

Survei Tentang Analisis Sentimen Pada Media Sosial

Putri Agung Permatasari¹, Linawati², Lie Jasa³

[Submission: 05-03-2021, Accepted: 27-05-2021]

Abstract— Social media has now become an important part of everyday life not only for personal needs but can be used in business, as well as many things that can be done. Social media used such as Facebook, Twitter, Youtube, Instagram, Likened, and Whatsapp. With the existence of social media, the amount of data available is in the form of images, comments in the form of text or emoticons, videos, and so on, so that people are free to have an opinion. With the analysis of the sentiment of opinion that is developing and many on social media, it can produce useful data and information. In sentiment analysis, data classification algorithms are needed including the Naive Bayes Classifier, Support Vector Machine, K-NN, RNN, C4.5, Lexicon Based, LDA Based Topic Modeling, and several other algorithms. This article examines some of the sentiment analysis literature on social media. Currently, the social media that is often used in analysis is Twitter and algorithm users that can increase the level of accuracy are the Naive Bayes Classifier algorithm and the Support Vector Machine. The results of calculating different data classification accuracy can be seen in the test data in the study.

Intisari—Media sosial saat ini telah menjadi bagian penting dalam kehidupan sehari-hari tidak hanya untuk kebutuhan pribadi melainkan bisa di gunakan dalam bisnis, serta banyak hal yang bisa dilakukan. Media sosial yang digunakan seperti Facebook, Twitter, Youtube, Instagram, Likened, dan Whatsapp. Dengan adanya media sosial tersebut banyaknya data yang ada berupa gambar, comment berupa text atau emoticon, video, dan lainnya, sehingga masyarakat bebas beropini. Dengan adanya analisis sentimen opini yang berkembang dan banyak di media sosial tersebut dapat menghasilkan data dan informasi yang bermanfaat. Dalam analisis sentimen diperlukannya algoritma klasifikasi data diantaranya Naive Bayes Classifier, Support Vector Machine, K-NN, RNN, C4.5, Lexicon Based, LDA Based Topic Modeling, dan beberapa algoritma lainnya. Artikel ini menelaah beberapa literature analisis sentimen pada media sosial. Saat ini media sosial yang sering digunakan dalam analisis adalah Twitter dan pengguna algoritma yang dapat meningkatkan tingkat akurasi adalah algoritma Naive Bayes Classifier dan Support Vector Machine. Hasil perhitungan akurasi klasifikasi data berbeda-beda terlihat pada data uji pada penelitian tersebut.

Kata Kunci— Analisis Sentimen, Teks Mining, Naive Bayes Classifier

¹Mahasiswa, Magister Teknik Elektro Universitas Udayana Gedung Pascasarjana Universitas Udayana, Jl. PB Sudirman Denpasar-Bali 80234 INDONESIA (telp:0361-555225; e-mail: agungputri.p@gmail.com).

^{2,3}Dosen, Magister Teknik Elektro Universitas Udayana Gedung Pascasarjana Universitas Udayana, Jl. PB Sudirman Denpasar-Bali 80234INDONESIA (telp:0361-555225) e-mail: linawati@umud.ac.id; liejasa@umud.ac.id

Dinamisnya perkembangan teknologi informasi ke dalam bentuk digital telah merubah gaya hidup manusia. Teknologi Putri Agung Permatasari: Survei Tentang Analisis Sentimen...

informasi khususnya internet semakin mudah dikonsumsi dan telah menjadi kebutuhan sebagian besar orang. Menurut Kemenkominfo Republik Indonesia, pengguna internet di Indonesia telah mencapai setidaknya 82 juta pengguna atau lebih dari 1/3 jumlah penduduk Indonesia. Besarnya jumlah pengguna internet di Indonesia menempatkan Indonesia pada peringkat ke-8 dalam hal banyaknya jumlah pengguna dibandingkan negara-negara lainnya [1]. Tujuan pengguna internet di Indonesia sangat beragam mulai dari hanya untuk berinteraksi dan berkomunikasi dengan kerabat melalui media sosial hingga meningkatkan produktivitas kerja.

Media sosial yang sering digunakan masyarakat baik untuk produktivitas kerja ataupun mendukung bisnisnya yaitu Facebook, Twitter, Youtube, Instagram, LinkedIn, Twitter, dan Whatsapp. Media sosial tersebut memiliki beberapa kemudahan untuk pengguna seperti fitur yang memungkinkan pengguna berkomentar atau menanggapi unggahan pemilik akun media sosial. Dengan media sosial pengguna bisa mengungkapkan ekspresi ataupun opininya, baik itu berupa ungkapan positif, negatif, dan netral. Tentu hal ini akan berpengaruh terhadap entitas yang diberikan opini atau komentar. Banyaknya unggahan yang ada di media sosial butuh adanya analisis sentimen, yang berfungsi untuk mengklasifikasi data yang terstruktur maupun yang tidak terstruktur.

Saat ini banyak penelitian yang berkaitan tentang analisis sentimen. Analisis sentimen pada media sosial dapat memberikan informasi yang bermanfaat bagi entitas yang diberikan opini. Analisis sentimen dapat mengklasifikasikan kalimat opini berupa kalimat positif, negatif, dan netral. Sehingga dapat menghasilkan sebuah informasi bagi entitas baik itu perusahaan maupun instansi. Untuk mendukung analisis sentimen tersebut ada berbagai metode yang digunakan diantaranya Naive Bayes Classifier, Support Vector Machine, K-NN, RNN, C4.5, Lexicon Based, LDA Based Topic Modeling, dan beberapa algoritma lainnya.

Beberapa literatur yang ada mengenai analisis sentimen khususnya terkait media sosial menjelaskan bahwa metode Naive Bayes Classifier merupakan metode sederhana yang mempunyai tingkat akurasi dari hasil klasifikasi yang tinggi dimana tingkat akurasinya dipengaruhi oleh banyaknya data uji [2]. Naive Bayes Classifier dapat dilakukan di perusahaan atau instansi yang sangat dipengaruhi oleh ekspresi ataupun opini masyarakat yang dituangkan dalam fitur media sosial seperti kolom comment.

Artikel ini mencoba menelaah literatur yang membahas tentang analisis sentimen pada media sosial. Hasil dari telaah ini diharapkan dapat membantu peneliti dan pengembang untuk mengetahui perkembangan dan tingkat akurasi klasifikasi pada beberapa algoritma yang digunakan dalam beberapa media sosial sehingga dapat mendapatkan nilai akurasi yang tinggi dan informasi yang tepat.

p-ISSN:1693 – 2951; e-ISSN: 2503-2372



Sistematika penyajian artikel ini adalah sebagai berikut: Bagian 2 memberikan informasi media sosial dan karakteristik. Bagian 3 menjelaskan mengenai metode dan bahan yang digambarkan melalui skematik penelitian. Bagian 4 memberikan tinjauan singkat tentang analisis sentimen beserta metode klasifikasi. Bagian 5 membahas tentang penelitian terkait penggunaan metode *Naïve Bayes Classifier* dan garis rencana penelitian kedepan.

II. MEDIA SOSIAL DAN KARAKTERISTIK

Media sosial merupakan media digital yang paling sering digunakan oleh masyarakat saat ini sebagai alat komunikasi jarak jauh dan dua arah [3]. Kumpulan sistem informasi dan aplikasi teknologi berbasis internet yang ada pada media sosial memungkinkan seseorang berbagi informasi dengan komunitas virtualnya. Salah satu media sosial yang terkenal adalah Twitter dan Instagram. Twitter termasuk dalam kelompok *microblogging* berbasis teks. Twitter sendiri telah berdiri lebih dari 10 tahun. Secara global, pengguna media sosial di tahun 2020 sudah mencapai 160 juta pengguna aktif [4]. Sedangkan Instagram termasuk dalam kelompok *media sharing* atau media berbagi berbasis visual seperti foto dan video. Twitter dan Instagram memiliki pengguna yang sangat besar di Indonesia.

Zimmerman dan Ng (2017) menjelaskan beberapa kelompok media sosial sebagai media pemasaran [5] :

1) *Social content-sharing service*.

- *Webblog / blog*, merupakan sebuah situs yang memiliki sistem informasi pengelolaan konten sehingga memudahkan penggunanya mengunggah (*update* dan *change*) konten secara digital. Blog dilengkapi berbagai fitur sehingga menjadikannya berfungsi sebagai media sosial seperti *comment*, *tag*, *link*, *subscription* dan sebagainya. Umumnya blog digunakan untuk menulis jurnal keseharian penulisnya ataupun tips ataupun berita yang terkait dengan industri dari pemilik *blog*. Contoh dari *blog* adalah Wordpress, Blogger, Medium dan Tumblr.
- *Media (Video dan Photo) Sharing*, merupakan jenis media sosial yang mendapatkan banyak perhatian pengguna saat ini. Hal ini disebabkan karakter pengguna kaum milenial yang sangat visual. Sehingga platform berbagi media visual mengalami *trend* yang sangat positif dalam beberapa tahun terakhir. Media yang bisa dibagikan sangat beragam dari suara, gambar dan video.

2) *Social-bookmarking services*, merupakan kelompok media sosial yang sudah cukup lama ada. Media sosial ini memiliki fitur untuk *vote/recommend* dan *bookmark* unggahan yang ada di media sosial tersebut sehingga pengguna lainnya dapat melihat daftar situs yang kita dan pengguna lain rekomendasikan. Contoh media sosial ini adalah StumbleUpon, Delicious, dan sebagainya.

3) *Social news services*, merupakan salah satu kelompok media sosial yang fokus merekomendasikan daftar artikel (*recommended list*) dari situs berita, blog ataupun website. Pengguna lainnya dapat melakukan voting terhadap daftar artikel tersebut sehingga artikel-artikel dalam daftar memiliki

masing-masing nilai atau peringkat. Contoh media sosial ini adalah Dig, Reddit dan lain-lain.

4) *Social geolocation and meeting services*, merupakan kelompok media sosial yang membagikan informasi ruang sesungguhnya (*real-space*) alih-alih ruang *cyber*. Media sosial ini memanfaatkan Global Positioning System (GPS). Contoh media sosial ini adalah Foursquare dan Meetup.

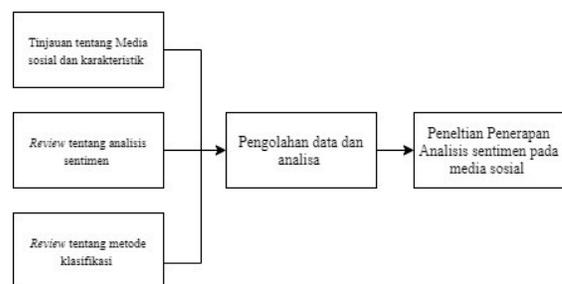
5) *Social-networking services*, merupakan salah satu kelompok sosial yang paling banyak digemari saat ini. *Social-networking* memiliki berbagi dan bertukar informasi personal seperti foto, video, pesan dan audio kepada pengguna lainnya. Dari sudut pandang bisnis, media sosial ini memiliki berbagai fitur yang memungkinkan penggunanya melakukan kegiatan pemasaran tertarget. Hal ini dimungkinkan dengan adanya *business account* dan *platform* beriklan serta *big data* yang dimiliki media sosial tersebut sehingga dapat mengklasifikasikan datanya sesuai kebutuhan perusahaan pengiklan.

- *Full networks*, contohnya Facebook dan Google+.
- *Short message network*, contohnya Twitter
- *Professional network*, contohnya LinkedIn
- *Speciality network*, contohnya Quora

6) *Community-building services*, merupakan contoh media sosial yang bertujuan untuk membangun komunitas. Media sosial ini merupakan salah satu tipe yang sudah lama. Contohnya adalah TripAdvisor, Yelp, Wikipedia dan sebagainya.

III. BAHAN DAN METODE

Penelitian ini melakukan tinjauan terhadap analisis sentimen pada media sosial. Khususnya penelitian yang dilakukan dengan banyak nya data atau *text*. Data yang digunakan berasal dari artikel jurnal, publikasi *conference*, dan dokumen lainnya. Data ini diperoleh melalui mengindeks atau mesin pencari seperti Google Scholar, ResearchGate, IEEE Xplore, dan Jurnal Unud MITE (Majalah Ilmiah Teknologi Elektro). Skematik penelitian yang dilakukan ditunjukkan pada gambar 1.



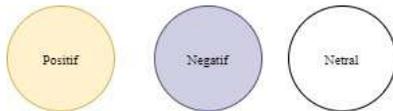
Gambar 1: Skematik Penelitian

Penelitian dilakukan melalui beberapa tahapan. Tahapan pertama melakukan tinjauan tentang media sosial dan karakteristik, *review* tentang analisis sentimen, dan *review* tentang metode klasifikasi. Selanjutnya pada tahapan kedua dilakukan pengolahan data dan analisa dari hasil telaah *literature*.

IV. ANALISIS SENTIMEN

Analisis sentimen adalah proses untuk memahami dan mengelompokan suatu kata yang dibagi menjadi beberapa kelas. Analisis sentimen juga disebut *opinion mining* yang dapat diartikan menggali opini dan emosi dari data uji. Tujuan dari analisis sentimen itu sendiri adalah untuk mengekstrak atribut dan komponen dari beberapa komentar yang ada di media sosial dan sehingga dapat menentukan beberapa kelas positif, negatif dan netral [6].

Dalam *text mining* dapat digunakan beberapa teknik pembelajaran. Diantara beberapa metode pembelajaran, *Naive Bayes Classifier* dinilai sebagai metode yang memiliki kemampuan yang baik dalam melakukan klasifikasi data dibandingkan metode klasifikasi lainnya terutama dalam hal akurasi. Metode ini dapat digunakan untuk menganalisis data berbasis teks yang sangat besar. Analisis sentimen juga dapat melibatkan pembelajaran bahasa dengan menggunakan *Part of Speech (POS) Tagging*. Metode ini memberikan label (*tag*) kata-kata pada sebuah kalimat. Berikut gambar pengelompokan klasifikasi data dari sistem analisis sentimen. Sistem analisis sentimen ditunjukkan pada gambar 2.



Gambar 2: Sistem Analisis Sentimen

A. Support Vektor Machines (SVM)

Support Vector Machines (SVM) merupakan salah satu metode pembelajaran tersupervised yang dapat menganalisis data dan mengenali pola. SVM umumnya digunakan untuk melakukan klasifikasi serta regresi. Selain itu, SVM dapat melakukan prediksi dan penilaian pada sebuah sistem. Para peneliti SVM menggunakannya sebagai pengelompokan dasar dalam menentukan superioritas *classifier* yang diusulkan [5]. SVM akan menilai volume kemunculan kata atau mengklasifikasikan polaritas dari opini atau pernyataan.

Teknik SVM akan mengelompokkan data dalam dua bagian atau lebih. Selanjutnya data yang telah dikelompokkan akan dipisahkan dengan garis dimana garis ini disebut *hyperplane*. Keberadaan SVM ditujukan untuk optimasi garis tersebut.

Jika diketahui *training vector* sebagai berikut

$X_i \in R^c, 1, \dots, n$ dalam dua kelas, dan vector $y \in \{1, -1\}^n$, metode SVM menyelesaikan permasalahan berikut :

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \quad (1)$$

Subject to

$$y_i (w^T \phi(x_i) + b) \geq 1 - C_i, C_i \geq 0, i = 1, \dots, n$$

Putri Agung Permatasari: Survei Tentang Analisis Sentimen...

Dengan elemen keduanya :

$$\min_{w,b,c} \frac{1}{2} w^T w + c \sum_{i=1}^n c_i \quad (2)$$

Subject to

$$y^T \alpha = 0$$

Dimana e $0 \leq \alpha \leq C, i = 1, \dots, n$ adalah vector dari keseluruhan, $C > 0$ adalah batas atas, ϕ adalah n dari n positif *semidefinite matrix*.

B. Naive Bayes Classifier (NBC)

Naive Bayes Classifier merupakan teknik pembelajaran algoritma *data mining*. Teknik yang digunakan didasarkan pada metode probabilitas dan statistik. Dalam klasifikasi Naive Bayes terdapat dua proses penting yaitu *learning (training)* dan *testing* [7]. Naive Bayes Classifier dapat digunakan untuk mengklasifikasi sebuah opini positif maupun negatif. Hasil dari penelitian-penelitian yang ada menunjukkan efektivitas metode Naive Bayes Classifier sebagai salah satu metode yang paling baik untuk pelatihan. Selain itu hasil klasifikasinya juga memiliki tingkat akurasi yang baik [8]. Proses algoritma klasifikasi Naive Bayes meliputi:

- 1) Input data, yang dimaksud dengan data disini yaitu data dari sumber yang ditentukan sebelumnya yang berisi data berbasis tekstual dimana umumnya berupa opini masyarakat.
- 2) Data *preprocessing*, yang merupakan tahapan awal pengolahan data berbasis tekstual dalam analisis sentimen. Di tahapan ini, digunakan beberapa metode yaitu :

- *parsing* atau *tokenizer* untuk memisahkan data tekstual menjadi beberapa token,
- *cleansing* adalah proses selanjutnya yaitu menghilangkan kata yang tidak relevan,
- *normalization* merupakan tahapan untuk normalisasi kalimat atau teks,
- *case folding* merupakan tahapan merubah bentuk dasar kata menjadi kata-kata yang berkarakter dan seragam.

C. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) merupakan salah satu metode yang digunakan untuk melakukan klasifikasi terhadap objek data pembelajaran yang memiliki jarak dekat dengan objek tersebut. K-NN adalah algoritma yang menghitung kemiripan pada tiap data uji dan semua data awal untuk menghitung daftar *nearest neighbor*-nya [9]. K-NN bertujuan untuk mengklasifikasikan objek berdasarkan atribut dan data *training*. Data klasifikasi tidak menggunakan data untuk dicocokkan dan hanya berdasarkan memori. Di dalam *query* terdapat titik yang akan ditemukan sejumlah k titik training yang posisinya paling dekat dengan titik *query*. Klasifikasi ketetangaan sebagai nilai prediksi dari sample uji yang baru.

V. TEKNIK DAN PROSES KLASIFIKASI

p-ISSN:1693 – 2951; e-ISSN: 2503-2372



Analisis sentimen dalam proses nya menggunakan metode klasifikasi untuk menentukan tingkat akurasi serta algoritma yang digunakan. Berikut beberapa teknik dan proses klasifikasi beberapa algoritma yang dapat di telaah dalam *review* artikel analisis sentimen pada media sosial.

A. Proses Klasifikasi Naïve Bayes Classifier

Klasifikasi Naive Bayes terhadap dokumen tekstual yaitu menampilkan setiap dokumen dengan atribut yang bermakna demikian sehingga $X_1, X_2, X_3, \dots, X_n$ memiliki makna dimana X_1 untuk kata pertama, X_2 adalah kata kedua, dan sebagainya. Kemudian Naïve Bayes Classifier akan mencari nilai probabilitas tertinggi dari [10]:

$$V_{MAP} = \underset{V_j}{\operatorname{argmax}} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad (1)$$

Jika nilai dari $P(X_1, \dots, X_n)$ adalah konstan untuk semua kategori V_j maka dari persamaan (1) dapat ditulis :

$$\underset{V_j}{\operatorname{argmax}} P(X_1, X_2, X_3, \dots, X_n | V_j) P(V_j) \quad (1)$$

Sehingga dari persamaan (2) dapat ditulis sebagai :

$$\underset{V_j}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \quad (2)$$

Keterangan :

V_{MAP}	: semua kategori yang diujikan
V_j	: kategori tweet, dengan :
V_1	: sentimen positif
V_2	: sentimen negatif
V_3	: sentimen netral
$P(X_i V_j)$: probabilitas X_i pada kategori
$P(V_j)$: probabilitas dari V_j

Pada saat pelatihan untuk $P(X_i | V_j)$ menggunakan rumus Chi Square yaitu :

Dan nilai $P(V_j)$ dapat dihitung menggunakan persamaan berikut (1) :

$$P(V_j) = \frac{|\text{docs } j|}{|\text{contoh}|} \quad (1)$$

Keterangan:

$|\text{docs } j|$ = jumlah dokumen setiap kategori

$|\text{contoh}|$ = jumlah dokumen dari semua kategori

$$P(X_i | V_j) = \frac{(N_{00} + N_{11} + N_{10} + N_{01})(N_{00} N_{11} + N_{10} N_{01})^2}{(N_{11} + N_{01})(N_{11} + N_{10})(N_{10} + N_{00})(N_{01} + N_{00})}$$

B. Text Mining

Text mining atau dikenal juga *text analysis* merupakan algoritma untuk mencari informasi dari data tekstual yang terstruktur. *Text mining* dapat membantu peneliti dalam memahami maksud dari sebuah pernyataan (data tekstual) tanpa harus membaca semua pernyataan tersebut. Pengelolaan data pada *text mining* terbagi dalam beberapa tahap. Tahapan ini dikenal dengan *Text Preprocessing* yang terdiri dari [11]:

- *Transform Cases*
- *Filter Stopword*
- *Filter Tokenize*
- *Stemming*

C. Proses Klasifikasi K-Nearest Neighbor

K-NN menghitung jarak atau kemiripan. Berikut adalah algoritma K-NN [12]:

1. k adalah jumlah *nearest neighbor* dan D adalah kumpulan awal
2. Untuk setiap data awal $z=(x,y)$
3. Menghitung $d(x,x)$ jarak antara z dengan tiap data $(x,y) \in D$
4. Pilih $D_z \subseteq D$, sejumlah k data dengan jarak terdekat dengan z
5. Menghitung jarak *Euclidean* antara data barudengan jarak semua data *training*. Terlampir cara perhitungan *Euclidean distance* adalah dengan persamaan sebagai berikut :

$$D(X, Y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2}$$

(1)

Keterangan :

D = Jarak Antara dua titik x dan y

X = Data Uji

Y = Sample

N = Dimensi data

6. Kemudian urutkan hasilnya berdasarkan jarak K yang ditentukan jika $K=3$ artinya dipilih 3 jarak terkecil dari hasil tersebut.
7. Selanjutnya menggunakan mayoritas atribut kelas pada 3 tetangga terdekat yang sudah dipilih untuk menentukan prediksi kelas pada data baru tersebut. Misalkan pada 3 tetangga terdekat memiliki 2 atribut kelas positif dan 1 atribut kelas negative maka kelas pada data baru tersebut adalah positif.

VI. PEMBAHASAN

Berdasarkan dari telaah yang telah dilakukan terdapat 21 buah dokumen yang membahas tentang analisis sentimen. Lebih detail dari setiap *paper* akan dijelaskan dibawah ini.

A. Penerapan Analisis Sentimen

Penelitian [13] terkait analisis sentimen yang bertujuan untuk mengetahui persepsi masyarakat terkait kenaikan harga jual rokok melalui media sosial. Media sosial yang digunakan yaitu Twitter dengan metode klasifikasi *Naïve Bayes Classifier* dan *Lexion Based*. Dengan jumlah data latihan yang sama sebanyak 350 buah dari opini masyarakat di Twitter, metode klasifikasi *Naïve Bayes Classifier* dapat memberikan nilai presentase *precision*, *recall*, dan *accuracy* yang lebih tinggi daripada metode *Lexion Based*. Namun dari hasil ketepatan pada metode klasifikasi Naïve Bayes Classifier sangat tergantung dengan jumlah data latihan yang digunakan, dan mendapatkan nilai akurasi sebesar 80%.

Penelitian [14] membahas tentang deteksi kemacetan di Jakarta. Dalam penelitian ini dinyatakan bahwa besarnya volume kendaraan serta sedikitnya ketersediaan fasilitas kendaraan umum menjadi salah satu penyebab kemacetan di Jakarta. Dengan adanya keluhan kemacetan tersebut sering kali pengguna jalan *posting* keluhan mereka di Twitter yang biasa disebut *tweets*. Setiap *tweets* dapat disimpan dan dapat diunduh untuk dianalisa. Aplikasi yang dirancang terhubung dengan Hadoop, Flume, Hive, dan Twitter API untuk melakukan *streaming* data Twitter secara *real-time*. Aplikasi ini dilengkapi berbagai fitur diantaranya fitur pertama untuk melihat isi data, fitur kedua untuk pencarian data tabel *tweet* pada tabel dan fitur

ketiga untuk menampilkan hasil analisa berdasarkan data pada tabel. Penelitian yang dilakukan mendapatkan nilai akurasi sebesar 80%.

Penelitian [15] yang membahas tentang konten media sosial Instagram yang dikelola oleh e-Commerce Shopee Indonesia dengan memanfaatkan analisis sentimen dan LDA-Based *Topic Modeling*. Dalam melakukan bisnis pada era digital memahami konsumen merupakan salah satu cara perusahaan untuk dapat terus melihat *feedback* yang diberikan konsumen terhadap produk maupun jasa yang diberikan perusahaan. Metode yang digunakan adalah menggunakan *Naive Bayes Model*, serta pembuatan *Topic Modeling* dengan menggunakan *LDA-Based Topic Modeling*. Sentimen yang terbentuk dalam konten komentar Instagram Shopee dan *Flashsale11:11* adalah negatif. Data yang dilakukan dibedakan menjadi dua yaitu data pengujian 30% dan data pelatihan 70%. Sehingga dalam proses pengkalsifikasian *Naive Bayes Classifier* memiliki tingkat akurasi sebesar 88,73%.

Penelitian [16] membahas tentang kicauan (*tweets*) masyarakat di media sosial Twitter tentang opini terhadap pemerintahan Presiden Joko Widodo. Banyak opini berupa pujian dan kritikan. Dengan menggali informasi tersebut serta melakukan klasifikasi *text data* yang diperlukan untuk analisis sentimen dengan menggunakan metode *Naive Bayes Classifier*. Penelitian ini membantu menganalisis opini masyarakat tersebut. Hasil dari penelitian ini bahwa klasifikasi pada data uji memberikan nilai akurasi sebesar 97%, nilai akurasi untuk sentimen negatif sebesar 96% serta positif sebesar 98%.

Penelitian [17] membahas analisis sentimen yang dapat dilakukan dengan menggunakan berbagai macam data mulai dari dokumen ataupun opini dari media sosial. Dalam penelitian ini menggunakan media sosial Facebook. Penelitian ini menggunakan *POS Tagging* sebelum dilakukan klasifikasi teks. *POS Tagging* yaitu proses pebelan kata yang bertujuan untuk menentukan apakah termasuk opini atau *non* opini. Pelebelan kata ini menggunakan *Rule Based* dan *Hidden Markov Model*. Sedangkan analisis sentimen yang dilakukan dalam penelitian ini menggunakan algoritma *Naive Bayes Classifier*. Hasil penelitian yang dilakukan menunjukkan nilai akurasi yaitu 87,1%.

Penelitian [18] membahas tentang memantau komentar masyarakat terhadap produk secara *real time*. Serta membahas efek metode sebelum pemrosesan teks kinerja klasifikasi sentimen dalam dua jenis tugas klasifikasi dan menyimpulkan kinerja klasifikasi tersebut. Pengukuran dari pengklasifikasi sentimen Twitter di tingkatkan sebelum pemrosesan metode yang dilakukan yaitu memperluas dan mengganti negasi. Pengklasifikasi *Naive Bayes* dan *Random Forest* lebih sensitif daripada metode yang lain. Dari pembahasan diatas penelitian ini mempelajari metode pemrosesan yang berbeda sehingga mempengaruhi klasifikasi sentimen di Twitter. Hasil dari penelitian tersebut menunjukkan dengan menggunakan algoritma *Naive Bayes Classifier* menghasilkan tingkat klasifikasi akurasi sebesar 68,5%, sedangkan dengan menggunakan algoritma *SVM* nilai akurasi yang dihasilkan sebesar 79%.

Penelitian [19] membahas tentang evaluasi opini tentang topik terkait layanan nirkabel dengan menerapkan pemrosesan bahasa alami (NLP) ke geo-tag data Twitter. Metode dekteksi Putri Agung Permatasari: Survei Tentang Analisis Sentimen...

sentimen umum saat ini bukanlah topik spesifik. Dalam penelitian ini mengembangkan kerangka kerja sentimen khusus topik layanan nirkabel baru yang menghasilkan lebih tinggi akurasi daripada kerangka kerja NLP umum. Langkah pertama memberi sentimen baru yang disebut dengan *Signal Senti Word (SSW)* dan membandingkan kinerja dengan dua pustaka lainnya. Metode yang diterapkan dalam penelitian ini terdapat tiga metode pembelajaran mesin yaitu *Naive Bayes*, *SVM* dan *Recurrent Neural Network (RNN)* untuk membangun analisis sentimen. Hasil dari penelitian ini menunjukkan *SSW* memiliki keunggulan dalam akurasi dan keahlian klasifikasi sentimen. Serta hasil dari metode pembelajaran mesin dan menilai kemampuan menunjukkan bahwa *Naive Bayes* dan *SVM* telah memperoleh manfaat yang dramatis dari pemfilteran *SSW* dan *Naive Bayes* dapat mencapai skor tertinggi 72%. Namun, *RNN* mungkin masih memiliki potensi untuk Twitter analisis sentimen dan rencana untuk melakukan pekerjaan lebih lanjut.

Penelitian [20] membahas tentang memprediksi kepribadian proaktif. Dengan 901 peserta dipilih dengan metode *cluster sampling*, teks pertanyaan jawaban singkat ditargetkan dan sosial peserta teks postingan media (Weibo) diperoleh sementara label kepribadian proaktif peserta dievaluasi oleh para ahli. Untuk membuat klasifikasi terdapat lima algoritma pembelajaran mesin disertakan *Support Vector Mesin (VSM)*, *XGBoost*, *K-Nearest-Neighbours (KNN)*, *Naive Bayes (NB)* dan *Regresi Logistik (LR)*. Tujuh indikator berbeda yang meliputi Sensitivitas (SEN), Kekhususan (SPE), Nilai Prediktif Positif (PPV), Nilai Prediktif Negatif (NPV) dan area bawah Kurva (AUC), dikombinasikan dengan validasi silang hierakis juga digunakan untuk membuat komprehensif evaluasi model. Hasil dari penelitian ini teknologi penambangan teks menunjukkan nilai yang luar biasa dalam memprediksi kepribadian proaktif individu, khususnya untuk mengidentifikasi individu dengan kepribadian proaktif rendah. Akurasi dan spesifisitas terbaik masing-masing mencapai 84,2% dan 96,9%.

Penelitian [21] membahas tentang pertumbuhan eksponensial permintaan organisasi bisnis dan pemerintah mendorong peneliti untuk menyelesaikan penelitian dalam sentimen analisis. Penelitian ini memanfaatkan empat pengklasifikasi pembelajaran mesin yang baik yaitu *Naive Bayes*, *J48*, *BFTree* dan *OneR* untuk pengoptimalan analisis sentimen. Penelitian dilakukan dengan menggunakan tiga kumpulan data yang dikompilasi secara manual. *Naive Bayes* cukup cepat namun *OneR* lebih menunjukkan dalam tingkat akurasi akurasi sebesar 91,3%. Pemrosesan metodologi yang diusulkan terbatas untuk mengekstrak kata-kata asing, *emoticon* dan memanjang kata-kata dengan sentimen yang sesuai.

Penelitian [22] ini dilakukan pada aplikasi Ruang Guru di Twitter. Untuk melakukan pengecekan keberhasilan aplikasi adalah dengan analisis sentimen. Pada penelitian ini menggunakan metode pengklasifikasi data menggunakan algoritma *Naive Bayes Classifier*, *Support Vector Machine (SVM)*, *K-Nearest Neighbour (K-NN)*, dan *feature selection* dengan algoritma *Particle Swarm Optimization (PSO)*. Berdasarkan dari hasil penelitian tersebut bahwa pengguna *feature selection PSO* dapat meningkatkan akurasi



pengklasifikasian, sehingga dari pengujian secara keseluruhan algoritma *SVM (PSO)* memiliki nilai akurasi yang paling tinggi sebesar 78,55%, *K-NN (PSO)* memiliki nilai akurasi sebesar 77,22 %, dan untuk *NBC (PSO)* nilai akurasi adalah 67,32% dan dapat diartikan klasifikasi yang dilakukan gagal.

Penelitian [23] ini mereview media massa menggunakan metode C4.5. Penelitian ini bertujuan untuk pemantauan dan meningkatkan tingkat kepuasan pada konsumen. Beberapa konsumen meluapkan opini pada media sosial seperti Facebook, Twitter atau situs media lainnya. Hasil dari penelitian ini diperoleh dengan menggunakan metode *forward selection* dan algoritma C4.5 menghasilkan nilai akurasi sebesar 84,00% lebih baik dari penelitian sebelumnya yang dimana menghasilkan nilai akurasi sebesar 80,00%. Sehingga ada peningkatan nilai akurasi dengan menggunakan algoritma C4.5.

Penelitian [24] ini melakukan analisis sentimen terhadap konten radikal, banyaknya orang yang yang berpaham radikal terhadap orang lain. Penyebaran konten radikal tersebut dilakukan pada media sosial Twitter. Dengan menggunakan metode *Support Vector Machine (SVM)* diharapkan mengetahui konten radikal positif dan konten radikal negatif. Berdasarkan penelitian tersebut tingkat hasil nilai akurasi yang paling tinggi adalah 70% dengan menggunakan data latih sebanyak 80 dokumen dan data uji yang dilakukan 20 dokumen.

Penelitian [25] ini melakukan pengambilan opini masyarakat di media sosial Twitter tentang tanggapan pemindahan ibu kota negara Indonesia dengan menggunakan metode klasifikasi *Naive Bayes Classifier*. Data yang digunakan di Twitter dengan *hashtag #IbukotaPalangkaraya* dan *#YukUsullbukotaDimana* dari data tersebut adanya tahap *Pre-Processing*. Sehingga hasil dari penelitian tersebut dengan 200 data *tweet* menghasilkan nilai akurasi rata-rata sebesar 89,86%. Dengan menggunakan data latih dan data uji 50:50, kesimpulan yang diperoleh apabila data latih lebih besar dari data uji maka hasil akurasi yang dihasilkan lebih tinggi.

Penelitian [26] ini melakukan analisis sentimen untuk mengetahui opini masyarakat tentang pelayanan maskapai penerbangan pada media sosial Twitter dengan menggunakan algoritma *Support Vector Machine (SVM)*. Sehingga dapat membantu perusahaan maskapai dalam evaluasi serta membantu masyarakat dalam pemilihan maskapai yang tepat. Hasil dari penelitian ini dengan menggunakan algoritma *SVM* adalah dengan jumlah iterasi maksimum sebanyak 50 kali dengan menggunakan *lexicon based* didapatkan akurasi paling baik sebesar 40%, *precision* sebesar 40%, *recall* sebesar 100%, dan *f-measure* sebesar 57,14%. Hasil juga dapat dipengaruhi oleh parameter yang optimal dan fitur yang digunakan.

Penelitian [27] ini dilakukan untuk mengetahui oleh-oleh favorit yang berada di Yogyakarta pada media sosial Instagram dan Twitter sehingga dapat di rekomendasikan. Penelitian ini melakukan dua metode yaitu *lexicon based* dan *Support Vector Machine (SVM)*. Hasil dari penelitian ini dengan menggunakan data uji yang sama yaitu 50. Hasil dari metode *lexicon based* tingkat akurasi yang diperoleh lebih tinggi daripada metode *SVM* dengan akurasi sebesar 87,78% namun hasil dari *recall* metode *SVM* lebih baik daripada *lexicon based* sebesar 100%.

Penelitian [28] ini bertujuan untuk mengetahui opini masyarakat di media sosial Twitter tentang viral Desa Penari dengan menggunakan metode *lexicon based*. Proses data dibagi menjadi tiga bagian yaitu positif, negatif, dan netral.

Pengumpulan data menggunakan API *search* Twitter dengan data *tweet* yang diambil Desa Penari yaitu 1000 buah komentar. Dengan menghasilkan komentar positif sebanyak 3.3%, negatif 20% dan netral sebesar 76,7%.

Penelitian [29] ini tentang opini masyarakat terhadap film pada media sosial berbahasa Indonesia dengan menggunakan metode *Naive Bayes Classifier* dengan perbaikan tidak baku. Dengan tujuan sebagai mengevaluasi dalam menonton film serta meningkatkan produksi film. Pada pengujian proses *pre-processing* dilakukan sebanyak 6 tahapan dan perbaikan kata tidak baku dengan penambahan normalisasi dapat memberikan hasil klasifikasi yang lebih baik yaitu sebesar 98,33% apabila dibandingkan dengan yang hanya dilakukan perbaikan kata baku dengan menggunakan kamus_katabaku menghasilkan nilai akurasi sebesar 91,67%.

Penelitian [30] ini tentang opini masyarakat terhadap pilkada pemilihan wakil rakyat dan diungkapkan melalui media sosial Twitter. Teknik klasifikasi untuk mengetahui opini positif dan negatif dengan menggunakan perbandingan dua metode klasifikasi *SVM* dan *k-means*. Hasil dari penelitian ini menyimpulkan bahwa metode algoritma *SVM* tingkat akurasi yang dihasilkan lebih tinggi 98% sedangkan dengan menggunakan metode *clustering* dengan algoritma *k-means* yaitu 82%.

Penelitian [31] ini bertujuan untuk mengetahui banyaknya opini masyarakat terhadap terhadap Covid -19 yang berkembang dan penyebaran yang sangat tinggi di Indonesia , serta belum adanya obat untuk menyembukan virus tersebut. Adanya opini tersebut masyarakat menungkan opini dalam media sosial. Untuk mengetahui pengklasifikasi menjadi positif dan negatif penelitian ini menggunakan perbandingan antara algoritma *Naive Bayes Classifier* dan *K-Nearest Neighbor*. Hasil dari penelitian tersebut didapatkan nilai akurasi pada algoritma *Naive Bayes Classifier* sebesar 85% dan pada algoritma *K-Nearest Neighbor* sebesar 82%.

Penelitian [32] ini dilakukan untuk mengetahui komentar masyarakat tentang pemilihan calon Gubernur Jawa Timur 2018 pada media sosial Twitter sehingga mendapatkan nilai akurasi paling optimal dan maksimal. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*. Penelitian tersebut menggunakan dua akun yang berbeda yaitu Khofifah dan Guslul. Dari hasil dataset akurasi yang paling tinggi sebesar 77% yang dilakukan di Twitter Khofifah.

Penelitian [33] bertujuan untuk mengetahui bagaimana opini masyarakat tentang pengguna aplikasi Gojek di media sosial Twitter. Metode klasifikasi data menggunakan algoritma *SVM* untuk mendapatkan opini positif, negatif dan netral. Hasil dari pelabelan data secara manual menggunakan metode *SVM* menghasilkan tingkat akurasi keseluruhan sebesar 79,19 %.

Berdasarkan hasil *review* sebelumnya menunjukkan bahwa tingkat akurasi klasifikasi penelitian berbeda, dalam pemilihan data dari media sosial dan algoritma yang digunakan.

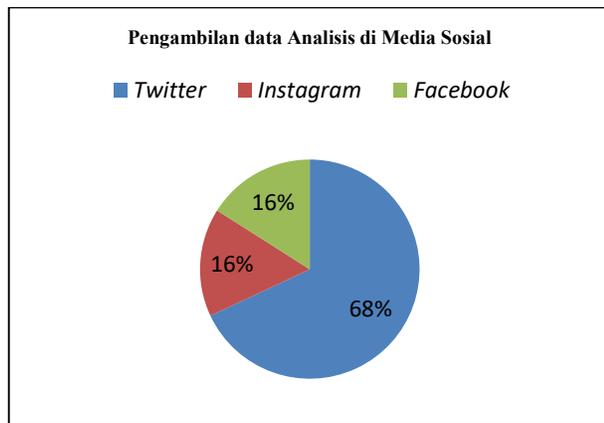
B. Media Sosial yang digunakan

Beberapa literatur yang sudah dibahas data yang sering diambil pada media sosial. Sebaran penelitian yang menggunakan data pada media sosial dapat dilihat pada tabel 1.

TABEL I
SEBARAN MEDIA SOSIAL YANG DILAKUKAN

Media Sosial	Penelitian
Twitter	[13],[14],[16],[18],[19],[22],[23],[24],[25],[26],[27],[28],[29],[30],[31],[32],[33]
Facebook	[17],[23],[27],[29]
Instagram	[15],[23],[27],[29]

Berdasarkan hasil *mapping* dari penggunaan pengambilan data analisis sentiment di media sosial, penggunaan data yang sering dilakukan dalam Analisis Sentimen adalah pada media sosial Twitter, dikarenakan teks yang digunakan dalam media sosial Twitter adalah 140 karakter dalam satu kali *tweet*. Sehingga dapat memudahkan bagi peneliti untuk mengambil data. Dengan menggunakan API (*Application Programming Interface*) memberikan kemudahan untuk mengambil data dari Twitter. Proses penyaringan kata dari hasil *tweet* sampai dataset tidak memakan waktu yang banyak. Jumlah penelitian pengguna pengambilan data pada media sosial ditunjukkan pada gambar 3.



Gambar 2: Pengambilan data Analisis di Media Sosial

Gambar 2 merupakan jumlah pengambilan data di media sosial Twitter, Instagram dan Facebook yang menunjukkan 68% lebih banyak mengambil data pada media sosial Twitter. Dalam melakukan analisis sentimen tidak hanya dilakukan di *Youtube, Web, e-commerce*. Pemilihan media sosial dalam penelitian dapat dilakukan sesuai kebutuhan sehingga data yang ada dapat membantu proses analisis sentimen.

C. Metode Klasifikasi yang digunakan

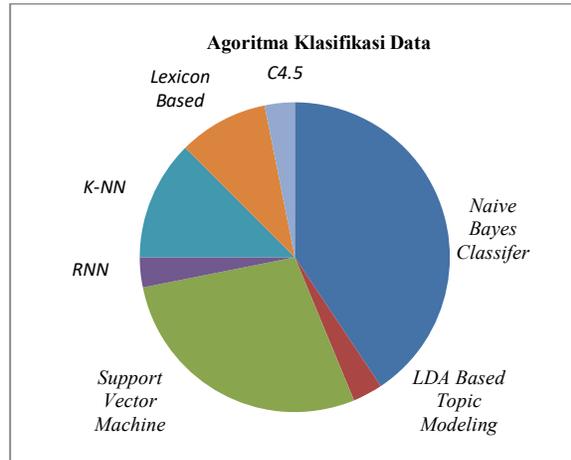
Berdasarkan dari hasil telaah *literature* yang telah dilakukan, maka dapat dilakukan *mapping* terhadap pemilihan metode dan algoritma klasifikasi data yang digunakan dilihat pada tabel 2.

TABEL 2
 ALGORITMA KLASIFIKASI DATA

Algoritma Klasifikasi Data	Penelitian
<i>Naive Bayes Classifier</i>	[13],[14],[15],[16],[17],[18],[19],[20],[21],[22],[25],[29],[31]
<i>Support Vector Machine (SVM)</i>	[18],[19],[20],[22],[24],[26],[27],[30],[33]
<i>Recurrent Neural Network</i>	[19]

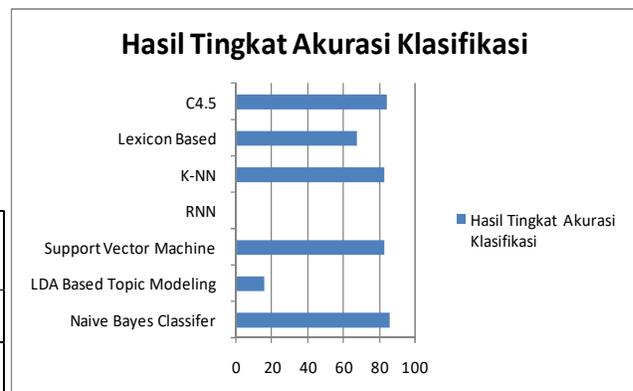
<i>LDA Based Topic Modeling</i>	[15]
<i>Lexicon Based</i>	[13],[26],[28]
<i>C4.5</i>	[23]
<i>K-Nearest Neighbor (KNN)</i>	[20],[22],[30],[31]

Tabel 2 merupakan hasil *mapping* terhadap pemilihan metode dan algoritma klasifikasi data dalam analisis sentimen. Jumlah pemilihan algoritma terlihat pada gambar 3.



Gambar 3: Metode Klasifikasi Data

Gambar 3 merupakan jumlah dari penggunaan metode dan algoritma klasifikasi data yaitu *Naive Bayes Classifier*, *LDA Based Topic Modeling*, *Support Vector Machine*, *RNN*, *K-NN*, *Lexicon Based* dan *C4.5*. Berdasarkan *literature* yang ada adalah *Naive Bayes Classifier*. Menurut Andi (2018) mengatakan bahwa algoritma *Naive Bayes Classifier* memiliki nilai akurasi yang sangat baik daripada algoritma sebelumnya ini dikarenakan algoritma tersebut mudah digunakan, hanya membutuhkan proses satu kali iterasi data *training*, dan jumlah data uji bisa dilakukan dalam jumlah kecil [34].



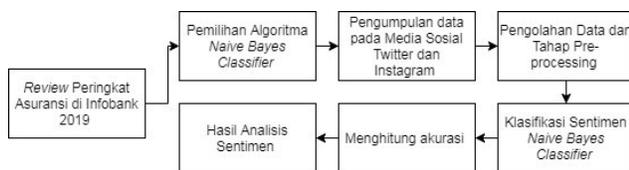
Gambar 4: Hasil Rata-Rata Nilai Akurasi



Gambar 4 merupakan hasil akurasi rata-rata dari algoritma yang digunakan. Algoritma *Naive Bayes Classifier* menghasilkan nilai akurasi rata-rata sebesar 85,08%. Beberapa literatur yang sudah dibahas untuk mendapatkan nilai akurasi tinggi dipengaruhi juga pada jumlah data latih, sehingga akurasi yang dihasilkan maksimal. Menurut Mochammad Haldi mengungkapkan bahwa nilai akurasi juga dapat dipengaruhi apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi akan menghasilkan nilai yang sama [35]. Sehingga untuk pemakaian algoritma *Naive Bayes Classifier* dapat dilihat dari keperluan data apa saja yang akan diolah. Algoritma *Naive Bayes* sangat baik dalam analisis teks data yang banyak tidak diperkenankan untuk tampilan gambar. Selain algoritma *Naive Bayes Classifier* dari review penelitian sebelumnya yang banyak digunakan adalah algoritma *Support Vector Machine* pemakaian algoritma tersebut dapat menghasilkan nilai akurasi yang maksimal. Seperti penelitian pada pemantauan opini masyarakat terhadap produk yang menggunakan perbandingan dengan algoritma *NBC* dan *SVM* hasil dari penelitian tersebut algoritma *SVM* lebih tinggi nilai akurasi dibanding *NBC*. Ini dikarenakan data yang diambil tidak hanya berupa teks melainkan gambar emosional atau *emoticon* dapat dijadikan data.

D. Rencana Penelitian

Beberapa penelitian yang telah dilakukan dalam penerapan analisis sentimen pada media sosial. Penelitian dilakukan dengan memilih data di media sosial Twitter dan Instagram. Selain itu pengguna algoritma yang digunakan *Naive Bayes Classifier* untuk menghasilkan tingkat akurasi yang maksimal dan bahasa *code* pemrograman yang dapat dimengerti. Berdasarkan hal tersebut, maka muncul sebuah rencana penelitian dengan melakukan Analisis Sentimen Masyarakat Terhadap Perusahaan Asuransi Terbaik di Infobank 2019 Berdasarkan Opini dari Twitter dan Instagram Menggunakan *Naive Bayes Classifier*. Tahapan dari rencana penelitian ditunjukkan pada Gambar 5.



Gambar 5: Alur Rencana Penelitian

Dalam rencana penelitian ini, tahapan pertama dilakukan *review* peringkat asuransi di Infobank 2019, penelitian yang dilakukan hanya ke 10 asuransi di Indonesia. Selanjutnya tahapan kedua dilakukan pemilihan algoritma *Naive Bayes Classifier*. Pada tahapan ketiga dilakukan pengumpulan data pada Media Sosial di Twitter dan Instagram dengan menggunakan *hashtag* ke nama 10 perusahaan asuransi. Tahapan keempat dari data yang sudah diambil di media sosial akan diolah data dan melakukan tahap *pre-processing*. Tahap kelima dilakukan klasifikasi sentimen melalui algoritma *Naive Bayes* sehingga mendapatkan kelas positif, negatif, dan netral. Tahap keenam melakukan perhitungan hasil akurasi klasifikasi

data sehingga mendapatkam hasil analisis sentimen terhadap perusahaan Asuransi.

VII. KESIMPULAN

Berdasarkan dari hasil telaah yang dilakukan, dalam melakukan analisis sentimen pada media sosial dapat dilakukan pemilihan media sosial untuk mencari opini dan data yang diperlukan sehingga mendapatkan informasi yang berguna. Selanjutnya pemilihan algoritma klasifikasi yang tepat sehingga proses perhitungan akurasi mendapatkan hasil yang maksimal.

Algoritma yang digunakan seperti *Naive Bayes* dan *Support Vector Machine* dapat meningkatkan hasil akurasi yang sangat tinggi. Dimana rata-rata *literature* menggunakan algoritma tersebut dengan pengambilan data di media sosial Twitter. Hasil akurasi dari masing-masing algoritma berbeda-beda, perbedaan terlihat pada pengambilan data uji disetiap penelitian.

VIII. REFERENSI

- [1] Kemkominfo : Pengguna Internet di Indonesia, dari URL :https://kominfo.go.id/index.php/content/detail/3980/Kemkominfo%3A+Pengguna+Internet+di+Indonesia+Capai+82+Juta/0/berita_satker
- [2] Afshoh, Fauziah. 2017. Analisa Sentimen menggunakan Naive Bayes untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter . Universitas Muhammadiyah, Surakarta.
- [3] Sekawanmedia (2020): Pengaruh Pengguna Media Sosial dan manfaat untuk bisnis, dari URL:<https://www.sekawanmedia.co.id/media-sosial-untuk-bisnis/>
- [4] S. Kamran, M. Shaikh, A. Naseem, and P. Kamble, "Exploiting Social Media Data for Traffic Monitoring Using the Techniques of Data Mining," *Int. J. Innov. Res. Comput. Commun*
- [5] Zimmerman, Jan, dan Ng, Deborah for Dummies. 2017. *Social Media Marketing : All-in-One*. Canada: John Wiley & Sons.
- [6] Kadek Ary B., Made Sudarma, dan Wayan G. A." Analisis Sentimen Pada Video di Media Sosial Youtube Menggunakan *STRUCT-SVM*". *Majalah Ilmiah Teknologi Elektro*. Vol. 18.No1. Jan-April 2019. DOI: <https://doi.org/10.24843/MITE.2019V18i01.P17>.ISSN: 1693-2951.
- [7] Informatikalogi.com dari URL: <https://informatikalogi.com/algoritma-naive-bayes/>
- [8] V. Ambassador Flores, Lic Jasa, dan Linawati. "Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram". *Majalah Ilmiah Teknologi Elektro*. Vol 19. No 1 Jan-Jun 2020. DOI: <https://doi.org/10.24843/MITE.2020.v19i01.P07>
- [9] Elok N.H., dan M. Balya I.A." Pengembangan Sistem Analisis Sentimen Berbasis Java Pada Data Twitter Terhadap Omnibus Law Menggunakan Algoritma *Naive Bayes* dan *K-Neares Neighbor (K-NN)*". *JIP* Vol.7. Feb 2021. ISSN: 2614-6371.
- [10] I Komang Dharmendra, K. Oka Saputra, N. Pramaita. " Analisa Sentiment Untuk Opini Alumni Pada Perguruan Tinggi" *Majalah Ilmiah Teknologi Elektro*, Vol.18, No. 2, Mei – Agustus 2019. DOI: <https://doi.org/10.24843/MITE.2019.v18i02.P11>
- [11] M. Akbar Maulana, A. Setyanto, dan M.P Kurniawan. "Analisis Sentimen Media Sosial Universitas AMIKOM Yogyakarta sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM. *ISSN : 2302-3805, 2018.*
- [12] G. Ayu Vida Matrika Giri " Klasifikasi Musik Berdasarkan *Genre* dengan Menggunakan K-Nearest Neighbor" *Jurnal Ilmu Komputer* Vol. XI No2. P-ISSN: 1979-5661. E-ISSN: 2622-321X. 2018.
- [13] Afshoh, Fauziah. 2017. Analisa Sentimen menggunakan Naive Bayes untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter . Universitas Muhammadiyah, Surakarta.
- [14] Nadika, Tifani. 2018 . Analisa Sentimen Menggunakan Data Twitter, Flume , Hive , Pada Hadoop dan Java untuk Deteksi Kemacetan di Jakarta.Sekolah Tinggi Ilmu Ekonomi (STIE), Jakarta.

- [15] Kabiru, Irawan Noor, dkk. 2019. Analisa Konten Media Sosial E-Commerce pada Instagram Menggunakan Metode Sentimen Alysis dan LDA-Based Topic Modeling (Studi Kasus : Shopee Indonesia). Universitas Telkom, Bandung.
- [16] Mahardhika, Yonathan Sari, dan Zuliarso, Edri. 2018. Analisis Sentimen Terhadap Pemeritahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Classifier. Univeritas Stikubank, Semarang.
- [17] Putu Sri Merta, Linawati, dan S. Komang Oka. 2019. "Pengguna Metode *Naive Bayes Classifier* pada Analisis Sentimen *Facebook* Berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, P-ISSN: 1693-2951, 2019. DOI: <https://doi.org/10.24843/MITE.2019.v18i01.P22>.
- [18] Z. Jianqiang and G. Xiaolin, 2017. *Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis*. *IEEE Access* vol 5.
- [19] W. Qi, R. Procter, J. Zhang and W. Guo. 2019. " *Mapping Consumer Sentiment Toward Wireless Services Using Geopatial Twitter Data*". *IEEE Access* Vol 7.
- [20] P. Wang, Y. Yuan, Y. Si, G. Zhu, X. Zhan, J. Wang, and R. Pan. " *Classification of Proactive Personality : Text Mining Based on Weibo Text and Short-Answer Questions Text*". *IEEE Access* vol 8. May 2020.
- [21] Jaspreet, S, Gurvinder, S, and Rajinifer, S. " *Optimization of Sentiment Analysis Using Machine Learning Classifiers*. DOI 10.1186/s13673-017-0116-3 open Access 2017.
- [22] Angelina P.G, dkk. " Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi" *Jurnal TEKNOINFO* vol.14 No.2, 2020. ISSN: 2615-224X.
- [23] Arif, R dan M. Rifqi Tsani. " Analisis Sentimen *Review* Media Massa Menggunakan Metode C4.5 Berbasis *Forward Selection*" . *Smart Compt* Volume 8.No2. Juni 2019, P-ISSN: 2089-676X.
- [24] Ferdi A, Indriati, dan Putra P.A. " Analisis Sentimen Konten Radikal di Media Sosial Twitter Menggunakan Metode *Support Vector Machine (SVM)*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol.3 No.1. Januari 2019. E-ISSN: 2548-964X.
- [25] Amar P. Natasuwarna. " Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes Classifier". *Seminar Nasional Sistem Informasi dan Teknik Informatika Sensitif 2019*.
- [26] Arsyia M.P, Imam C, dan Putra.P. " Analisis Sentimen tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan *i Support Vector Machine (SVM)* ". *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol3. No.3, Maret 2019, hlm 2789-2797. E-ISSN: 2548-964X.
- [27] Hidayatullah, H, Debby G.P, Wilis K. "Metode *Lexicon Based* dan *Support Sector Machine* Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit". *Seminar Nasional Inforematika 2018*. UPN "Veteran" Yogyakarta. ISSN: 1979-2328.
- [28] Rifiana Arief dan Karel Imanuel. " Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter dengan Metode *Lexicon Based* ". *Jurnal Ilmiah Matrik*, Vol 21 No.3, Desember 2019. ISSN: 1411-1624. e-ISSN: 2621-8089.
- [29] Prananda A, Rizal S., M. Ali F. " Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku". *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol1. No.12. Desember 2017. hlm 1733-1741. e-ISSN: 2548-964X.
- [30] Atika R., Aris M., dan Junta Z. " Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksana Pilkada Serentak Menggunakan Algoritma Support Vektor Machine". *Jurusan Teknik Informatika*. Universitas Dian Nuswanto 2017. ISSN : 1978-8282.
- [31] Ardime L. F., Rima D.R., dan Nia Annisa F.T. " Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter" Vol.1x No.1 (2021) FEB 2021. ISSN Media Elektronik: 12345-XYZ
- [32] Ghulam Asrofi Buntoro. " Analisis Sentimen calon Gubernur Jawa Timur 2018 dengan Metode Naive Bayes Classifier". *JIPN*. Volume 4. No.1 Maret 2019. e-ISSN 2541-3724.
- [33] Nur Fitriyah, Budi W., dan Di Asih I.M. " Analisis Sentimen Gojek Pada Media Sosial Twitter dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Gaussian*, Vol. 9, No. 3. 2020. ISSN: 2339-2541.
- [34] Andi Krisna Dewmawan dari URL : <https://www.adhikrisnadermawan.com/2018/07/27/klasifikasi-naive-bayes/>



{ Halaman ini sengaja dikosongkan }