27 November 2019 ISSN : 2715-6982.

Penelitian ini membahas tentang sentimen analisis mengenai tweet #Telkomsel atau @Telkomsel di Twitter menunjukkan bahwa banyak masyarakat memberikan “sentimen negatif” terhadap layanan Telkomsel. Hasil pengolahan terhadap 151 data latih disimpulkan bahwa hasil klasifikasi sentimen yang didapat adalah sentimen negatif sebanyak 51 *tweets*, sentimen netral sebanyak 51 *tweets*, dan sentiment positif sebanyak 49 *tweets* di mana netral adalah opini yang tidak bersentimen. Tingkat akurasi dalam penentuan kategori adalah sebesar 70.21% dan *micro average* 70.20% dalam penentuan sentimen memiliki tingkat *precision* 70.11% dan *recall* 70.33%.

12. Analisis Sentimen *Hatespeech* Pada Twitter Dengan Metode *Naïve Bayes* *Classifier* Dan *Support Vector Machine*

Nama penulis jurnal tersebut adalah Buntoro dengan tahun terbit Jurnal Dinamika Informatika Volume 5, Nomor 2, September 2016 ISSN 1978-1660.

Penelitian ini membahas tentang Analisis sentimen *tweet* Bahasa Indonesia dengan tagar *Hatespeech* dapat membantu menentukan sentimen pada *tweet* opini

Bahasa Indonesia yang ada di Twitter. Setelah dilakukan analisis sentimen terlihat banyak *tweet* opini yang sebenarnya tidak masuk kategori *Hatespeech* tapi diberi tagar *Hatespeech*. Nilai akurasi tertinggi didapat dengan 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%. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Support Vector Machine* (SVM) lebih tinggi akurasinya untuk klasifikasi *sentiment tweet HateSpeech* Bahasa Indonesia dibandingkan metode klasifikasi *Naïve Bayes Classifier* (NBC). Untuk penelitian selanjutnya perlu dikembangkan *stopword list* dan *stemmer* Bahasa Indonesia yang mampu meningkatkan akurasi dalam analisis sentimen.

13. Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter

Nama penulis jurnal ini adalah Fathan Hidayatullah & A.Sn dengan tahun terbit Seminar Nasional Informatika 2014 (semnasIF 2014) ISSN: 1979-2328.

Penelitian ini membahas tentang menganalisis *tweet* berbahasa Indonesia yang membicarakan tentang tokoh publik. Analisis dilakukan dengan melakukan klasifikasi *tweet* yang berisi sentimen masyarakat tentang tokoh tertentu. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*. *Naive Bayes Classifier* dikombinasikan dengan fitur untuk dapat mendeteksi negasi dan pembobotan menggunakan *term frequency* serta TF-IDF. Klasifikasi *tweet* pada penelitian ini diperoleh berdasarkan kombinasi antara kelas sentimen dan kelas

kategori. Klasifikasi sentimen terdiri dari positif dan negatif sedangkan klasifikasi kategori terdiri dari kapabilitas, integritas, dan akseptabilitas. Hasil pengujian pada aplikasi yang dibangun dan pada *tools* RapidMiner memperlihatkan bahwa akurasi dengan *term frequency* memberikan hasil akurasi yang lebih baik daripada akurasi dengan fitur TF-IDF. Metode *Support Vector Machine* menghasilkan akurasi performansi yang lebih baik daripada metode *Naive Bayes* baik dalam klasifikasi sentimen maupun dalam klasifikasi kategori. Namun demikian, secara keseluruhan penggunaan metode *Support Vector Machine* dan *Naive Bayes* sama-sama memiliki performansi yang cukup baik untuk melakukan klasifikasi *tweet*.

14. Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi *Naive Bayes*

Penulis dari jurnal tersebut adalah Natasuwarna dengan tahun terbit Jurnal Seminar Nasional Sistem Informasi dan Teknik Informatika SENSITIF 2019.

Jurnal ini membahas tentang keputusan pemerintah Indonesia mengenai pemindahan ibukota memperoleh tanggapan secara positif dan negatif oleh masyarakat di media sosial diantaranya melalui Twitter. Data analisis sentimen diambil dari opini masyarakat yang menggunakan bahasa Indonesia dari *tweet* media sosial Twitter secara *crawling*. Pencarian kata yang digunakan adalah #IndonesiaIbuKotaBaru, #IbukotaPalangkaraya, dan #YukUsulIbukotaDimana. Tahapan kegiatan penelitian terdiri dari pengumpulan data melalui media sosial Twitter, *pre-processing data*, seleksi fitur, klasifikasi dengan *Naive Bayes*, dan evaluasi. *Pre-processing data* mencakup *normalization*, *case folding*, *convert*

emoticon, *convert negation*, *tokenizing*, *stopword removal*, dan *stemming*. Seleksi fitur adalah dengan memanfaatkan *term frequency*. Sedangkan untuk evaluasi menggunakan *cross validation* dan *confusion matrix* dengan cara mendapatkan perhitungan akurasi. Tujuan penelitian ini adalah melakukan analisis sentimen opini masyarakat sehubungan dengan keputusan pemindahan ibukota Republik Indonesia dengan klasifikasi dua kelas yaitu positif dan negatif. Jumlah data yang digunakan sebanyak 200 data *tweet* yang terbagi atas 100 opini positif dan 100 opini negatif dengan menggunakan lima rasio perbandingan data latih dan data uji. Penelitian ini menghasilkan akurasi rata-rata sebesar 89,86%.

15. Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online

Menggunakan Algoritma *Support Vector Machine* Dan *Naïve Bayes*

Penulis dari jurnal tersebut adalah Hidayat dengan tahun terbit JURNAL ELEKTRONIK SISTIM INFORMASI DAN KOMPUTER (JESIK) VOL.1 No.1 Januari-Juni 2015 ISSN.

Jurnal ini membahas tentang Manfaat analisis sentimen dalam dunia politik antara lain untuk membantu dalam menganalisis kebijakan publik pemerintah serta memberikan efisiensi waktu dan efisiensi kerja bagi para penyedia berita dalam mengklasifikasikan berita dan membantu para pencari berita untuk mendapatkan wacana berita politik harian yang mereka inginkan. Proses pada analisis sentimen diawali dengan *preprocessing*, dilanjutkan dengan pembobotan kata, kemudian penghitungan *cosine similarity*, dan klasifikasi. *Preprocessing* terdiri dari beberapa tahap yaitu *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*. Metode

pembobotan kata yang digunakan adalah *Term Frequency Inverse Document Frequency* (TF-IDF) dan menggunakan *Support Vector Machine* (SVM). *Naive Bayes Classifier* (NBC), sebagai metode klasifikasinya. Adalah suatu metode pengklasifikasian berdasarkan mayoritas dari *polarity document subjectivity* yang dihasilkan dari *crawling*. Metode ini bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan *training sample*. Pengujian akurasi dari Analisis Sentimen Terhadap Wacana Politik Pada Media Masa *Online* Berbahasa Inggris dengan metode NB menghasilkan rata-rata akurasi sebesar 59,98 % dan nilai tertinggi akurasi sebesar SVM 90,50%.

16. Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi *Naïve Bayes*

Penulis dari jurnal tersebut adalah Nanang Ruhyana dengan tahun terbit Jurnal IKRA-ITH Informatika Vol 3 No 1 Maret 2019 ISSN 2580-4316.

Jurnal ini membahas tentang menganalisis hasil sentimen pada twitter untuk kebijakan ganjil genap tersebut, sehingga dapat mengelompokkan pengguna twitter menjadi opini positif dan negatif, penelitian ini menggunakan metode data mining untuk klasifikasi dengan algoritma *Naïve Bayes Classifier*, *Naive Bayes Classifier* (NBC) merupakan metode yang berdasarkan atas probabilitas *bayes* untuk melakukan pengelompokan data. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap, dalam penelitian ini menghasilkan *accuracy* 86,67%, *precision* 71,43% dan *recall* 80,00%.

17. Analisis Sentimen Persepsi Masyarakat Terhadap Pemindahan Ibukota Baru Di Kalimantan Timur Pada Media Sosial Twitter

Penulis dari jurnal tersebut adalah Icha Adellia Safra & Eri Zuliarso dengan tahun terbit Proceeding SENDIU 2020 ISBN: 978-979-3649-72-6.

Jurnal ini membahas tentang mengetahui persepsi Masyarakat Indonesia di Twitter mengenai pemindahan Ibukota dengan menggunakan kata kunci #IbuKotaBaru dan #IbuKotaPindah setelah itu data disimpan dalam database *MySQL*, lalu dilakukan proses *text processing*. Proses klasifikasi *text* dibagi menjadi kelas sentimen positif dan negatif, algoritma *Naïve Bayes Classifier* dibutuhkan untuk itu. Data yang digunakan sebesar 200 data *tweets*, yang terdiri dari 159 data *training* dan 41 data *testing* menghasilkan akurasi sebesar 78%. Visualisasi tab tabel data asli, tabel data test, *histogram*, *wordcloud* serta *confusion matrix* ditampilkan menggunakan *R Shiny*. Aplikasi web ini dapat diakses pengguna lainnya melalui internet.

18. Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda *Naive Bayes*

Penulis dari jurnal tersebut adalah Fajar Priyono, Surti Kanti, Iqbal Dzulfihar I, Imam Amirulloh, Endang Sri P, Alvi, Didi Rosiyadi dengan tahun terbit *Journal of Electrical and Electronics Engineering Vol. 1, No. 2, December 2016, Pages 38-45*.

Jurnal ini membahas tentang analisa sentimen opini ujian nasional berbasis komputer. Sampel yang di ambil sebanyak 181 kalimat sentimen yang di olah menggunakan algoritma *Naive Bayes* dengan mengelompokkan data menjadi tiga

kelas Sentimen Positif, Netral, Negatif. Hasil pengolahan data menunjukkan kelas sentimen netral memiliki nilai tertinggi sebesar 79% dan nilai terendah di peroleh kelas negatif dengan nilai 0.09%. Sedangkan tingkat akurasi ketiga kelas sentimen mencapai 100%.

19. Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada Twitter Menggunakan Metode *Naive Bayes* dan KNN

Penulis dari jurnal tersebut adalah Muhammad Syarifuddin dengan tahun terbit Lisensi CC Atribusi-NonKomersial 4.0 Internasional VOL. 15. NO. 1 AGUSTUS 2020 P-ISSN: 0216-6933 | E-ISSN: 2685-807X.

Jurnal ini membahas tentang analisis sentimen opini masyarakat guna menyelaraskan dan memberi pandangan baru mengenai suatu isu tentang *COVID-19*, metode yang digunakan adalah algoritma *Naive Bayes* dan KNN, penelitian ini berfokus pada perbandingan hasil klasifikasi metode *Naive Bayes* dan KNN, serta mengetahui kecenderungan opini masyarakat di twitter. Subjek diambil menggunakan API twitter sebanyak 1098 opini dengan kata kunci “*COVID-19*”, Dimulai dari klasifikasi opini positif atau negatif, *data cleansing*, *preprocessing*, hingga didapatkan hasil akhir. Dari hasil pengujian diketahui metode *Naive Bayes* memiliki nilai lebih tinggi, salah satunya adalah tingkat *accuracy* sebesar 63.21%, sedangkan metode KNN sebesar 58.10%, dan didapatkan pula kecenderungan opini masyarakat di Twitter condong positif, hal tersebut dapat dilihat dari jumlah opini positif sebesar 610 sedangkan negatif 488, ditunjang dengan hasil pengujian *precision* di metode *Naive Bayes* dengan nilai positif lebih tinggi dari pada negatif yaitu 66.40%: 58.94%.

20. Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode *Naive*

Bayes

Penulis dari jurnal tersebut adalah Didik Garbian Nugroho, dkk dengan tahun terbit Prosiding SNST ke-7 Tahun 2016 ISBN 978-602-99334-5-1.

Jurnal ini membahas tentang Analisis sentimen pada jasa ojek online merupakan proses mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang tentang pelayanan ojek online yang tertulis. Di media sosial masyarakat mengeluarkan beragam opini tentang pelayanan dari transportasi ini dengan jumlah yang banyak, sehingga terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian terdahulu mengenai analisis sentimen pada Twitter menggunakan metode *Naive Bayes* dengan data sebanyak 3000 data yang terdiri dari 1000 kalimat netral, 1000 kalimat positif dan 1000 kalimat negatif dengan hasil akurasi hingga 88%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen ke dalam sentimen positif, netral atau negatif serta menyalurkan opini tersebut ke setiap jasa yang bersangkutan dengan opini yang muncul. Hasil yang didapatkan dari akurasi *naive bayes* memperoleh ketepatan 80%.

Berdasarkan uraian diatas, ada beberapa persamaan antara penelitian sebelumnya dengan penelitian yang sedang dilakukan yaitu metode yang digunakan menggunakan metode Klasifikasi dan ada beberapa yang menggunakan algoritma *Naive Bayes*. Keterbaruan dari penelitian ini yang berbeda dengan penelitian sebelumnya adalah mengetahui perbandingan dari keempat data set tentang PSBB di

DKI Jakarta tersebut seberapa banyak mendapatkan sentimen positif, negatif, dan netral.

2.11 Penelitian Terkait

Penelitian terkait ini dijadikan perbandingan dengan penelitian yang akan dilakukan sehingga dapat diketahui perbedaan apa saja yang ada pada penelitian ini.

Nama penulis dari penelitian tersebut adalah Haqqizar and Larasyanti, 2019 dengan judul “Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di *Twitter* Dengan Metode *Naive Bayes*”. Tujuan dari penelitian ini adalah tentang mengetahui sentimen publik terhadap objek tertentu yang disampaikan di *Twitter* dalam Bahasa Indonesia, sehingga sangat berguna untuk melakukan riset pasar terhadap opini publik. Hasil dari penelitian pengolahan data 151 kicauan menyimpulkan bahwa sentimen menghasilkan klasifikasi sentimen negatif sebanyak 51 *tweet*, sentimen netral sebanyak 51 *tweet*, dan sentimen positif sebanyak 49 *tweet*. Tingkat akurasi dalam penentuan kategori adalah 70,21% dan rata-rata mikro 70,20% dalam penentuan sentimen memiliki tingkat presisi 70,11% dan tingkat pengembalian 70,33%. Kedekatan dari penelitian ini adalah penggunaan salah satu metode yaitu *Naive Bayes* serta menggunakan *tools* pendukung yaitu *RapidMiner*.

Nama penulis dari penelitian tersebut adalah (Fajar Priyono, Surti Kanti, Iqbal Dzulfiqar I, Imam Amirulloh, Endang Sri P, Alvi, 2017) dengan judul “Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer Menggunakan Metoda *Naive Bayes*”. Tujuan dari penelitian ini adalah tentang analisa sentimnen

opini ujian nasional berbasis komputer dengan sampel data sebanyak 181 kalimat sentimen yang diolah menggunakan algoritma *Naive Bayes* dengan mengelompokkan data menjadi tiga kelas sentimen positif, negatif, dan netral. Hasil dari penelitian ini menunjukkan kelas sentimen netral memiliki nilai tertinggi sebesar 79% dan nilai terendah diperoleh kelas negatif dengan nilai 0,09%, sedangkan tingkat akurasi ketiga kelas sentimen mencapai 100%. Kedekatan dari penelitian ini adalah pengelompokkan ketiga sentimen positif, negatif, dan netral serta penggunaan *tools* RapidMiner.

Nama penulis dari penelitian tersebut adalah (Syarifuddin, 2020) dengan judul “Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada *Twitter* Menggunakan Metode *Naive Bayes* Dan KNN”. Tujuan dari penelitian ini adalah tentang menganalisa opini masyarakat di *Twitter* akan adanya *Covid-19* dan penelitian ini berfokus pada perbandingan hasil algoritma *Naive Bayes* dan KNN dengan data sampel sebanyak 1098 opini dengan kata kunci “*COVID-19*”. Hasil dari pengujian diketahui metode *Naive Bayes* memiliki nilai lebih tinggi sebesar 63,21%, sedangkan metode KNN mendapatkan nilai sebesar 58,10% dan didapatkan pula kecenderungan opini masyarakat di *Twitter* condong positif. Kedekatan dari penelitian ini adalah penelitian pada sosial media *Twitter* dengan menggunakan salah satu algoritma yaitu *Naive Bayes* dan menggunakan *tools* pendukung yaitu RapidMiner.