

BAB II

LANDASAN TEORI

2.1. Media Sosial

Media sosial merupakan jaringan *online* yang memungkinkan penggunanya untuk berinteraksi, berbagi, dan menciptakan isi jejaring sosial secara mudah [14]. Melalui media sosial manusia dapat menjalin sebuah interaksi seperti komunikasi satu sama lain sebagai seorang pengguna. Dengan adanya media sosial, memberikan kemudahan bagi orang – orang untuk ikut terhubung dengan yang lainnya.

Media sosial dapat menjadi sebuah wadah bagi penggunanya untuk mengemukakan pendapatnya [10]. Dengan menggunakan media sosial seseorang dapat menyampaikan isi pesan atau pendapat mereka dengan banyak orang. Akibatnya media sosial dapat digunakan untuk wadah mengekspresikan diri.

Berdasarkan kedua pemahaman diatas, media sosial dapat menjadi pintu bagi para penggunanya untuk menjalin hubungan secara mudah dan tanpa jarak batas maupun waktu. Hal tersebut juga dapat menjadi kesempatan bagi orang -orang untuk mengeluarkan suaranya secara publik untuk mencapai banyak orang. Sehingga media sosial dapat dikatakan sebagai wadah orang berkomunikasi tanpa batas.

2.2. Twitter

Twitter merupakan salah satu media sosial yang menggunakan internet untuk menghubungkan pengguna - penggunanya secara global. Sebagai media sosial, *Twitter* tergolong dalam kategori *microblog* karena berpusat pada aktivitas *posting* pesan singkat berupa *tweets* [15]. Dimana, sebagai *microblog*, *Twitter* membatasi penggunanya untuk membuat *posting* sebesar 140 kata.

Twitter merupakan media sosial yang memberikan layanan komunikasi dalam jaringan sehingga memungkinkan penggunanya untuk menuliskan pesan singkatnya untuk dipublikasikan secara meluas [16]. Seperti media sosial pada umumnya, *Twitter* juga menggunakan sistem berbasis *real-time* dalam memberikan penggunanya layanan komunikasi.

Seiring dengan perkembangan teknologi, jumlah pengguna *Twitter* juga terus meningkat setiap saat. Negara Indonesia berhasil menduduki peringkat ketiga dari Asia dalam jumlah pengguna *Twitter* terbanyak [9]. Faktor dari penyebab peningkatan pengguna tersebut adalah masyarakat Indonesia yang aktif dalam membuat *posting tweet* [17].

2.3. Analisis Sentimen

Analisis sentimen adalah sebuah proses yang bertugas untuk mendeteksi, mengekstraksi, dan mengklasifikasikan pendapat atau opini dari seseorang dalam bentuk tekstual [18]. Bentuk teks opini tersebut dapat dikategorikan ke dalam positif, negatif, atau netral. Analisis sentimen ini digunakan untuk menghasilkan pengelompokkan teks ke dalam kategori sifatnya.

Penggalian pendapat sering juga disebut sebagai analisis sentimen merupakan sebuah teknik yang melibatkan perhitungan computer untuk mengelompokkan teks berdasarkan sifat dari opininya [16]. Tujuan dari analisa tersebut adalah untuk menentukan suatu nilai opini terhadap topik tertentu. Sehingga hasil analisis sentimen dapat membentuk sebuah identifikasi penilaian terhadap sentimen.

Analisis sentimen dibutuhkan atas dasar untuk mencari tahu sikap seseorang terhadap suatu topik atau polaritas kontekstual keseluruhan dokumen [11]. Dengan menganalisis sentimen, kita dapat memahami respon pendapat masyarakat terhadap topik tertentu. Sehingga mempermudah manusia dalam memahami hasil akhir sentimen dari keseluruhan konteks topik.

2.4. *Text Mining*

Penggalian teks atau *text mining* merupakan salah satu teknik dari penggalian data yang berupa tekstual [11]. Tujuan dari penggalian teks ini adalah untuk menafsirkan makna – makna yang terkandung dalam suatu konteks tektual. Sehingga semua orang dengan mudah mampu memahami konteks yang disampaikan oleh orang lain.

Text mining merupakan sebuah proses untuk melakukan penarikan informasi dari sumber data menggunakan identifikasi dan eksplorasi [19]. Sumber data yang dimaksud merupakan sekumpulan teks yang tidak terstruktur. Penggunaan *text mining* dalam penarikan informasi tersebut dapat menghasilkan sebuah pola menarik jika informasi yang digunakan memiliki nilai.

Dalam penyelesaian masalah, *text mining* pada umumnya sering berhubungan dengan beberapa subyek lainnya seperti, analisis sentimen. Dimana, penggalian data digunakan sebagai tahap awal dalam melakukan analisis sentimen. Artinya, penambangan teks merupakan suatu proses yang menghasilkan penemuan informasi, relasi dan fakta yang tersembunyi di dalam teks [20].

2.5. Text Preprocessing

Data yang terdapat pada media sosial banyak yang belum terstruktur sehingga membutuhkan sebuah proses untuk merapikannya [19]. *Text Preprocessing* merupakan proses yang melibatkan hasil data yang telah dikumpulkan untuk dipersiapkan dalam bentuk teks yang siap diolah dalam analisis sentimen. Tujuan dari persiapan data sebelum di proses ini adalah untuk mengubah data yang berantakan menjadi data yang tata. Sehingga, proses preparasi tersebut merupakan faktor yang penting dalam mempengaruhi tingkat akurasi dan proses pelaksanaan analisa sentimen. Preparasi yang dilakukan terdiri atas tahap - tahap berikut:

a) *Data Cleansing*

Data Cleansing atau pembersihan data merupakan tahap untuk membersihkan hal - hal yang mengandung karakter non-alfabetis pada data teks yang akan diolah. Tujuan dari pembersihan data ini adalah untuk mengurangi *noise* pada saat dilakukannya proses pelatihan model SVM [12]. Contoh dari *data cleansing* dapat dilihat pada Tabel 2.1 berikut:

Tabel 2.1. Contoh Proses Data Cleansing

Sebelum <i>Data Cleansing</i>	Sesudah <i>Data Cleansing</i>
Aku sedang berada di @waterpark	Aku sedang berada di waterpark

b) *Case Folding*

Case Folding merupakan tahap untuk mengubah seluruh huruf pada data teks menjadi huruf kecil. Tujuan dari *case folding* ini adalah untuk membuat konsistensi huruf dalam mempermudah pencarian sehingga dibutuhkan sebuah standar berupa huruf kecil [21]. Contoh dari *data cleansing* dapat dilihat pada Tabel 2.2 berikut:

Tabel 2.2. Contoh Proses *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
HaRI InI	hari ini

c) Tokenisasi

Tokenisasi merupakan tahap untuk memisahkan setiap kata - kata dalam kalimat teks menjadi terpisah. Tujuan dari melakukan tokenisasi ini adalah untuk menganalisis teks lebih lanjut [21]. Contoh dari tokenisasi dapat dilihat pada Tabel 2.3 berikut:

Tabel 2.3. Contoh Proses Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
hari ini	“hari”, “ ini”

d) Pembobotan TD-IDF

Term Frequency - Inverse Document Frequency merupakan cara atau strategi yang dilakukan untuk menentukan bobot pada suatu kata sehingga dapat meningkatkan kinerja analisis sentimen dan proses teks mining lainnya pada kategorisasi teks [22]. Pembobotan ini berfungsi untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, akan dilakukan pembobotan atau perhitungan [20]. Semakin banyak *term* yang muncul pada sebuah koleksi dokumen, semakin tinggi nilai atau bobot *term* tersebut [23]. Perhitungan nilai bobot menggunakan rumus TF-IDF adalah sebagai berikut [22]:

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t$$

Rumus 2.1. Persamaan TD-IDF

$$IDF_t = \log \frac{N}{DF_t}$$

Rumus 2.2. Inverse Document Frequency

Keterangan:

t = kata - kata yang dihitung

d = bobot kalimat

$TF - IDF_{t,d}$ = kalimat bobot terhadap kata

$TF_{t,d}$ = *Term Frequeuncy*

IDF_t = *Inverse Document Frequency*

N = *jumlah kalimat*

DF_t = *jumlah kata yang terulang*

2.6. *K-fold Cross Validation*

K-fold Cross Validation merupakan salah satu metode untuk melakukan validasi dalam menghitung kinerja sesungguhnya yang dihasilkan dari model yang telah dibuat [22]. Tujuan dari validasi *K-fold* ini adalah untuk meningkatkan tingkat akurasi dengan menghilangkan *bias* atau *noise* pada kata. *K-fold* bekerja dengan menggunakan iterasi yang telah ditentukan untuk mengganti *data testing* dan *data training* sehingga validasi yang dihasilkan memiliki akurasi yang bagus.

Dalam melakukan validasi atau pengujian model yang telah dibuat, *K-fold Cross Validation* memiliki pengaruh terhadap nilai akurasi [12]. Dimana semakin banyak nilai iterasi yang dilakukan akan mengurangi *noise* pada analisis sehingga menghasilkan nilai akurasi yang tinggi. Pengurangan *noise* (data yang tidak

berguna) merupakan salah satu faktor penting dalam melakukan validasi keakuratan model.

Pemahaman lain mengenai *K-fold Cross Validation* adalah sebuah proses penilaian yang dilakukan dengan membentuk pembagian antara data *testing* dan data *training* [21]. Dengan menggunakan *K-fold Cross Validation*, hasil pembagian data memiliki performa yang baik. Hal tersebut dikarenakan *K-fold Cross Validation* dapat menentukan jumlah iterasi terbaik sesuai dengan akurasi yang dihasilkan pada setiap pembagian [22]. Contoh *K-fold* dapat dilihat pada Tabel 2.4 berikut:

Tabel 2.4. Contoh Proses *K-fold Cross Validation*

Iterasi ke - 1	Iterasi ke - 2	Iterasi ke - 3
<i>data testing: 1</i>	<i>data testing: 2</i>	<i>data testing: 3</i>
<i>data training: 2,3</i>	<i>data training: 1,3</i>	<i>data training: 1,2</i>

2.7. *Rapidminer Studio*

Rapidminer studio merupakan salah produk perangkat lunak yang dikembangkan oleh perusahaan "*Rapidminer Inc*". Dalam kegunaannya *rapidminer* dibuat untuk data ilmu pengetahuan yang dapat menyediakan lingkungan terpadu dari melakukan penambahan teks hingga menerapkan pembelajaran mesin [24]. *Rapidminer* juga dilengkapi dengan alat – alat yang dapat digunakan untuk penelitian, pembelajaran, hingga pengembangan sebuah aplikasi.

Rapidminer merupakan sebuah perangkat lunak yang mempermudah penggunaannya dalam melakukan pengolahan data [25]. *Rapidminer* mempermudah penggunaannya dalam mengelola data karena *rapidminer* sudah menerapkan penggunaan prinsip dan algoritma yang diperlukan dalam penggalian data. Menjadikannya salah satu perangkat lunak pilihan terbaik untuk melakukan ekstraksi data.

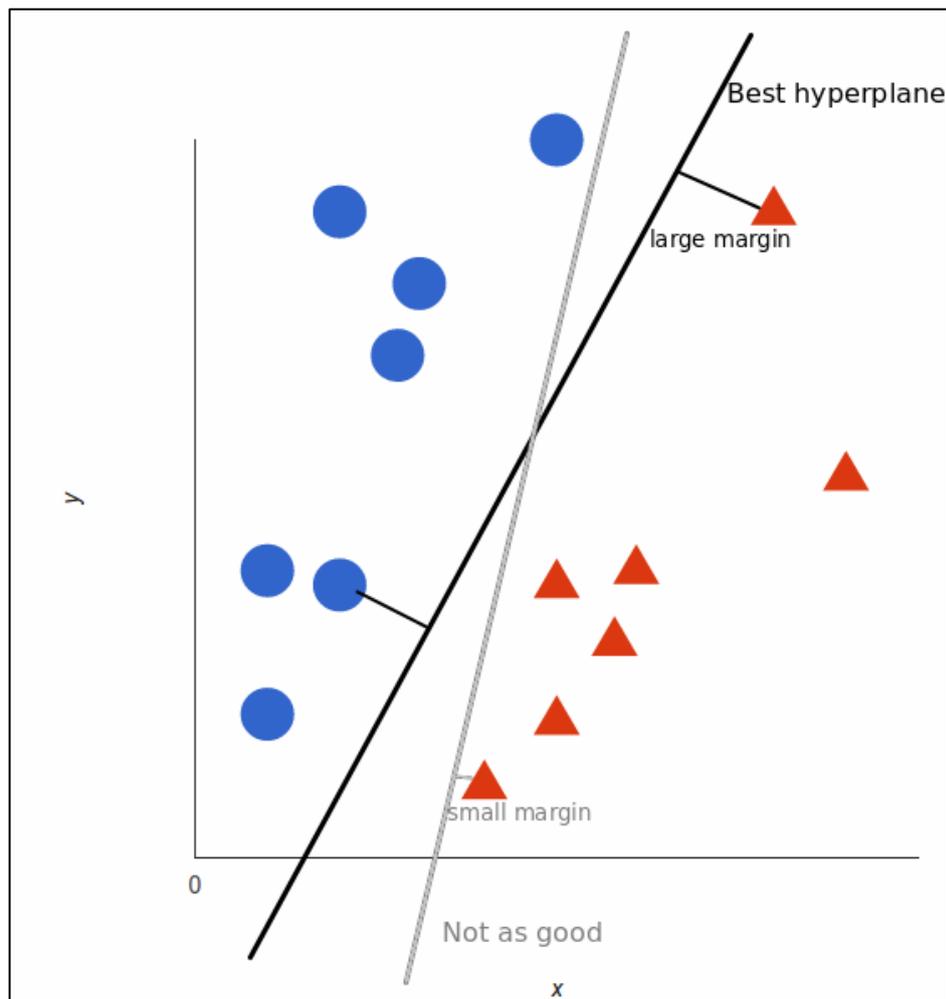
Kemampuan *rapidminer* dalam mengelola data juga memiliki akurasi yang tepat dengan perhitungan manual menjadikannya perangkat lunak yang baik untuk digunakan [26]. Hal tersebut juga didukung oleh *tools* yang disediakan *rapidminer* sangatlah luas. Meliputi dari pengambilan data, persiapan data, penerapan algoritma, pengecekan hasil, dan visualisasi hasil [27].

2.8. Support Vector Machine

Support Vector Machine (SVM) adalah salah satu algoritma yang pertama kali diperkenalkan oleh Vapnik pada tahun 1992 [28]. Algoritma ini digunakan untuk melakukan pembelajaran mesin. Dimana mesin akan melakukan komputerasi dalam membuat sebuah garis pembatas (*hyperplane*) terbaik untuk memisahkan dua buah kelas. Seiring dengan perkembangannya, SVM kini dapat memperluas perhitungannya untuk melakukan klasifikasi kelas yang banyak.

SVM bekerja dengan mengubah data pelatihan menjadi dimensi yang lebih tinggi (*vectors*) melalui pemetaan nonlinier [29]. Kemudian menggunakan dimensi baru yang telah dibuat, akan dicari berbagai macam fungsi pemisah (*classifier*) yang dapat digunakan untuk melakukan pemisahan dua kelas berbeda. Selanjutnya

SVM melakukan perbandingan fungsi pemisah yang telah ditemukan untuk menentukan fungsi pemisah terbaik (*hyperplane*). Contoh visualisasi dari cara kerja SVM dapat dilihat pada Gambar 2.1:



Gambar 2.1. SVM Menentukan *Hyperplane* Terbaik

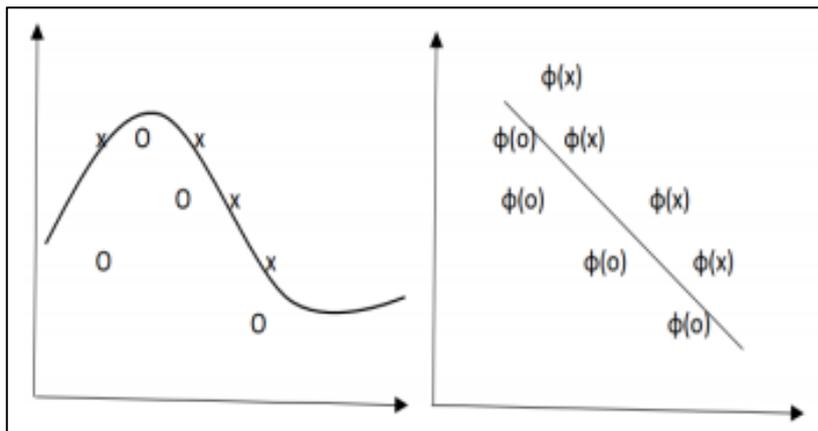
Sumber: [30]

Keistimewaan yang dimiliki SVM adalah kemampuannya dalam menetapkan *hyperplane* (pemisah) *input* data nonlinier berdimensi tinggi [15]. Hal tersebut dikarenakan SVM dapat menggunakan fungsi kernel pada kasus yang tidak linier

menjadikannya suatu dimensi linier. Dalam SVM terdapat berbagai fungsi kernel yang umumnya dipakai [29], yaitu:

- Linear = $X^T X$
- Polinomial = $(X^T X_i + 1)^p$
- Radial Basis Function (RBF) = $\exp(-1/2a^2 ||X - X_i||^2)$
- Sigmond = $\tanh(\beta^T X_i + \beta_i), \beta, \beta_i \in \Re$

Contoh hasil dari penerapan fungsi kernel untuk mengubah kasus nonlinier menjadi linier dapat dilihat pada Gambar 2.2 dibawah:



Gambar 2.2. Penerapan Fungsi Kernel Pada Kasus Nonlinier

Sumber: [29]

2.9. Naïve Bayes

Naïve Bayes adalah salah satu metode pengklasifikasian yang melibatkan sekumpulan probabilitas [31]. Algoritma ini bekerja dengan melakukan perhitungan probabilitas dengan menjumlahkan kombinasi dan frekuensi nilai dari

dataset yang dimiliki. Kemudian dilanjutkan dengan teorema *bayes*, yaitu mengasumsikan ketergantungan variabel satu dengan variabel lainnya sehingga dapat dibentuk klasifikasi.

Cara kerja yang sederhana, cepat, dan mampu menghasilkan akurasi yang tinggi menjadikannya keunggulan bagi algoritma *naïve bayes* [32]. Hal tersebut membuat *naïve bayes* menjadi salah satu algoritma klasifikasi yang terumum yang digunakan. Namun, pada penerapannya juga terdapat keterbatasan utama yang dimana algoritma tidak selalu dapat memenuhi asumsi ketergantungan antara atribut yang ingin dilakukan klasifikasi.

Naïve bayes adalah metode klasifikasi yang tidak terikat oleh aturan - aturan [33]. Dimana algoritma komputer tersebut bekerja menggunakan teori probabilitas dalam menemukan peluang terbesar untuk kemungkinan pembuatan klasifikasi. Hal - hal tersebut membuat *naïve bayes* dikenal sebagai *Idiot's Bayes* dan *Independence Bayes*.

2.10. Logistic Regression

Regresi Logistik (*Logistic Regression*) merupakan sebuah metode untuk menghasilkan relasi variabel respon dengan variabel bebas [34]. Dalam praktiknya regresi logistik dapat digunakan untuk dua jenis klasifikasi. Jenis pertama adalah regresi logistik biner dimana pembuatan klasifikasi terdiri atas dua (2) kelas. Kemudian jenis kedua adalah multinomial yang memungkinkan algoritma melakukan klasifikasi untuk jumlah kelas yang lebih dari dua (2).

Regresi logistik tidak sama dengan regresi linier pada umumnya dimana logistic bekerja tanpa melakukan asumsi hubungan antar variabel respon dan terikat secara linier. Regresi logistik berasal pada analisis regresi dan digunakan ketika variabel respon merupakan variabel dikotomi [13]. Dalam pengerjaannya regresi logistik melakukan asumsi - asumsi yang harus dipenuhi, yaitu: variabel terikat bersifat dikotomi, variabel bebas tidak perlu memiliki persamaan tetapi harus terpisah, dan tidak memerlukan asumsi kenormalan ganda [35].

Klasifikasi regresi logistik merupakan salah satu metode tradisional [36]. Dimana pada contoh realitanya adalah regresi logistik sering dilakukan pada kasus klasifikasi calon nasabah baru kartu kredit, klasifikasi wilayah bersarkan perekonomiannya. Perhitungan klasifikasi regresi logistik dilakukan dengan rumus - rumus sebagai berikut [37]:

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

Rumus 2.3. Regresi Logistik

Kemudian diperoleh persamaan peluangnya menjadi:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Rumus 2.4. Persamaan Peluang Regresi Logistik

2.11. Confusion Matrix

Confusion Matrix adalah alat untuk memvisualisasi kualitas klasifikasi yang dihasilkan menggunakan penggunaan matrix [38]. Dalam visualisasinya, *confusion*

matrix terdiri dari empat (4) faktor yang menentukan performa klasifikasi. Berikut adalah rincian empat (4) faktor berdasarkan [15], yaitu:

1. TP = *True Positive* = Prediksi Positif yang benar
2. FP = *False Positive* = Prediksi Positif yang salah (eror klasifikasi)
3. TN = *True Negative* = Prediksi Negatif yang benar
4. FN = *False Negative* = Prediksi Negatif yang salah (eror klasifikasi)

Rumus untuk perhitungan *performance metric* yang umum digunakan dalam *confusion matrix* terdiri dari [39]:

1. Akurasi = Keakuratan klasifikasai

$$\frac{TP + TN}{TP + FN + FP + TN}$$

Rumus 2.5. Akurasi *Performance Metric*

2. *Precision* = Rasio prediksi benar positif dengan keseluruhan yang diprediksi positif

$$\frac{TP}{TP + FP}$$

Rumus 2.6. *Precision Performance Metric*

3. *Recall* = Rasio prediksi benar positif dengan keseluruhan data yang benar positif

$$\frac{TP}{TP + FN}$$

Rumus 2.7. Recall Performance Metric

Tabel 2.5 adalah contoh visualisasi dari *confusion matrix*:

Tabel 2.5. Contoh Confusion Matrix

	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Confusion Matrix menjadi penting karena dapat digunakan untuk mengevaluasi hasil klasifikasi. Dimana akurasi yang dimiliki pada suatu hasil klasifikasi berpengaruh terhadap performa dari sebuah klasifikasi [40]. Sehingga *confusion matrix* dapat dipakai untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda.

Confusion matrix adalah sebuah teknik yang digunakan untuk memperkirakan objek yang benar atau salah [20]. Pada praktiknya, *confusion matrix* bekerja dengan membandingkan hasil prediksi klasifikasi dengan informasi kelas yang asli. Kemudian menampilkan perbandingan hasil tersebut ke dalam bentuk matrix.

2.12. ROC Curve (AUC)

Kurva ROC (*ROC Curve*) merupakan teknik yang banyak sekali digunakan dalam penilaian hasil prediksi dan digunakan untuk menilai hasil kinerja klasifikasi sebuah algoritma [33]. Dimana kurva ROC berfungsi sebagai gambaran visualisasi menggunakan rata - rata *True Positif* dengan *False Negatif*. Kegunaan dari visualisasi tersebut adalah melihat pemetaan klasifikasi yang dibentuk.

Kurva *Receiver Operating Characteristic* atau ROC dalam praktiknya digunakan sebagai visualisasi yang mengekspresikan data *confusion matrix* [41]. Sehingga menghasilkan sebuah kurva garis biru dan garis merah AUC atau *Area Under Curve*. Kegunaan dari AUC tersebut adalah membentuk sebuah perbatasan dibawah kurva garis biru yang menandakan performa algoritma dalam membandingkan objek ke dalam klasifikasi.

Dalam menentukan performa klasifikasi, kurva ROC juga sering digunakan dalam perhitungan numerik dan untuk mengukur akurasi klasifikasi menggunakan AUC. Hasil yang dikeluarkan AUC adalah berupa angka 0 hingga 1 untuk klasifikasi terbaiknya. Tabel 2.6 adalah tingkat akurasi AUC yang dirumuskan ke dalam lima (5) kategori [38].

Tabel 2.6. Katogori Klasifikasi AUC

Nilai AUC	Kategori Klasifikasi
0.90 - 1.00	Excellent Classification
0.80 - 0.90	Good Classification
0.70 - 0.80	Fair Classification
0.60 - 0.70	Poor Classification
0.50 - 0.60	Fail Classification

2.13. Tinjauan Pustaka

Tabel 2.7. Penelitian - penelitian Terdahulu

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
1.	Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma <i>Support Vector Machine</i> - Jurnal RESTI Vol. 3 No. 2 (2019) 243-250	Heru Sukma Utama, Didi Rosiyadi, Bobby Suryo Prakoso, dan Dedi Ariadarma (2019)	<i>Support Vector Machine</i> dan <i>K-Nearest Neighbor</i>	Hasil pengelolaan data menggunakan algoritma SVM memiliki hasil tingkat <i>confusion matrix</i> yang bagus berupa akurasi, presisi, dan <i>recall</i> yang baik. [11]	Penerapan algoritma <i>Support Vector Machine</i> , metode pengambilan data dari <i>Twitter</i> .
2.	Pengaruh <i>Semantic Expansion</i> pada <i>Naïve Bayes</i>	Muhamad Satria Adhi, Muhammad Zidny	<i>Naïve Bayes</i>	Penambahan metode <i>semantic expansion</i> dalam proses	Metode pengambilan data dari <i>Twitter</i> dan pemahaman

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
.	<i>Classifier</i> untuk Analisis Sentimen Tokoh Masyarakat - Jurnal RESTI Vol. 3 No. 2 (2019) 141 - 147	Naf'an, dan Elisa Usada (2019)		algoritma <i>Naïve Bayes</i> mampu meningkatkan tingkat akurasi. Dimana tingkat nilai akurasi memiliki penambahan sebesar 2% untuk melakukan klasifikasi teks. [42]	cara kerja algo <i>Naïve Bayes</i> .
3.	Analisis Sentimen Pemindahan Ibu Kota Negara dengan <i>Feature Selection</i> Algoritma <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> - Jurnal RESTI Vol. 4 No. 3 (2020) 504 - 512	Faried Zamachsari, Gabriel Vangeran Saragih, Susafa'ati, dan Windu Gata (2020)	<i>Support Vector Machine</i> dan <i>Naïve Bayes</i>	Hasil tingkat akurasi analisis sentimen menggunakan algoritma <i>Support Vector Machine</i> memiliki akurasi yang lebih tinggi tanpa menerapkan <i>feature selection</i> . Akan tetapi, berbanding terbalik pada kasus analisis sentimen menggunakan algoritma <i>Naïve Bayes</i> . Dimana tingkat akurasi dari <i>Naïve Bayes</i> mengalami peningkatan ketika	Metode pengambilan data dari <i>Twitter</i> dan pemahaman hasil perbandingan algoritma <i>Support Vector Machine</i> dengan <i>Naïve Bayes</i> .

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
				menggunakan <i>feature selection</i> . [31]	
4.	Komparasi Algoritma <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> Untuk Analisa Sentimen <i>Review Film</i> - Jurnal PILAR Nusa Mandiri Vol. 14 No. 2	Elly Indrayuni (2018)	<i>Support Vector Machine</i> dan <i>Naïve Bayes</i>	<i>Naïve Bayes</i> memiliki kelebihan yaitu sederhana, cepat dan memiliki akurasi yang tinggi. Sementara SVM memiliki nilai akurasi yang lebih tinggi dan akurat dibandingkan <i>Naïve Bayes</i> [32].	Pemahaman dan perbandingan algoritma <i>Support Vector Machine</i> dengan <i>Naïve Bayes</i> .
5.	Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode <i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i> – Jurnal Majalah Ilmiah Teknologi Elektro, Vol. 18, No. 1, Januari - April 2019	Dwi Ardiada, Made Sudarma, Dwi Giriantari (2019)	<i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i>	<i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i> dapat digunakan untuk melakukan klasifikasi dan memiliki akurasi yang bagus. [19]	Pemahaman algoritma <i>Support Vector Machine</i> .

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
6.	<i>Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine - International Journal of Computer Science and Software Engineering (IJCSSE) Vol. 6, Issue 8</i>	Fiktor Imanuel Tanesab, Irwan Sembiring, dan Hindriyanto Dwi Purnomo (2017)	<i>Support Vector Machine</i>	Dalam penelitian ini ditemukan bahwa hasil klasifikasi nilai dari pembobotan menggunakan <i>Lexicon Based</i> memiliki nilai True Positive rate sebesar 91,1% berdasarkan komentar yang diambil dari tahun 2015 hingga 2016. [43]	Penerapan algoritma <i>Support Vector Machine</i> dan pemahaman pengaruh pembobotan.
7.	<i>Implementing Support Vector Machine Sentiment Analysis to Students' Opinion toward Lecturer in an Indonesian Public University - Journal of Sustainable Engineering: Proceedings Series 1(2) 2019</i>	Daniel Febrian Sengkey, Agustinus Jacobus, dan Fabian Johanes Manoppo (2019)	<i>Support Vector Machine</i>	Hasil penggunaan SVM untuk mengklasifikasi opini pelajar terhadap pengajar di universitas umum Indonesia memiliki tingkat akurasi sebesar 74% dengan nilai kappa 0,46. Dimana nilai tersebut termasuk dalam kategori cukup moderat. Namun, hal tersebut tidak	Penerapan algoritma <i>Support Vector Machine</i> .

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
				menutup kemungkinan untuk peningkatan akurasi dengan menambah tahapan - tahapan pada <i>preprocessing</i> . [44]	
8.	<i>Sentiment Analysis of Public Opinion on The Go-Jek Indonesia Through Twitter Using Algorithm Support Vector Machine - IOP Conf. Series: Journal of Physics: Conf. Series 1462 (2020) 012063</i>	H Syahputra, L K Basyar, dan A A S Tamba (2020)	<i>Support Vector Machine</i>	Menggunakan algoritma SVM ditemukan hasil akurasi sebesar 91,8% dengan 1977 fitur data pelatihan. Selain itu, kesalahan klasifikasi dalam dokumen dapat disebabkan oleh suatu kelas yang berisi kata - kata yang sama dengan kelas lainnya sehingga pembobotan kata pada kategori lain menjadi lebih besar dari kelas yang seharusnya. [45]	Penerapan algoritma <i>Support Vector Machine</i> .
9.	<i>A Naive Bayes</i>	Rein Rachman	<i>Naive Bayes</i>	Hasil data cleansing	Pemahaman dan

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
	<i>Sentiment Analysis for Fintech Mobile Application User Review in Indonesia - International Journal of Advanced Trends in Computer Science and Engineering Vol. 8 No. 5</i>	Putra, Monika Evelin Johan, dan Emil Robert Kaburuan (2019)		mampu mempengaruhi hasil perbedaan akurasi yang sedikit lebih rendah menggunakan algoritma <i>Naïve Bayes</i> . [46]	penerapan algoritma <i>Naïve Bayes</i> .
10	Komparasi Algoritma Klasifikasi Menentukan Kelulusan Mata Kuliah Pada Universitas Budi Luhur - Jurnal String Vol.3 No.1 Agustus 2018	Frastian, Nahot Hendrian, Senna Valentino, V.H. (2018)	<i>Decision Tree, Naïve Bayes, Random Forest</i>	<i>Decision Tree, Naïve Bayes, Random Forest</i> memiliki tingkat akurasi yang tinggi dimana akurasi tertinggi diraih oleh <i>decision tree</i> . [33]	Pemahaman <i>naïve bayes</i> .
11	Klasifikasi Pemilihan Program Studi di Fakultas MIPA Universitas Lambung Mangkurat Menggunakan Regresi Logistik Multinomial - Jurnal Matematika	Silvi Risaria Dewi, Nur Salam, Dewi Sri Susanti (2018)	<i>Logistic Regression</i>	<i>Regresi logistic</i> dapat digunakan untuk memberikan visualisasi klasifikasi pada mahasiswa untuk pemilihan fakultas [34].	Pemahaman algoritma <i>Logistic regression</i> .

No	Judul dan Nama Jurnal	Pengarang	Metode Penelitian	Kesimpulan	Adopsi dari Penelitian
	Murni dan Terapan "epsilon" Vol.12 No.2 Desember 2018				
12	Komparasi 5 Metode Algoritma <i>Data Mining</i> Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan - Jurnal Techno Nusa Mandiri Vol.XIII, No. 1 Maret 2016	Sari Dewi (2016)	<i>Neural Nertwork, Decision Tree, Naïve Bayes, K-Nearest Neighbor, Logistic Regressio n</i>	Kelima algoritma yang digunakan mampu menghasilkan <i>Good Classification</i> kecuali <i>Naïve Bayes</i> karena tergolong dalam <i>Fair Classification</i> [13].	Pemahaman, penerapan algoritma <i>Logistic Regression</i> dan <i>Naïve Bayes</i> .