

Komparasi Pengaruh Model Klasifikasi Naive Bayes dan Support Vector Machine pada Analisis Data Sentimen di Bidang Pendidikan

Riri Fajriah¹, Denni Kurniawan²

^{1,2}Program Studi Magister Ilmu Komputer

Fakultas Teknologi Informasi, Universitas Budi Luhur, Indonesia

¹2211600941@student.budiluhur.ac.id, ²denni.kurniawan@budiluhur.ac.id

Article Info

Article history:

Received Feb 9, 2024

Revised May 15, 2024

Accepted May 20, 2024

Keywords:

Analisis Sentimen

Naive Bayes

Support Vector Machine

Komparasi

Bidang Pendidikan

ABSTRACT

Some of the problems that often occur in the application of text mining processing are related to data pre-processing, feature selection, classification models, class imbalance, language dynamics, evaluation, and validation. The application of supervised machine learning techniques to classification models such as Naive Bayes and Support Vector Machines in sentiment analysis requires special attention to ensure accurate and reliable results. The purpose of this study is to evaluate how the Naive Bayes and Support Vector Machine classification models affect sentiment analysis, based on the results of several studies using these classification models in the field of education. The method in this study is a literature review with a search for research literature, inclusion and exclusion criteria, screening and selection process of research results, and assessment of research results that have been done before. The results of the literature review in this study show that the Naive Bayes and Support Vector Machine classification models can provide good accuracy values in sentiment analysis research, but combining the two models with an ensemble approach further improves accuracy. As for sentiment analysis research, there are several important factors that need to be considered, such as research contributions, data mining implementation methods, influencing parameters, data selection and extraction processes, data evaluation mechanisms, and the risk of failure. All of these factors are expected to be considered as a conceptual framework that will support success in sentiment analysis research for future research in the field of education.

Copyright © 2024 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Riri Fajriah,

Fakultas Teknologi Informasi,

Universitas Budi Luhur,

Jl. Ciledug Raya, RT.10/RW.2, Petukangan Utara, Pesanggrahan, Jakarta Selatan, 12260

Email: 2211600941@student.budiluhur.ac.id

1. PENDAHULUAN

Analisis sentimen adalah proses menganalisis teks atau data untuk menentukan sentimen atau perasaan yang terkandung di dalamnya. Analisis sentimen biasanya digunakan dalam pengolahan bahasa alami atau yang dikenal dengan *Natural Language Processing* untuk menentukan apakah sebuah teks menunjukkan perasaan positif, negatif, atau netral tentang subjek tertentu. Dalam penelitian terkait dengan analisa sentimen metode pengolahan data yang tepat dilakukan dengan menggunakan model *supervised* klasifikasi [1]. Salah satu jenis model dalam *data mining* yang digunakan untuk memprediksi kelas atau label instance berdasarkan atributnya. Dalam analisis sentimen yang telah disebutkan sebelumnya, model *supervised* klasifikasi dapat

digunakan untuk memprediksi sentimen (positif, negatif, atau netral) dari teks berdasarkan atribut yang diekstrak dari teks.

Pemanfaatan analisis sentimen umumnya dilakukan pada bidang pemasaran. Hal ini terkait dengan kebutuhan perusahaan untuk mengevaluasi sentimen masyarakat atau pelanggan terkait dengan produk dan layanan yang diberikan. Pada bidang pemasaran sumber data analisis sentimen dapat diambil dari *platform social media* seperti *twitter*. Akan tetapi pada penelitian yang berbeda metode analisis sentimen juga diimplementasikan di bidang pendidikan. Teknik pengumpulan data yang dilakukan umumnya adalah menggunakan survey dari pihak-pihak yang memiliki keterlibatan dalam proses pengajaran seperti pengajar dan siswa [2]. Tujuan penelitian adalah tolak ukur keberhasilan proses pengajaran berdasarkan hasil evaluasi data sentimen siswa agar meningkatkan serta memperbaiki kualitas dan kepuasan pengajaran di masa depan.

Dengan pemahaman bahwa analisis sentimen memberikan kontribusi pada proses evaluasi kinerja operasional suatu bisnis di berbagai bidang seperti pemasaran dan pendidikan, maka upaya penelitian terkait dengan topik analisis sentimen perlu pengkajian dari perspektif metode pada *data mining* yang dapat diimplementasikan. *Data mining* adalah proses yang mengekstraksi pola atau informasi berguna dari kumpulan data yang besar, kompleks, dan seringkali tidak terstruktur. Tujuan utama dari proses ini adalah untuk menemukan informasi atau pengetahuan yang tersembunyi atau berharga dari kumpulan data tersebut, yang kemudian dapat digunakan untuk membuat keputusan yang lebih baik, memecahkan masalah bisnis, atau memprediksi hasil di masa depan [3]. Metodologi yang dapat menjadi panduan dalam mengimplementasikan data mining adalah menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Tahapan-tahapan penting dalam penelitian *data mining* terdiri dari beberapa proses yaitu : *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation* dan *Deployment* [4].

Salah satu proses yang dilakukan dalam implementasi metode CRISP-DM adalah *Modelling*. Pada proses ini dilakukan pemilihan teknik pemodelan untuk menentukan analisa data mining yang sesuai berdasarkan tujuan bisnis dan masalah yang ingin diselesaikan [5]. Pada tujuan bisnis menganalisis data sentimen diperlukan pendekatan model klasifikasi. Hal ini dikarenakan teknik machine learning akan mengklasifikasi narasi text dari suatu opini bermuatan sentimen dengan pelabelan positif, negatif maupun netral dan tingkat akurasi akan dipengaruhi dari hasil probabilitas kemunculan labelling data dengan klasifikasi yang sama. Salah satu model klasifikasi yang dapat digunakan untuk menganalisis data sentimen adalah menggunakan model algoritma *Naive Bayes* [6]. Algoritma *Naive Bayes* adalah pengklasifikasi probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi nilai dari kumpulan data yang diberikan. Algoritma ini menggunakan teorema *Bayes* dan menganggap bahwa nilai variabel kelas diperhitungkan untuk semua variabel independen. Dalam skema penelitian *artificial neural network* didapatkan hasil bahwa hasil kinerja algoritma *Naive Bayes* bergantung pada pengukuran akurasi dari hasil klasifikasi data yang dianalisa [7]. Pada beberapa hasil penelitian implementasi *Naive Bayes* memiliki kelemahan dimana tingkat akurasi model *Naive Bayes Classifier* tidak terlalu baik apabila dibandingkan model klasifikasi lainnya, namun salah satu strategi untuk mengatasi hal ini adalah dengan menggunakan pendekatan *ensemble* yaitu dimana menggabungkan *Naive Bayes Classifier* dengan metode *Support Vector Machine* dan dalam penelitian yang sudah dilakukan terjadi peningkatan nilai akurasi dengan penggabungan model algoritma *Naive Bayes Classifier* dan *Support Vector Machine* pada evaluasi data analisis sentimen [8].

SVM (Support Vector Machine) merupakan salah satu metode *machine learning* untuk *classification*. *SVM* memanfaatkan *kernel function* yaitu : *linear kernel, polynomial kernel, radial basis function kernel* dan *sigmoid kernel* atau *hyperbolic tangent kernel*. Implementasi *SVM* diberbagai bidang yaitu : *text and hypertext classification, image classification, recognizing handwritten characters, biological science, protein classification* dan lain sebagainya. Dikarenakan pemanfaatannya tersebut maka *SVM* menjadi salah satu model algoritma yang tepat untuk menganalisis data sentimen berbasis *text*. Tujuan objektif algoritma *SVM* adalah menemukan optimal *hyperlane* atau fungsi pemisah yang memisahkan dua buah *class*, dengan cara memaksimalkan margin antar dua *class*. Margin adalah jarak antara *hyperlane* dengan input *vector* data terdekat. Input *vectors* data yang bersentuhan dengan *boundary* disebut *support vectors*. Kemampuan ini yang memberikan kontribusi bahwa *SVM* cocok digunakan untuk teknik *ensemble* untuk memperbaiki kinerja model klasifikasi lainnya seperti *Naive Bayes* karena *SVM* akan membantu memaksimalkan *margin* antar *class* pada data *analisis sentimen* serta mengklasifikasikan secara akurat [9]. Dalam sebuah penelitian menggunakan *SVM (Support Vector Machine)* dan *Naive Bayes* untuk melakukan *analisis sentimen* pada dataset ulasan produk, diketahui bahwa model *SVM* adalah model pembelajaran terawasi yang dapat mencapai hasil yang baik dalam kategorisasi teks, sementara model *Naive Bayes* adalah model klasifikasi probabilistik sederhana berdasarkan teorema *Bayes* dengan asumsi "naif" yang kuat. Menurut penelitian ini, *SVM* digunakan untuk menemukan batas terbaik yang memisahkan antara sampel pelatihan positif dan negatif, sementara *Naive Bayes* digunakan untuk mengklasifikasikan ulasan berdasarkan probabilitas kemunculan kata-kata tertentu dalam ulasan tersebut. Hasil penelitian menunjukkan bahwa *SVM* cenderung memberikan hasil yang lebih akurat daripada *Naive Bayes* dalam klasifikasi dokumen. Namun, meskipun *SVM* dapat memberikan hasil yang

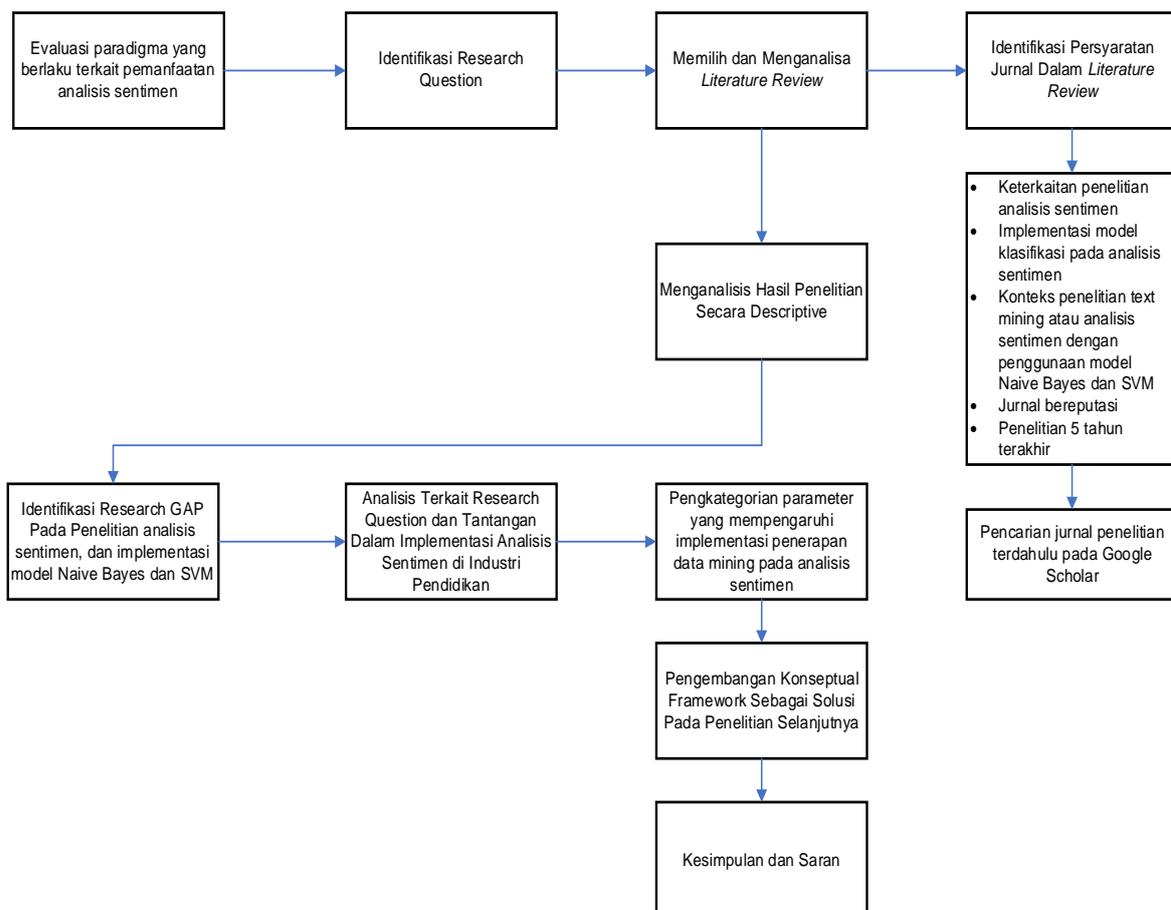
lebih baik dalam beberapa kasus, *Naive Bayes* juga dapat memberikan hasil yang baik, terutama varian *Bernoulli*. Dalam konteks analisis sentimen, kedua varian *Naive Bayes* (*Multinomial* dan *Bernoulli*) dapat mencapai hasil yang baik, meskipun SVM cenderung lebih akurat daripada *Naive Bayes* ketika dibandingkan dengan *Random Forest* [10].

Tujuan *literature review* yang dilakukan pada penelitian ini adalah didasarkan pada proses evaluasi mengenai faktor-faktor *Research GAP* pada beberapa penelitian terkait dengan analisa sentimen dan implementasi model klasifikasi *Naive Bayes* dan *Support Vector Machine*. Selain itu, berdasarkan hasil analisa dari *literature review* tersebut diharapkan dapat memberikan pemahaman mengenai faktor-faktor yang mempengaruhi analisa sentimen dengan menggunakan model klasifikasi *Naive Bayes* dan *Support Vector Machine*. Penelitian ini juga diharapkan memberikan kontribusi dengan menghasilkan suatu *Konseptual Framework* yang bertujuan untuk menyediakan panduan terstruktur dan sistematis untuk memahami hubungan antara berbagai elemen yang terlibat dalam penelitian mengenai sentimen analisis khususnya di bidang pendidikan. Hasil *Konseptual Framework* pada penelitian ini tidak hanya membantu mengorganisir dan mensintesis informasi dari *literature review* yang telah dilakukan, tetapi juga memberikan dasar yang kuat untuk evaluasi lebih lanjut dan pengembangan penelitian di masa mendatang mengenai analisis sentimen.

Selain itu, *domain goal* pada penelitian ini bertujuan memberikan wawasan bagi penelitian selanjutnya mengenai faktor-faktor penting yang perlu diperhatikan pada penelitian analisis sentimen di bidang Pendidikan seperti apa kontribusi penelitian, bagaimana proses implementasi metode data mining pada proses penelitian analisis sentimen, faktor-faktor yang mempengaruhi keberhasilan penelitian, tahap evaluasi hasil analisa data sentimen serta pertimbangan resiko kegagalan yang dapat menghambat proses penelitian analisa data sentimen di bidang pendidikan.

2. METODE

Metode penelitian yang dilakukan pada analisa *paper review* pada penelitian ini dilakukan melalui beberapa proses dengan tahapan sebagai berikut :



Gambar 1. Kerangka Penelitian

Berdasarkan informasi pada Gambar 1 tersebut diatas maka proses penelitian yang dilakukan melalui beberapa tahapan sebagai berikut :

1. Evaluasi paradigma yang berlaku saat ini terkait pemanfaatan analisis sentimen

Berdasarkan beberapa penelitian dapat dipahami manfaat dari analisis sentimen adalah terkait dengan :

Tabel 1. Pemanfaatan Analisis Sentimen

Sumber : [11]

Manfaat Analisis Sentimen	Keterangan
Memahami Opini dan Minat Pengguna	Analisis sentimen dapat membantu dalam memahami pemikiran, minat, dan opini pengguna Twitter terhadap berbagai topik atau peristiwa. Hal ini dapat memberikan wawasan berharga bagi peneliti dan aplikasi yang memerlukan pemahaman mengenai pandangan masyarakat terhadap suatu topik tertentu.
Penerapan dalam Berbagai Konteks dan Domain	Analisis sentimen Twitter dapat diterapkan dalam berbagai konteks dan domain, seperti kesehatan, kerusuhan, penjualan saham, polusi udara, dan manajemen bencana. Dengan demikian, informasi yang diperoleh dari analisis sentimen dapat digunakan untuk berbagai keperluan dan bidang.
Pemrosesan Data Besar	Analisis sentimen juga dapat memberikan solusi terbaik dalam menganalisis dan memproses kumpulan data besar. Dengan menggunakan pendekatan yang tepat, analisis sentimen dapat membantu dalam mengolah informasi yang masif yang dihasilkan oleh platform seperti Twitter.

Berdasarkan informasi pada Tabel 1 dapat dipahami bahwa pemanfaatan analisis sentimen memberikan kontribusi untuk pemahaman mengenai pandangan masyarakat untuk isu-isu yang berlangsung di masyarakat. Pertimbangan ini dirasa menjadi tolak ukur untuk memungkinkan proses implementasi analisis sentimen dibidang pendidikan yang memiliki keterkaitan dengan fungsi esensial di masyarakat.

Pada suatu penelitian Analisis sentimen pada bidang pendidikan dapat dimanfaatkan untuk berbagai tujuan, seperti evaluasi kinerja guru, analisis umpan balik dari siswa untuk mendeteksi masalah yang mungkin mereka hadapi dalam pembelajaran, pelacakan emosi siswa berdasarkan kategori emosi Plutchik sebagai sumber umpan balik informatif bagi instruktur, serta intervensi waktu nyata dalam kelas berdasarkan umpan balik siswa untuk mengatasi masalah seperti kebingungan dan kebosanan yang memengaruhi keterlibatan siswa. Analisis sentimen juga dapat digunakan untuk mewakili sikap siswa terhadap pembelajaran secara kuantitatif. Dengan menggabungkan analisis sentimen dengan faktor-faktor lain seperti kehadiran dan prestasi, sistem seperti SENSE dapat menentukan kemajuan akademis setiap siswa dari tiga faktor tersebut. Hal ini memungkinkan lembaga pendidikan untuk mengatasi tuntutan dalam jaringan sekolah yang padat dan tetap sejalan dengan penggunaan teknologi canggih seperti yang dilakukan perusahaan, karena sekolah seharusnya tidak diperlakukan secara berbeda dari sektor bisnis lainnya. Dengan demikian, analisis sentimen dapat membantu dalam memahami dan mengukur sikap siswa, memberikan umpan balik yang lebih baik kepada guru dan siswa, serta mendukung pengambilan keputusan yang lebih baik dalam konteks pendidikan [12].

2. Identifikasi *Research Question*

Beberapa hal mendasar yang menjadi *research question* pada penelitian ini berdasarkan pemahaman dari beberapa hasil *literature review* adalah :

- Apakah model klasifikasi *Naive Bayes* dan *Support Vector Machine* dapat optimal digunakan pada analisis sentimen ?
- Faktor-faktor apa saja yang mempengaruhi keberhasilan penelitian analisis sentimen ?
- Bagaimana strategi menghadapi tantangan dalam penelitian selanjutnya mengenai implementasi metode analisis sentimen pada industri pendidikan ?

Pertimbangan hasil identifikasi *research question* akan dimanfaatkan sebagai landasan pada penelitian lanjutan mengenai analisis sentimen di bidang pendidikan.

3. Memilih dan Menganalisa *Literature Review*

Dalam pemilihan dan analisa *literature review* ada beberapa parameter yang menjadi persyaratan bagi peneliti yaitu :

- Jurnal-jurnal yang terkait dalam *literature review* terkait dengan topik analisis sentimen sebagai proses pengembangan teknologi data mining khususnya pada evaluasi teknik *Natural Language Processing* (NLP) pada penelitian *text mining*.
- Spesifikasi model algoritma penelitian yang dievaluasi dari beberapa jurnal adalah terkait dengan implementasi model klasifikasi *Naive Bayes* dan *Support Vector Machine*.
- Jurnal yang digunakan dalam analisa *literature review* bersumber dari jurnal bereputasi seperti science direct, IEEE, Sinta, Elsevier dan lain sebagainya.

d. Pencarian jurnal menggunakan mesin pencarian di *Google Scholar* yang memiliki korelasi dengan tujuan penelitian.

Tahapan selanjutnya dalam proses analisa *literature review* dijelaskan pada hasil dan pembahasan sampai dengan kesimpulan dan saran dari penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1. Research GAP

Pembahasan hasil analisa *Research GAP* dari beberapa tinjauan *literature review* pada penelitian ini dijelaskan oleh Tabel 2 berikut:

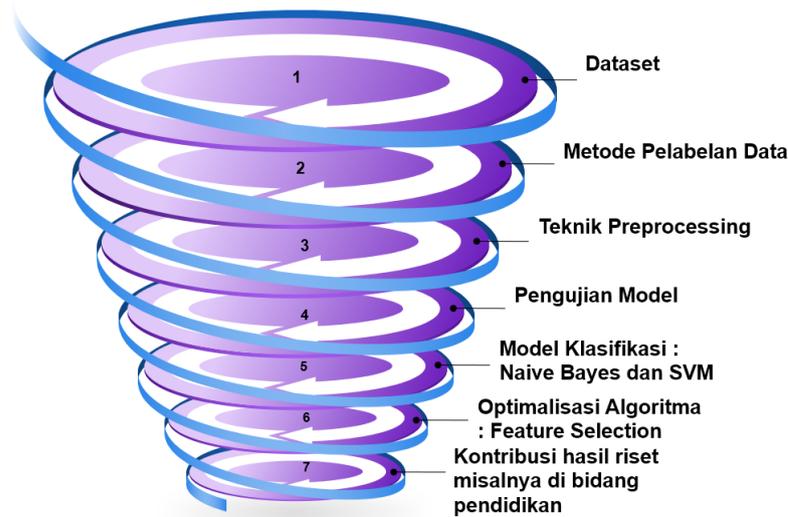
Tabel 2. Analisa *Research GAP* Tinjauan *Literature Review*

No	Jurnal	Masalah	Metode	Hasil	Research GAP
1.	[13]	Bagaimana implementasi analisis sentimen dengan pendekatan <i>Deep Learning</i> pada beberapa penelitian?	<i>Systematic Mapping Study</i>	Faktor kelemahan yang perlu diantisipasi dalam penelitian analisis sentimen dengan pendekatan <i>deep learning</i> : <ul style="list-style-type: none"> ▪ Keterbatasan dataset ▪ Dukungan sumber daya dalam labelling data. ▪ Format narasi <i>text</i> tidak terstruktur sehingga perlu tahap <i>pre-processing data</i>. ▪ Perlunya ada standarisasi perangkat dan metode analisis sentimen. 	Penelitian tidak menjelaskan apakah implementasi model dan lingkup area penelitian memiliki pengaruh dalam implementasi analisis sentimen
2.	[2]	Bagaimana pengaruh opini siswa terkait peningkatan kualitas pendidikan ?	<i>Aspect-Based Sentiment Analysis</i> (ABSA)	Analisis sentimen siswa merespons pengajaran dan pembelajaran, sehingga dapat membantu dalam meningkatkan kualitas pendidikan.	Dataset melalui <i>survey</i> kepada siswa tidak mengevaluasi parameter penilaian misalnya kualitas pengajaran, infrastruktur pendukung, kurikulum dsb.
3.	[12]	Bagaimana mengukur analisa kualitatif capaian pembelajaran melalui pendapat guru dan siswa ?	<i>Student Performance Quantifier using Sentiment analysis</i> dengan <i>Valence Aware Dictionary and Sentiment Reasoner</i>	Penelitian analisis sentimen merupakan solusi analisa kualitatif pada pengukuran capaian pembelajaran. Hasil akurasi analisis sentimen selain dengan penerapan model klasifikasi juga didukung dengan strategi optimisasi model dengan pemanfaatan teknik <i>Valence Aware Dictionary and Sentiment Reasoner</i>	Penelitian menggunakan dataset kecil sehingga belum dapat dilihat pengaruh model optimisasi algoritma efektif atau tidak dengan penggunaan dataset yang lebih besar
4.	[14]	Bagaimana menentukan polaritas konsumen berdasarkan hasil analisis sentimen ?	Multinomial Naive Bayes (MNB) dan Penerapan <i>Particle Swarm Optimization</i> (PSO)	Nilai akurasi model <i>Naive Bayes Classifier</i> (NBC) setelah dioptimalkan menggunakan <i>Particle Swarm Optimization</i> (PSO) adalah antara 89% hingga 91.76%.	Dalam penelitian tidak dijelaskan apakah selain model PSO dapat mengupayakan tingkat optimisasi pada proses <i>preprocessing</i> data sebelum uji model terkait <i>feature selection</i> dan <i>feature extraction</i> .
5.	[15]	Bagaimana strategi teknik dalam analisis sentimen ?	Metode <i>Literature Review</i>	Model klasifikasi seperti SVM, <i>Naive Bayes</i> dan <i>Neural Network</i> memiliki akurasi tertinggi dan dianggap sebagai metode pembelajaran dasar pada penelitian analisis sentimen.	Dalam penelitian tidak dijelaskan metode pengujian dari model klasifikasi apakah hanya komparasi akurasi dari penggunaan model atau dapat dilakukan teknik integrasi model dalam suatu analisis.
6.	[16]	Bagaimana menganalisis sentimen dalam teks berbahasa Tionghoa.?	Pendekatan hibrida yang menggunakan kamus sentimen dan algoritma <i>Naive Bayes</i> serta kamus Tionghoa dari Universitas Teknologi Dalian (DUT)	Hasil akurasi dari pendekatan ini mencapai 88.2%, yang menunjukkan peningkatan relatif sebesar 17.6% dibandingkan dengan pendekatan N-gram-NB, 20.0% dibandingkan dengan pendekatan N-gram-SVM, dan 14.8% dibandingkan dengan pendekatan TextCNN.	Penelitian tidak menggunakan teknik pelabelan manual, padahal untuk substansi data analisis sentimen selain Bahasa Inggris ada baiknya menggunakan multiple proses pelabelan data baik secara manual dan menggunakan kamus data.
7.	[17]	Analisis data sentimen penggunaan	Implementasi Model SVM dan <i>Naive Bayes</i>	Akurasi SVM memberikan hasil yang lebih baik dalam menganalisis sentimen pengguna terhadap e-wallet dibandingkan	Penelitian ini tidak adanya penggunaan teknik seleksi fitur untuk meningkatkan kinerja algoritma.

No	Jurnal	Masalah	Metode	Hasil	Research GAP
		model e-wallet di masyarakat ?		Naïve Bayes. SVM memiliki akurasi sebesar 98,6%.	
8.	[18]	Latar belakang permasalahan	Metode Naïve Bayes Vectorization	Improved <i>Naïve Bayes Vectorization Technique</i> telah menghasilkan peningkatan kinerja dalam klasifikasi dokumen teks menggunakan SVM. Teknik vectorisasi ini berhasil mengurangi dimensi data dan berkontribusi pada peningkatan kinerja pendekatan klasifikasi SVM.	Penelitian ini fokus pada penggunaan Improved Naïve Bayes Vectorization Technique dalam klasifikasi dokumen teks menggunakan SVM, namun tidak memberikan analisis mendalam mengenai batasan atau kelemahan dari pendekatan tersebut yang bisa menjadi lanjutan penelitian.
9.	[19]	Bagaimana analisa klasifikasi teks dipengaruhi oleh feature selection seperti metode Chi-Square ?	Metode <i>Chi-Square</i>	Metode Chi-square (yang disebut ImpCHI) dapat meningkatkan kinerja klasifikasi teks bahasa Arab. Penelitian ini juga menunjukkan bahwa penggunaan teknik <i>stemming</i> dan pemilihan fitur secara bersamaan dapat meningkatkan kinerja klasifikasi teks bahasa Arab dalam hal recall, presisi, dan <i>f-measure</i> .	Meskipun penelitian ini mengevaluasi kinerja metode pemilihan fitur dalam hal recall, presisi, dan <i>f-measure</i> , sebaiknya ada penambahan evaluasi tambahan seperti metode <i>confusion matrix</i> untuk mengevaluasi kinerja model klasifikasi.
10.	[20]	Bagaimana mengevaluasi kinerja teknik seleksi fitur yang berbeda untuk analisis sentimen ?	Metode integrasi <i>Logistic Regression</i> (LR), <i>Support Vector Machines</i> (SVM), <i>Decision Tree</i> , dan <i>Naive Bayes</i>	Hasil penelitian menunjukkan bahwa teknik seleksi fitur yang terbaik, ketika dilatih menggunakan metode ensemble, mencapai hasil yang luar biasa pada analisis sentimen. Teknik seleksi fitur yang dilatih menggunakan klasifikasi ensemble juga menunjukkan kinerja yang lebih baik.	Penelitian ini hanya mempertimbangkan teknik seleksi fitur untuk meningkatkan kinerja analisis sentimen, tanpa mempertimbangkan faktor-faktor lain yang dapat mempengaruhi kinerja analisis sentimen, seperti kualitas data, preprocessing data, atau pengaturan hiperparameter yang tepat.

Berdasarkan informasi pada Tabel 2, dapat dipahami disimpulkan beberapa hal terkait dengan penelitian dalam konteks analisis sentimen yaitu :

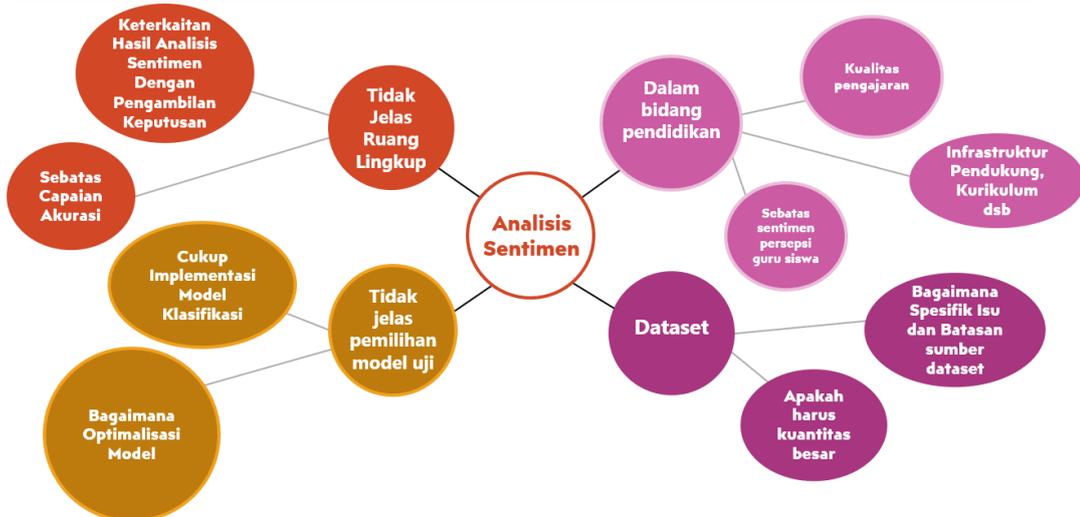
- Optimalisasi penerapan analisis sentimen dengan pendekatan *deep learning* dengan beberapa faktor seperti dataset, metode pelabelan data, *preprocessing* data berfungsi untuk mempersiapkan dataset dan memperbaiki data terkait data yang mungkin dalam kondisi tidak terstruktur dan akan mempengaruhi dalam metode *Natural Language Processing* (NLP) dalam pengujian model klasifikasi pada proses analisis sentimen.
- Dalam pemanfaatan di bidang pendidikan, analisis sentimen sangat bermanfaat untuk mengevaluasi capaian hasil pembelajaran dari pendapat pihak-pihak yang terlibat seperti siswa dan guru untuk meningkatkan kualitas pembelajaran dan pengajaran.
- Analisis sentimen sebagai pendekatan evaluasi kualitatif yang didukung dengan pemodelan klasifikasi pada implementasi machine learning.
- Implementasi model klasifikasi *Naive Bayes* akan mendapatkan nilai akurasi terbaik apabila diintegrasikan dengan metode optimalisasi algoritma.
- Model klasifikasi SVM lebih optimal dalam menampilkan nilai akurasi dari analisis sentimen dibandingkan dengan model klasifikasi *Naive Bayes*, akan tetapi apabila kedua model digabungkan seperti model *Naive Bayes Vectorization Technique* menghasilkan nilai akurasi lebih baik daripada sekedar dengan pemodelan SVM saja atau *Naive Bayes* saja.
- Strategi optimalisasi pemodelan klasifikasi *Naive Bayes* dan SVM pada analisis sentimen dapat menggunakan teknik *feature selection* seperti *Chi-Square*.



Gambar 2. Faktor-Faktor Yang Mempengaruhi Analisis Sentimen

Selain itu berdasarkan informasi pada Tabel 2, *Research GAP* yang perlu diperhatikan untuk disempurnakan dalam penelitian analisis sentimen adalah :

Research GAP



Gambar 3. *Research GAP* Hasil *Literature Review*

Pada Gambar 3. dapat dipahami berdasarkan hasil kajian *literature review* bahwa kesenjangan hasil penelitian terkait analisis sentimen dengan model klasifikasi SVM dan Naive Bayes yaitu :

- a. Beberapa penelitian belum menetapkan dengan jelas ruang lingkup penelitian seperti keterkaitan hasil analisis sentimen dengan proses pengambilan keputusan, karena sebatas untuk mengevaluasi capaian akurasi.
- b. Batasan penelitian terjadi pada beberapa studi kasus penelitian yang telah dilakukan adalah tidak jelasnya model uji khususnya dalam konteks pengujian text mining dengan model klasifikasi serta proses optimalisasi model.
- c. Pada beberapa penelitian analisis sentimen di bidang pendidikan umumnya mengevaluasi opini guru dan siswa belum banyak yang melakukan penelitian terkait dengan analisa sentimen bagaimana opini terkait keberadaan infrastruktur, kurikulum terkait capaian hasil pembelajaran dari perspektif siswa.
- d. Selain itu hasil analisa *research GAP* dari literature review mengenai penelitian mengenai analisa sentiment masih memiliki kekurangan dalam dataset terkait spesifik isu permasalahan yang diangkat dalam penelitian serta dataset.

Selain itu pada penelitian mengenai perbandingan dari implementasi Algoritma *Naive Bayes* dan SVM pada penelitian analisa sentiment dengan menggunakan dataset dari hasil *review* [21] didapatkan komparasi antara kedua model klasifikasi tersebut yaitu :

a. Kelebihan Model Klasifikasi Support Vector Machine (SVM)

1. SVM efektif dalam menangani dataset dengan dimensi tinggi dan memiliki kemampuan untuk menangani data yang tidak linier dengan menggunakan kernel.
2. SVM cenderung memberikan hasil yang lebih akurat dalam klasifikasi data yang kompleks dan tidak terstruktur.

b. Kekurangan Model Klasifikasi Support Vector Machine (SVM)

1. SVM membutuhkan waktu komputasi yang lebih lama terutama pada dataset yang besar.
2. SVM cenderung sulit untuk diinterpretasikan secara intuitif karena kompleksitas modelnya.

c. Kelebihan Model Klasifikasi Naive Bayes

1. *Naive Bayes* mudah diimplementasikan dan cepat dalam proses pelatihan dan klasifikasi.
2. *Naive Bayes* efektif dalam menangani dataset dengan fitur yang tidak terkait secara langsung.

d. Kekurangan Model Klasifikasi Naive Bayes

1. Naive Bayes diasumsikan bahwa semua fitur dalam dataset adalah independen, yang mungkin tidak selalu terjadi dalam praktiknya.
2. Naive Bayes cenderung kurang efektif dalam menangani data yang kompleks dan tidak linier.

Dengan mempertimbangkan kelebihan dan kekurangan dari kedua model tersebut, pemilihan model yang sesuai harus didasarkan pada karakteristik dataset dan tujuan analisis sentimen yang ingin dicapai.

3.2. Analisis Research Question

Pada penelitian ini Analisis *Reserach Question* dapat dijelaskan dengan dukungan hasil *literature review* pada penelitian sebelumnya adalah :

Tabel 3. Analisis *Research Question*

<i>Research Question</i>	Hasil
a. Apakah model klasifikasi SVM (<i>Support Vector Machine</i>) dan <i>Naive Bayes</i> dapat optimal digunakan pada analisis sentimen ?	<ol style="list-style-type: none"> a. Model SVM memiliki performa yang baik dalam klasifikasi sentimen . Selain itu, SVM telah terbukti efektif dalam mendeteksi teks dan menghasilkan hasil yang baik dengan dukungan optimalisasi model misalnya dengan <i>feature extraction</i> seperti <i>Global Vectors for Word Representation</i> [22]. b. Dalam penelitian ini, model klasifikasi <i>Naive Bayes</i> digunakan untuk analisis sentimen pada data Twitter terkait vaksinasi Covid-19. Metode Naive Bayes adalah metode klasifikasi teks yang didasarkan pada probabilitas kata kunci dalam membandingkan dokumen pelatihan dan dokumen uji. Meskipun tidak disebutkan secara eksplisit apakah model Naive Bayes optimal digunakan, namun hasil penelitian menunjukkan bahwa akurasi metode Naive Bayes adalah sebesar 82.51% . Oleh karena itu, dapat disimpulkan bahwa model klasifikasi <i>Naive Bayes</i> memberikan tingkat akurasi yang baik dalam analisis sentimen pada penelitian ini [23].
b. Faktor-faktor apa saja yang mempengaruhi keberhasilan penelitian analisis sentimen ?	<ol style="list-style-type: none"> a. Faktor-faktor yang mempengaruhi keberhasilan penelitian analisis sentimen adalah pemilihan metode analisis, kategori penelitian, jumlah dataset, serta kerangka kerja yang ditetapkan atau tahapan proses analisis data sentimen serta optimalisasi model klasifikasi dengan teknik seperti <i>feature selection</i> dan <i>feature extraction</i> [11].
c. Bagaimana strategi menghadapi tantangan dalam penelitian selanjutnya mengenai implementasi metode analisis sentimen pada industri pendidikan ?	<ol style="list-style-type: none"> a. Strategi menghadapi tantangan dalam penelitian dalam implementasi metode analisis sentimen di industri pendidikan yaitu : pengembangan model <i>machine learning</i>, penggunaan teknik <i>aspect-based sentiment analysis</i>, pengelolaan data yang lebih lanjut, dan pengembangan sentiment lexicon [2].

3.3 Kategori Parameter Yang Berpengaruh Pada Implementasi Data Mining

Kategori parameter yang berpengaruh pada implementasi *Data Mining* dalam penelitian yang telah dilakukan pada konteks penerapan analisis sentimen [24] yaitu meliputi :

1. **Preprocessing Data** : Meliputi langkah-langkah seperti tokenisasi, stemming, dan pembersihan data untuk mempersiapkan data mentah sebelum proses analisis.
2. **Sentiment Analysis** : Menganalisis sentimen dari teks yang diekstrak dari Twitter terkait protes petani dengan menggunakan model klasifikasi untuk mengidentifikasi polaritas dan subjektivitas.

3. **Teknik Analisis** : Penggunaan teknik misalnya seperti Bag of Words dan TF-IDF untuk mengonversi teks menjadi format numerik guna analisis lebih lanjut.
4. **Klasifikasi Data** : Penggunaan empat model klasifikasi yaitu *Naive Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* untuk tujuan prediksi sentimen.
5. **Visualisasi Data** : Penggunaan perpustakaan visualisasi untuk melakukan analisis mendalam terhadap data *Twitter* yang diekstrak.
6. **Evaluasi Model** : Penggunaan indikator kinerja seperti presisi, *recall*, akurasi, dan matriks kebingungan untuk mengevaluasi model klasifikasi yang digunakan.

Apabila kinerja salah satu parameter dalam implementasi *Data Mining* tidak berjalan optimal, hal ini dapat memiliki beberapa implikasi, antara lain :

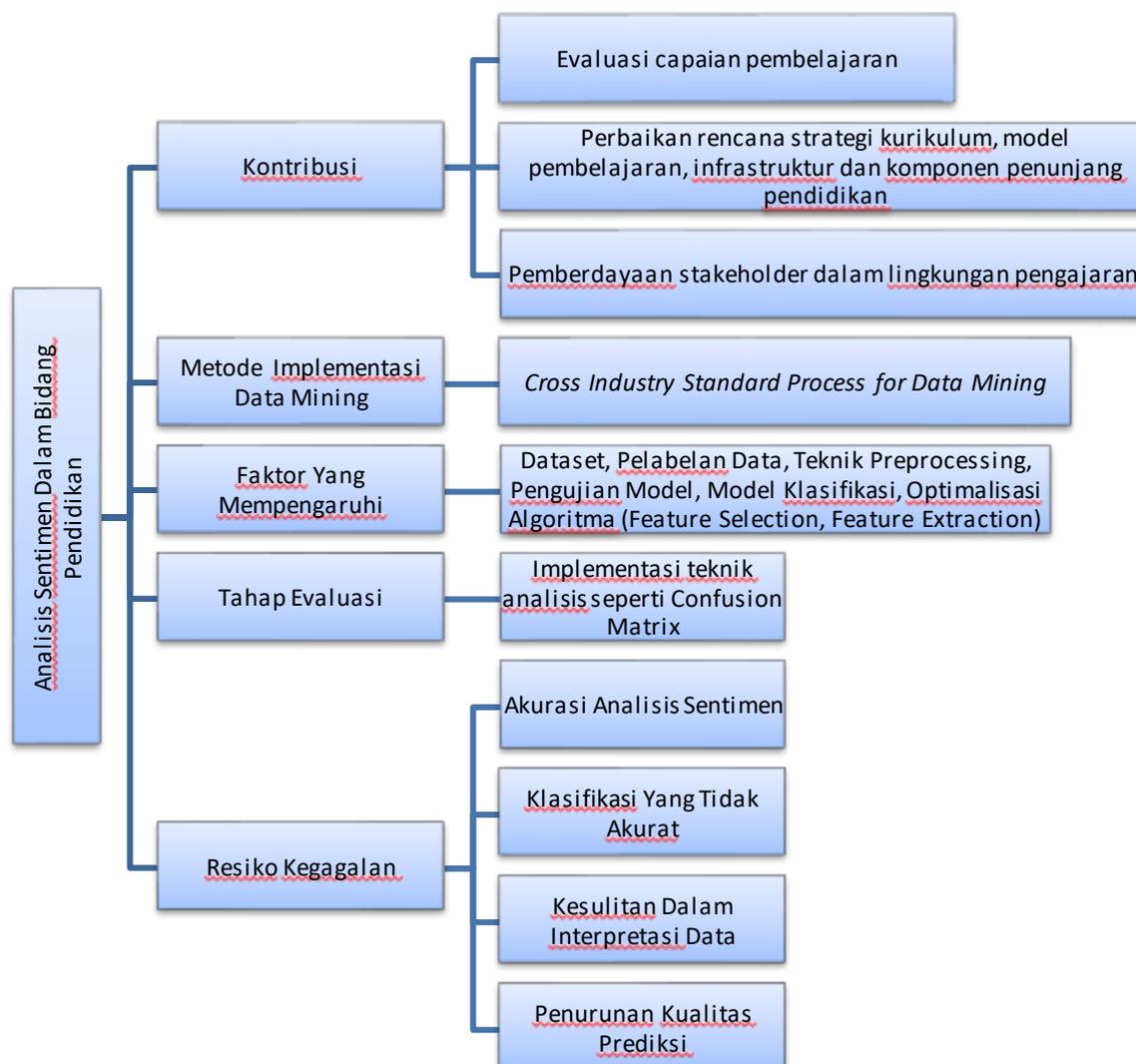
1. **Akurasi Analisis Sentimen** : Jika salah satu parameter seperti proses preprocessing data tidak dilakukan dengan baik, dapat mengakibatkan noise atau informasi yang tidak relevan tetap ada dalam data, yang pada gilirannya dapat mempengaruhi akurasi analisis sentimen.
2. **Klasifikasi Yang Tidak Akurat** : Jika model klasifikasi tidak dioptimalkan dengan baik, misalnya jika teknik analisis seperti Bag of Words atau TF-IDF tidak diterapkan dengan benar, maka hasil klasifikasi sentimen dapat menjadi tidak akurat dan tidak dapat diandalkan sebaiknya perlu diperbaiki dengan menggunakan metode *feature extraction* dan *feature selection* yang tepat.
3. **Kesulitan Dalam Interpretasi Data** : Jika parameter visualisasi data tidak berjalan optimal, interpretasi hasil analisis dapat menjadi sulit, sehingga pemahaman yang mendalam tentang sentimen publik terhadap protes petani mungkin tidak dapat dicapai.
4. **Penurunan Kualitas Prediksi** : Jika salah satu parameter dalam proses klasifikasi data tidak berjalan optimal, seperti pemilihan model klasifikasi yang tidak sesuai, dapat mengakibatkan penurunan kualitas prediksi sentimen, yang pada akhirnya dapat mempengaruhi keseluruhan hasil penelitian.

Oleh karena itu, penting untuk memastikan bahwa setiap parameter dalam implementasi *Data Mining* dijalankan dengan optimal agar hasil analisis sentimen yang diperoleh dapat menjadi lebih akurat dan bermanfaat.

3.4 Konseptual Framework

Analisa konseptual *framework* pada hasil literature review pada penelitian ini bertujuan untuk menyediakan panduan yang terstruktur dan sistematis untuk memahami hubungan antara berbagai elemen yang terlibat dalam penelitian mengenai analisis data sentimen dengan menetapkan kriteria yang jelas untuk mengevaluasi pada penelitian analisis data sentimen pada bidang pendidikan. Dengan demikian, kerangka konseptual menjadi alat yang sangat penting dalam menyusun dan menyajikan hasil literature review yang komprehensif dan sistematis.

Sebagai strategi yang dapat diusulkan pada penelitian selanjutnya terkait dengan konteks penelitian analisis sentimen dengan objektivitas pada penerapan di industri pendidikan maka hasil dari analisa *literature review* yang telah dilakukan peneliti merumuskan konseptual *framework* adalah :



Gambar 4. Konseptual *Framework*

Pada Gambar 4 tersebut hal-hal intensif yang perlu diperhatikan dalam melakukan penelitian terkait analisis sentimen adalah apa kontribusi dari penelitian analisis sentimen, penetrasi dengan metode *data mining*, perhatikan faktor yang mempengaruhi penelitian, tahap evaluasi serta resiko kegagalan penelitian. Berikut analisa yang dapat di evaluasi adalah :

1. Perspektif Kontribusi

Beberapa penelitian terkait analisis sentimen di bidang pendidikan sebatas dalam mengevaluasi tingkat akurasi hasil data sentimen dengan model klasifikasi, sebaiknya dalam penelitian yang nantinya akan dilakukan dengan konteks yang sama dapat memberikan kontribusi penelitian untuk mengevaluasi capaian pembelajaran, perbaikan rencana strategi kurikulum, model pembelajaran, infrastruktur dan komponen penunjang pendidikan, serta bisa juga mengarah pada kontribusi penelitian kepada aspek pemberdayaan *stakeholder* dalam lingkungan pengajaran.

2. Perspektif Metode Implementasi Data Mining

Hasil *literature review* dari beberapa penelitian menjelaskan siklus pada penerapan penelitian data mining dengan implementasi model klasifikasi seperti *Naive Bayes* dan *Support Vector Machine* akan lebih baik didukung dengan implementasi Metode *Cross Industry Standard Process for Data Mining*, dimana metode ini merupakan panduan umum yang membantu peneliti dalam menyusun dan melaksanakan proyek *data mining* secara sistematis dan terstruktur.

3. Perspektif Faktor Yang Mempengaruhi

Peneliti juga mengevaluasi dari hasil *literature review* bahwa pada penelitian analisis data sentimen perlu sangat memperhatikan komponen-komponen terkait dengan dataset, pelabelan data, teknik *preprocessing*, pengujian model, model klasifikasi, serta metode optimalisasi algoritma misalnya seperti pemanfaatan fitur *data selection* dan *data extraction* pada proses *pre-processing* dataset yang akan dievaluasi.

4. Perspektif Evaluasi

Berdasarkan hasil *literature review* yang telah dilakukan didapatkan bahwa analisa hasil evaluasi pengujian model klasifikasi seperti Naive Bayes dan SVM akan lebih baik disajikan dengan implementasi model visual menggunakan salah satunya Model *Confusion Matrix*.

5. Perspektif Resiko Kegagalan

Peneliti juga perlu memperhatikan aspek resiko kegagalan yang mungkin dapat terjadi dari proses penelitian terkait analisis data sentimen seperti permasalahan capaian akurasi data, penerapan model klasifikasi yang tidak sesuai, kesulitan dalam proses hasil interpretasi data serta penurunan kualitas prediksi hasil pengolahan data analisis sentimen yang telah dilakukan.

4. PENUTUP

Penelitian analisa *literature review* ini menyimpulkan bahwa berdasarkan beberapa penelitian yang telah dilakukan terkait analisis sentimen, maka ada beberapa hal yang perlu diperhatikan yaitu aspek kontribusi misalnya pada bidang pendidikan upayakan kontribusi tidak hanya sebatas evaluasi nilai akurasi saja, selain itu dikarenakan analisis sentimen merupakan salah satu topik penelitian di bidang *Data Mining* maka sebaiknya tahapan penelitian mengikuti Metode *Cross Industry Standard Process for Data Mining*.

Faktor-faktor dominan yang memiliki pengaruh dalam penelitian analisis sentimen misalnya aspek dataset, pelabelan data, teknik preprocessing data, pengujian model, model klasifikasi yang digunakan, serta aspek optimalisasi algoritma. Selain itu perlu diperhatikan metode dalam evaluasi data hasil pengujian model, misalnya bisa menggunakan teknik *Confusion Matrix Analysis* dan faktor-faktor yang dapat menjadi resiko kegagalan penelitian analisis sentimen perlu diantisipasi oleh peneliti agar tidak terjadi dan menghambat hasil penelitian.

Adapun *future research* yang sebaiknya dilakukan berdasarkan hasil penelitian *literature review* ini adalah penelitian dibidang data mining dengan konteks analisis data sentimen khususnya di ruang lingkup instansi pendidikan yang bertujuan mengevaluasi capaian pembelajaran, perbaikan rencana strategi kurikulum, model pembelajaran, infrastruktur dan komponen penunjang pendidikan, serta bisa juga mengarah pada kontribusi penelitian kepada aspek pemberdayaan *stakeholder* dalam lingkungan pengajaran. Penelitian sebaiknya menerapkan pengujian model klasifikasi seperti Naive Bayes dan SVM dengan dukungan metode optimalisasi algoritma agar capaian akurasi data lebih baik.

UCAPAN TERIMAKASIH

Alhamdulillah peneliti mengucapkan terima kasih kepada bimbingan Bapak Denni Kurniawan, S.T., M.TI., Ph.D atas bimbingannya dalam penelitian *paper review* yang telah dilakukan serta dukungan Program Studi Magister Ilmu Komputer Fakultas Teknologi Universitas Budi Luhur dalam proses penelitian.

DAFTAR PUSTAKA

- [1] A. A. Q. Aqlan, B. Manjula, and R. Lakshman Naik, *A study of sentiment analysis: Concepts, techniques, and challenges*, vol. 28. Springer Singapore, 2019. doi: 10.1007/978-981-13-6459-4_16.
- [2] P. A. and Y. K. M. Ganpat Singh Chauhan, "Aspect-Based Sentiment Analysis of Students' Feedback to Improve Teaching-Learning Process," *Asp. Sentim. Anal. Students' Feed. to Improv. Teaching-Learning Process*, vol. 2, no. January, pp. 83-93, 2019, doi: 10.1007/978-981-13-1747-7.
- [3] M. K. Jiawei Han, *Data Mining : Concepts and Techniques*, no. Vol 3. 2012. [Online]. Available: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=PT%0Ahttp://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52012PC0011:pt:NOT>
- [4] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048-3061, 2021, doi: 10.1109/TKDE.2019.2962680.
- [5] A. N. H. Ananda Kejora Rotty, Triwulandari Satitidjati Dewayana, "Cross-Industry Standard Process for Data Mining (CRISP-DM) Approach in Determining the Most Significant Employee Engagement Drivers to Sales at X Car Dealership," *Proc. 3rd Asia Pacific Int. Conf. Ind. Eng. Oper. Manag. Johor Bahru, Malaysia, Sept. 13-15, 2022*, pp. 3368-3379, 2023, doi: 10.46254/ap03.20220552.
- [6] Y. Indulkar and A. Patil, "Comparative study of machine learning algorithms for twitter sentiment analysis," *2021 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2021*, no. July 2021, pp. 295-299, 2021, doi: 10.1109/ESCI50559.2021.9396925.
- [7] A. Y. Mucahid Mustafa Saritas, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7(2), pp. 88-91, 2019, doi: 10.1039/b000000x.
- [8] K. Korovkinas and P. Danenas, "SVM and Naive Bayes Classification Ensemble Method for Sentiment

- Analysis,” *Balt. J. Mod. Comput.*, vol. 5, no. 4, pp. 398–409, 2017, doi: 10.22364/bjmc.2017.5.4.06.
- [9] V. H. Nhu *et al.*, “Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082749.
- [10] M. Guia, R. R. Silva, and J. Bernardino, “Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis,” *IC3K 2019 - Proc. 11th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag.*, vol. 1, no. Ic3k, pp. 525–531, 2019, doi: 10.5220/0008364105250531.
- [11] O. Y. Adwan, M. Al-Tawil, A. M. Huneiti, R. A. Shahin, A. A. Abu Zayed, and R. H. Al-Dibsi, “Twitter sentiment analysis approaches: A survey,” *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 15, pp. 79–93, 2020, doi: 10.3991/ijet.v15i15.14467.
- [12] J. Watkins, M. Fabielli, and M. Mahmud, “SENSE: A Student Performance Quantifier using Sentiment Analysis,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 0–5, 2020, doi: 10.1109/IJCNN48605.2020.9207721.
- [13] Z. Kastrati, F. Dalipi, A. S. Imran, K. P. Nuci, and M. A. Wani, “Sentiment analysis of students’ feedback with nlp and deep learning: A systematic mapping study,” *Appl. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/app11093986.
- [14] S. Khomsah, “Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews,” *J. Penelit. Pos dan Inform.*, vol. 10, no. 2, p. 157, 2020, doi: 10.17933/jppi.2020.100206.
- [15] P. Mehta and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [16] Z. Li, R. Li, and G. Jin, “Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary,” *IEEE Access*, vol. 8, pp. 75073–75084, 2020, doi: 10.1109/ACCESS.2020.2986582.
- [17] D. A. Kristiyanti, D. A. Putri, E. Indrayuni, A. Nurhadi, and A. H. Umam, “E-Wallet Sentiment Analysis Using Naïve Bayes and Support Vector Machine Algorithm,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012079.
- [18] R. P. M. Hajah T. Sueno, Bobby D. Gerardo, “Multi-class Document Classification using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3937–3944, 2020, doi: 10.30534/ijatcse/2020/216932020.
- [19] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, “Feature selection using an improved Chi-square for Arabic text classification,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 2, pp. 225–231, 2020, doi: 10.1016/j.jksuci.2018.05.010.
- [20] A. Madasu and S. Elango, “Efficient feature selection techniques for sentiment analysis,” *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 6313–6335, 2020, doi: 10.1007/s11042-019-08409-z.
- [21] A. M. Rahat, A. Kahir, and A. K. M. Masum, “Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset,” *Proc. 2019 8th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2019*, pp. 266–270, 2020, doi: 10.1109/SMART46866.2019.9117512.
- [22] A. R. FITRIANSYAH, “Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta - Bandung pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding,” *J. Tugas Akhir Fak. Inform.*, vol. 9, no. 5, pp. 6078–6083, 2022, [Online]. Available: <https://repository.telkomuniversity.ac.id/home/catalog/id/181602>
- [23] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, “Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.
- [24] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, 2021, doi: 10.1016/j.jjime.2021.100019.