

BAB II

LANDASAN TEORI

2.1 Sentimen Analisis

“Sentimen analisis adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan menggunakan teknik analisis teks” [16]. Sentimen analisis dapat dibagi menjadi 4 tipe yaitu :

1. *Fine-grained sentiment analysis*

Fine-grained sentiment analysis adalah tipe analisis yang memiliki penilaian spesifik yang biasanya banyak digunakan pada *review e-commerce* berupa lima bintang yang dimana bintang lima menyatakan sangat baik sedangkan bintang satu menyatakan sangat buruk[16]

2. *Emotion detection*

“Tipe analisis sentimen ini bertujuan untuk mendeteksi emosi, seperti kebahagiaan, kemarahan, kesedihan, dan frustrasi. Namun, salah satu kelemahan dari *emotion detection* adalah cara pengekspresian emosi orang yang berbeda-beda.” Sebagai contoh, kata ‘parah’ sebenarnya bermakna negatif. Namun jika seseorang mengatakan, “Makanan ini enak banget, parah,” maknanya menjadi positif. [16]

3. *Aspect-based sentiment analysis*

“*Aspect-based sentiment analysis* digunakan untuk mengetahui aspek apa yang mendapat penilaian positif, netral, atau negatif dari pelanggan. Misalnya dalam sebuah *review* produk, ada pelanggan yang memberikan komentar, “Bahan celana ini halus dan nyaman digunakan.” Maka bisa disimpulkan bahwa aspek yang mendapat penilaian positif dari pelanggan adalah bahan celana tersebut.” [16]

4. *Multilingual sentiment analysis*

Analisis sentimen *multilingual* digunakan untuk menganalisis kata-kata dalam berbagai bahasa, tetapi tipe analisis ini tidak mudah karena dibutuhkan daftar kata dari bermacam bahasa dan terus meng-*update* daftar tersebut sesuai perkembangan bahasa. [16]

2.2 *Data Mining*

Data mining merupakan proses pengumpulan informasi dari data yang berjumlah besar. *Data mining* memiliki banyak fungsi, namun fungsi utamanya yaitu fungsi *descriptive* dan fungsi *predictive*. “Fungsi deskripsi dalam data mining adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati”. Dengan melakukan fungsi tersebut diharapkan data tersebut bisa digunakan dalam perkembangannya dan mudah diketahui karakteristik dari data tersebut yang dimana bisa di temukannya

suatu pola tersembunyi pada data tersebut dan pola tersebut yang bisa dibilang sebagai karakteristik dari sebuah data. [17]

Fungsi prediksi merupakan fungsi yang bertujuan untuk menemukan pola tertentu dari suatu data yang pola tersebut didapatkan dari berbagai variabel yang ada pada data tersebut. Ketika pola didapatkan, maka pola tersebut digunakan untuk memprediksi variable lainnya yang belum diketahui nilai atau jenisnya.

2.3 *Text Mining*

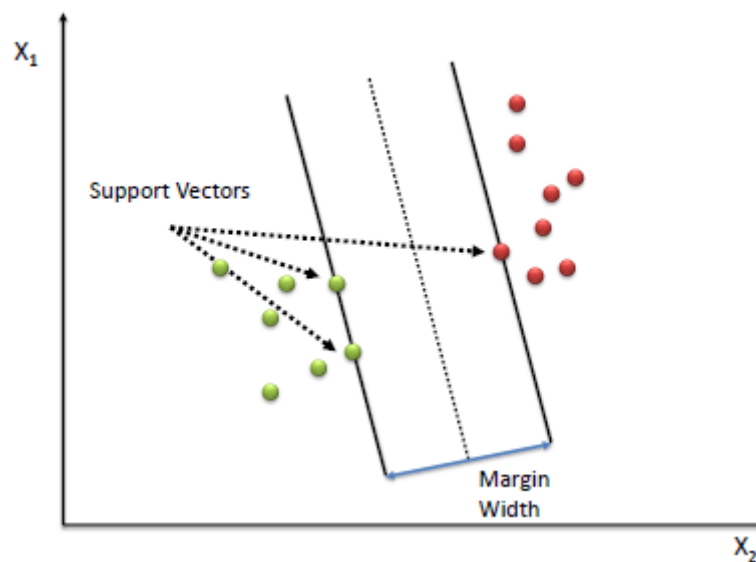
“Text mining adalah “proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data”. *Text Mining* telah mengalami pengembangan dikarenakan banyaknya platform data besar dan algoritma pembelajaran yang dapat menganalisis data yang tidak terstruktur secara besar.” [18]

2.4 *Support Vector Machine*

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti *Support Vector Classification*) dan regresi (*Support Vector Regression*). Dalam pemodelan klasifikasi, *SVM* memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik

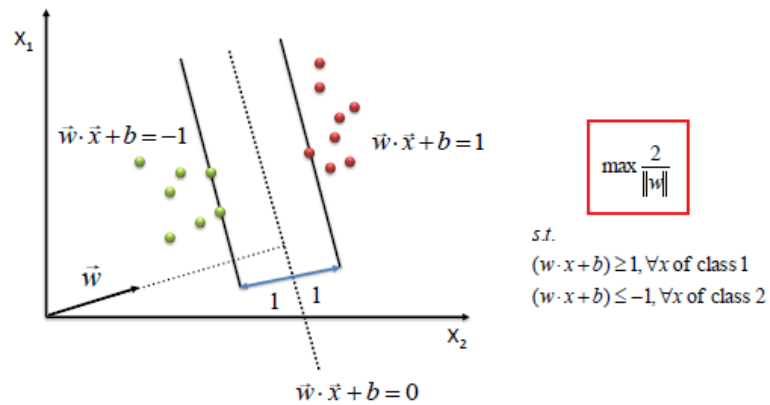
klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan *linear* maupun *non linear*. [19]

Support Vector Machine (SVM) melakukan klasifikasi dengan menemukan *hyperplane* yang memaksimalkan margin antara dua kelas. Vektor yang mendefinisikan *hyperplane* adalah vektor pendukung [20].

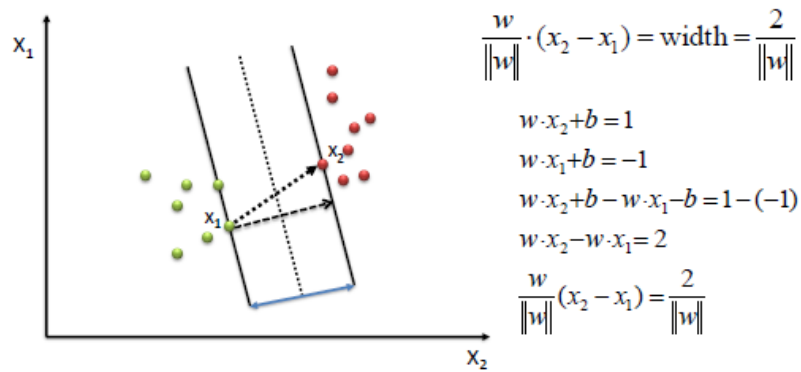


Rumus 2. 1 Rumus SVM

Untuk menentukan *hyperplane* optimal diperlukannya untuk memaksimalkan lebar margin (w).



Rumus 2. 2Rumus Pekerjaan untuk SVM



Rumus 2. 3 Rumus Lengkap untuk SVM

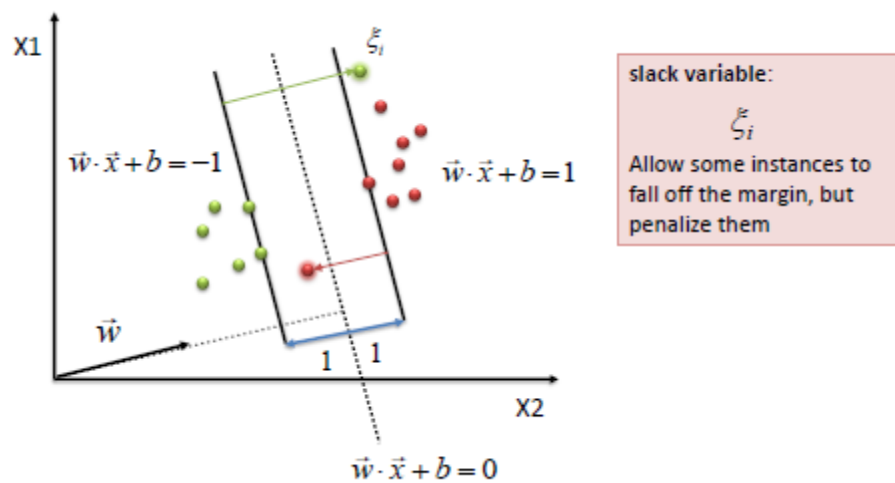
Untuk menemukan w dan b dengan menyelesaikan fungsi tujuan berikut menggunakan Pemrograman Kuadratik.

$$\min \frac{1}{2} \|w\|^2$$

$$s.t. y_i (w \cdot x_i + b) \geq 1, \forall x_i$$

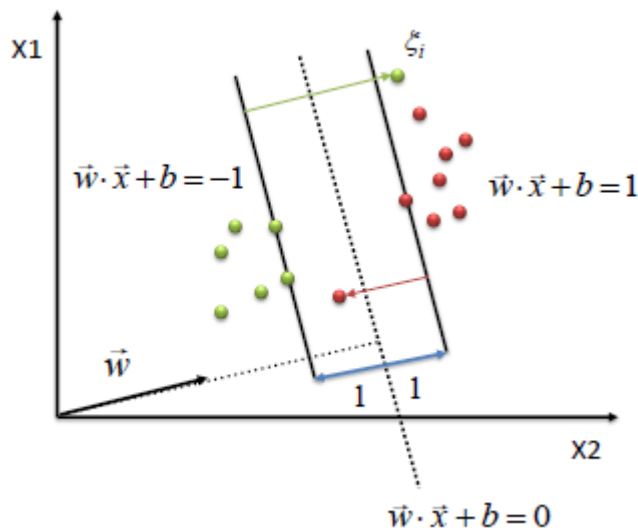
Rumus 2. 4 Rumus Kuadratik dari SVM untuk mencari w dan b

Keunggulan *SVM* adalah jika datanya dapat dipisahkan secara linier, terdapat nilai minimum global yang unik. Analisis *SVM* yang ideal harus menghasilkan *hyperplane* yang benar-benar memisahkan vektor (kasus) menjadi dua kelas yang tidak tumpang tindih. Namun, pemisahan yang sempurna mungkin tidak dapat dilakukan, atau dapat mengakibatkan model dengan begitu banyak kasus sehingga model tidak dapat diklasifikasikan dengan benar. Dalam situasi ini *SVM* menemukan *hyperplane* yang memaksimalkan margin dan meminimalkan kesalahan klasifikasi



Rumus 2. 5 Proses Pekerjaan SVM

Algoritme mencoba mempertahankan variabel slack ke nol sambil memaksimalkan margin. Namun, itu tidak meminimalkan jumlah kesalahan klasifikasi tetapi jumlah jarak dari *hyperplanes* margin.



Constraint becomes :

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \forall x_i$$

$$\xi_i \geq 0$$

Objective function

penalizes for misclassified instances and those within the margin

$$\min \frac{1}{2} \|w\|^2 + C \sum_i \xi_i$$

C trades-off margin width and misclassifications

Rumus 2.6 Proses Pekerjaan SVM Beserta Perhitungan

Cara termudah untuk memisahkan dua kelompok data adalah dengan garis lurus (1 dimensi), bidang datar (2 dimensi), atau bidang-N dimensi. Namun, ada situasi di mana kawasan nonlinier dapat memisahkan grup secara lebih efisien. SVM menangani ini dengan menggunakan fungsi *kernel (nonlinier)* untuk memetakan data ke dalam ruang yang berbeda di mana bidang *hiper (linier)* tidak dapat digunakan untuk melakukan pemisahan. Artinya, fungsi non-linier dipelajari oleh mesin pembelajaran linier dalam ruang fitur berdimensi tinggi sedangkan kapasitas sistem dikontrol oleh parameter yang tidak bergantung pada dimensi ruang. Ini disebut trik *kernel* yang berarti fungsi *kernel* mengubah data menjadi ruang fitur berdimensi lebih tinggi untuk memungkinkan dilakukannya pemisahan *linier*.

2.5 Naïve Baiyes

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes . Metode pengklasifikasian dg menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes . Ciri utama dr *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian [21].

Teorema Bayes memberikan cara untuk menghitung probabilitas posterior, $P(c/x)$, dari $P(c)$, $P(x)$, dan $P(x/c)$. Pengklasifikasi *Naïve Bayes* mengasumsikan bahwa pengaruh nilai suatu prediktor (x) pada kelas tertentu (c) tidak bergantung pada nilai prediktor lain. Asumsi ini disebut kebebasan bersyarat kelas. [22]

The diagram shows the Bayes' theorem formula:
$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$
 with four blue arrows pointing from labels to parts of the formula: 'Likelihood' points to $P(x | c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c | x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Rumus 2. 7 Rumus dari Naïve Bayes

- $P(c | x)$ adalah probabilitas posterior kelas (target) yang diberi prediktor (atribut).

- $P(c)$ adalah probabilitas prior kelas.
- $P(x | c)$ adalah kemungkinan yang merupakan probabilitas kelas prediktor yang diberikan.
- $P(x)$ adalah probabilitas prediktor sebelumnya

Pada model *ZeroR* tidak ada prediktor, pada model *OneR* akan mencoba mencari satu prediktor terbaik, *Naïve Bayes* memasukkan semua prediktor yang menggunakan aturan *Naïve Bayes* dan asumsi independensi antar prediktor.

2.6 *Random Forest*

“*Random Forest* adalah classifier yang terdiri dari classifier yang berbentuk pohon $\{h(x, \theta_k), k = 1, \dots\}$ dimana θ_k adalah random vector yang didistribusikan secara independen dan masing masing *tree* pada sebuah unit kan memilih class yang paling populer pada input x .” [23]

Saat menggunakan Algoritma *Random Forest* untuk memecahkan masalah regresi, akan digunakan *mean squared error (MSE)* untuk untuk menghasilkan cabang dari setiap node berdasarkan data.

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2$$

Rumus 2. 8 Rumus dari MSE

- Dimana N adalah jumlah titik data,
- f_i adalah nilai yang dikembalikan oleh model
- y_i adalah nilai sebenarnya untuk data titik i

Rumus ini menghitung jarak setiap node dari nilai sebenarnya yang diprediksi, membantu memutuskan cabang mana yang merupakan keputusan yang lebih baik untuk hasil dari *Random Forest*. Di sini, “yi” adalah nilai titik data yang uji pada simpul tertentu dan fi adalah nilai yang dikembalikan oleh pohon keputusan.

Saat melakukan *Random Forests* berdasarkan data klasifikasi, *Random Forest* akan menggunakan indeks *Gini*, atau rumus yang digunakan untuk memutuskan bagaimana node pada cabang pohon keputusan.

$$Gini = 1 - \sum_{i=1}^C (p_i)^2$$

Rumus 2. 9 Rumus dari Gini

Rumus ini menggunakan kelas dan probabilitas untuk menentukan *Gini* setiap cabang pada node, menentukan cabang mana yang lebih mungkin terjadi. Di sini, “pi” mewakili frekuensi relatif kelas yang amati dalam dataset dan “c” mewakili jumlah kelas. Alternatif lainnya yaitu menggunakan *entropi* untuk menentukan bagaimana *node* bercabang di pohon keputusan

$$Entropy = \sum_{i=1}^C -p_i * \log_2(p_i)$$

Rumus 2. 10 Rumus dari Entropi

Entropy menggunakan probabilitas hasil tertentu untuk membuat keputusan tentang bagaimana node harus bercabang. Berbeda dengan indeks *Gini*, indeks ini lebih intensif matematis karena fungsi logaritmik yang digunakan dalam penghitungannya.

2.7 Google Play

Google Play Service adalah paket *application programming interface (API)* yang memayungi beberapa aplikasi yang terpasang di *Android*. *Google Play Service* berfungsi sebagai pelayan sinkronisasi akun-akun *Google*, seperti *e-mail Gmail*, peta *Google Maps*, dan melacak lokasi (*GPS*) [24].

2.8 Aplikasi Bibit

“Bibit adalah aplikasi reksa dana untuk membantu investor pemula mulai berinvestasi” yang menggunakan hasil riset pemenang *Nobel Prize*, yang mampu menyesuaikan investasi para pengguna di reksa dana yang sesuai dengan kemampuan pengguna dalam melakukan investasi [25]. Didalam aplikasi Bibit terdapat sebuah portofolio yang akan dibuatkan sesuai dengan kondisi penggunaannya dan proses pembuatan portofolio untuk pengguna sebagai berikut [26]:

1. Auto Risk Profiling: pengguna diharuskan menjawab 6 pertanyaan mengenai kondisi keuangan pengguna agar Bibit

dapat menentukan profil risiko yang cocok dengan pengguna menggunakan algoritma tanpa bias.

2. Auto Financial Plan: Pengguna akan mendapatkan portofolio yang paling sesuai dengan jawaban yang telah dijawab.
3. Auto Rebalancing: Bibit akan secara otomatis melakukan alokasi yang optimal sesuai dengan perubahan yang ada seperti, usia, profil risiko, dan kondisi pasar.

Setelah setelah semua rekomendasi didapatkan oleh pengguna, setelah itu pengguna dapat melakukan investasi secara online pada platform Bibit yang memiliki manfaat sebagai berikut :

1. “Gratis Biaya Komisi. Semua pembelian di Bibit tidak dipotong biaya transaksi supaya keuntungan bisa lebih tinggi. Bibit dibayar oleh manajer investasi.”
2. “Mulai dari Rp 10.000. Investasi sudah seharusnya untuk semua kalangan. Mulai tanam duit kamu dari kecil atau sebanyak yang kamu inginkan. Lebih cepat kamu mulai, lebih besar pertumbuhan uang kamu.”
3. “Cairkan Kapan Saja Tanpa Penalti. Reksadana dapat dicairkan kapan saja dengan cepat tanpa kena penalti.”
4. “Auto Rebalancing. Alokasi portofolio saat ini berbeda dengan target alokasi yang optimal. Maka Bibit membantu

melakukan auto-rebalancing supaya lebih mendekati portofolio optimal.”

2.9 *Confusion matrix*

Confusion matrix adalah pengukuran performa untuk klasifikasi pembelajaran mesin yang dimana keluaran hasil kelasnya bisa lebih dari 2 kelas [27]. Terdapat empat istilah representasi hasil proses pada klasifikasi *confusion matrix* yaitu *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif* yang dimana contohnya akan diambil dengan analogi kehamilan yaitu :

1. *True Positive (TP)*:

Menghasilkan prediksi positif dan itu benar.

Memprediksi bahwa seorang wanita hamil dan wanita tsb memang benar hamil.

2. *True Negative (TN)*:

Menghasilkan prediksi negatif dan itu benar.

Memprediksi bahwa seorang pria tidak hamil dan benar karena pria tidak mungkin hamil.

3. *False Positive (FP)*: (Kesalahan Tipe 1)

Menghasilkan prediksi positif dan itu salah.

Memprediksi bahwa seorang pria hamil tetapi tidak mungkin pria bisa hamil.

4. *False Negative (FN)*: (Kesalahan Tipe 2)

Menghasilkan prediksi negatif dan itu salah.
 Memperkira bahwa seorang wanita tidak hamil
 tetapi sebenarnya wanita tersebut hamil.

2.10 Penelitian Terhadulu

Tabel 2. 1 Tabel Penelitian Terdahulu

Jurnal	Permasalahan	Metode dan Hasil	Kesimpulan
Nama Jurnal: Seminar Nasional Teknologi Informasi dan Multimedia 2018 Judul Jurnal: ANALISIS SENTIMEN TERHADAP OPINI MASYARAKAT INDONESIA MENGENAI BUKALAPAK Penulis: Wivadario Wilana, Sandy	Bagaimana Opini mayoritas masyarakat Indonesia terhadap Bukalapak ?	Metode : metode <i>Naïve Bayes</i> digambarkan dalam bentuk grafik, Hasil: “Terdapat 510 total data yang diolah dan menghasilkan output yang bernilai netral dengan probabilitas 50%, positif 37%, dan negatif 13%.”	“Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap opini masyarakat dilakukan dengan metode <i>Naïve Bayes</i> . Metode <i>Naïve Bayes</i> digunakan karena lebih akurat dalam klasifikasi data-data yang sudah diolah. Probabilitas yang dihasilkan metode <i>Naïve Bayes</i> dengan jelas menyatakan bahwa opini sentimen

Hanshe Hanasbey, Meylan Ribka Awinero, Jonathan Vorgorius Adeputra Modouw, Melkior N. N Sitokdana			masyarakat terhadap Bukalapak bernilai netral. Nilai probabilitas yang dihasilkan dari metode <i>Naïve Bayes</i> adalah 37%(Positif), 50%(Netral), dan 13%(Negatif).”
Nama Jurnal: Jurnal Ilmiah Ekonomi Bisnis Volume 24 No.3 Desember 2019 Judul Jurnal: ANALISIS HUBUNGAN ANTARA SENTIMEN INVESTOR DAN IMBAL HASIL PASAR SAHAM DENGAN PENDEKATAN ALIRAN DANA REKSA DANA DAN ANALISIS VECTOR	Apakah ada hubungan kausalitas dan dinamis antara sentimen investor dan imbal hasil pasar dengan melibatkan faktor inflasi ?	Metode: Hasil analisis menunjukkan Metode yang digunakan berbasis kuantitatif dengan analisis deskriptif, dengan menggunakan model Vector Autoregressive, analisis Fungsi Respons Impuls dan Variance Decomposition serta analisis Kausalitas Granger. Hasil: adanya “hubungan antara	“Berdasarkan uraian pada bagian pembahasan dapat disimpulkan bahwa terdapat hubungan antara sentimen investor dengan imbal hasil indeks pasar, namun pengaruhnya tidak signifikan. Pada hasil dari sisi IRF, respons yang fluktuatif dan cukup signifikan terlihat pada respons imbal hasil indeks pasar terhadap kejutan pada sentimen investor, sedangkan hasil dari

<p>AUTOREGR ESSIVE (VAR)</p> <p>Nama Penulis: Anggia Paramita Puti Kencana</p>		<p>sentimen investor dengan imbal hasil pasar, namun dampaknya tidak signifikan.”</p>	<p>VD yang menunjukkan bahwa imbal hasil indeks pasar saham mampu menjelaskan secara dominan terhadap fluktuasi yang terjadi pada imbal hasil indeks pasar saham sebagai akibat terjadinya kejutan, begitu pula pada fluktuasi akibat kejutan pada sentimen investor mampu dijelaskan cukup dominan oleh sentimen investor itu sendiri”</p>
<p>Nama Jurnal: Seminar Nasional Teknologi Informasi dan Multimedia 2018</p> <p>Judul Jurnal: ANALISIS SENTIMEN BERDASAR KAN OPINI MASYARAKAT PADA</p>	<p>Apakah opini pada media sosial terkhusus twitter dalam pelayanan yang disediakan oleh penyedia jasa ekspedisi barang memiliki respond yang baik ?</p>	<p>Metode pengumpulan data: Data yang digunakan berjumlah 200 buku, 150 data digunakan sebagai data latih, sedangkan 50 data digunakan sebagai data uji.</p>	<p>“Berdasarkan dari hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini telah menghasilkan sebuah sistem analisis sentimen jasa pengiriman barang menggunakan <i>Naive Bayes</i>. Sistem yang dibangun telah mampu</p>

<p>TWITTER MENGGUNAKAN AKAN <i>NAÏVE</i> <i>BAYES</i></p> <p>Nama Penulis: Teguh Ansyor Lorosae, Burhanudin Dwi Prakoso, Saifudin, Kusrini</p>		<p>Hasil: “Data yang digunakan sebanyak 6000 data yang terdiri dari 2000 kalimat netral, 2000 kalimat positif dan 2000 kalimat negatif dengan menghasilkan akurasi antara 70,43% - 88,03%.”</p>	<p>mentransformasi sentimen yang berupa teks opini pada fanpage Tiki sehingga dapat menampilkan informasi sentimen masyarakat terhadap jasa pengiriman barang yang bersifat positif, negatif dan netral. Penggunaan metode <i>Naive Bayes</i> classifier pada penelitian ini tidak menjamin ketepatan dalam proses klasifikasi. Akurasi pada sistem yang dibangun mencapai 84%. Tingkat akurasi akan semakin baik bila data latih yang digunakan lebih banyak, jumlah kategori pada setiap kelas sentimen pun akan berpengaruh.”</p>
<p>Nama Jurnal: Jurnal Warta Edisi : Jurnal TECHNO Nusa Mandiri Vol. 16, No. 1 Maret 2019</p>	<p>Apakah dengan menggunakan <i>Naive Bayes</i> dengan PSO lebih baik daripada</p>	<p>Metode pengumpulan data: penelitian kepustakaan (library research).</p>	<p>“Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan</p>

<p>Judul Jurnal: ANALISIS SENTIMEN OPINI PUBLIK BAHASA INDONESIA TERHADAP WISATA TMII MENGGUNAKAN NAÏVE BAYES DAN PSO</p> <p>Nama Penulis: Ratih Yulias, Retno Sari</p>	<p><i>Naïve Bayes</i> sendiri ?</p>	<p>Hasil: “data sebanyak 50 ulasan positif dan 50 ulasan negatif dimana hasil akurasi yang didapat menggunakan metode <i>Naïve Bayes</i> sebesar 70% sedangkan akurasi menggunakan <i>Naïve Bayes</i> dan PSO sebesar 94,02%”</p>	<p><i>Naïve Bayes</i> Dan PSO, membandingkan akurasi yang didapat dari penggunaan 2 metode yaitu <i>Naïve Bayes</i> dan <i>Naïve Bayes</i> dengan PSO. Setelah dilakukan eksperimen untuk analisis sentimen opini publik Bahasa Indonesia Terhadap Wisata TMII dengan menggunakan <i>Naïve Bayes</i> didapati akurasi sebesar 70% dengan menggunakan 4 fold Cross Validation. Sedangkan hasil eksperimen untuk analisis sentimen opini publik Bahasa Indonesia Terhadap Wisata TMII dengan menggunakan <i>Naïve Bayes</i> dengan PSO didapati akurasinya sebesar 94.02% dengan</p>
---	-------------------------------------	---	--

			menggunakan 9 Fold Cross Validation.”
Judul Jurnal: Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan <i>Random Forest Classifier</i>	Apakah dengan penambahan pendeteksi sarkasme didalam <i>Random Forest</i> dapat meningkatkan akurasi akhir ?	Metode Pengumpulan data : dengan melakukan <i>tweet crawling</i> yang dimana data didapatkan dari twitter Hasil : “Dengan menggunakan pendeteksian sarkasme pada analisis sentimen dapat meningkatkan nilai rata-rata dari akurasi, presisi, recall dan <i>f1score</i> walaupun ada satu nilai yang terjadi penurunan tetapi tidak mempengaruh i nilai rata- rata. Kenaikan nilai rata-rata akurasi sebesar 16,61 %, kenaikan nilai presisi sebesar	Dengan melakukan penambahan Pendeteksi sarkasme, dapat meningkatkan tingkat akurasi akhir dari <i>Random Forest</i> dimana sebelumnya sebesar 62,95% menjadi 77,79%
Nama Penulis : Debby Alita, Auliya Rahman			

		5,45 %, kenaikan nilai recall sebesar 9,64% dan kenaikan nilai <i>F1score</i> sebesar 11,27%.”	
--	--	--	--

Dari Penelitian terdahulu, berdasarkan metode analisis sentimen, maka dipilihnya *Support Vector Machine*, *Naïve Bayes* dikarena pada penelitian sebelumnya, akurasi dari kedua model tersebut sangatlah tinggi sehingga dipilihnya menjadi salah satu dari 3 model yang akan digunakan dan *Random Forest* karena pada penelitian terdahulu didapatkan tingkat akurasi yang sebelumnya dibawah 70% dapat meningkat jika ditambahkan metode lain sehingga *Random Forest* dipilih untuk digabungkan dengan *Support Vector Machine*, *Naïve Bayes* untuk mengetahui apakah hal yang serupa akan serupa akan terjadi namun dengan menggabungkan dua model algoritma yang berbeda namun memiliki akurasi yang lebih tinggi dari *Random Forest*. Berdasarkan metode prediksi, dipilihnya untuk membandingkan hasil dari penggabungan 3 model algoritma yaitu *Support Vector Machine*, *Naïve Bayes*, dan *Random Forest* yang akan membuat suatu prediksi baru yang akan dibandingkan hasil akurasi tersebut dengan akurasi algoritma masing-masing.