Faktor Exacta

Vol. 17, No. 4, Dec 2024, pp. 334~345
eISSN: 2502-339X, pISSN: 1979-276X, DOI: 10.30998/faktorexacta.v17i4.24751

Optimalisasi Model Klasifikasi *Naive Bayes* dan *Support Vector Machine* Dengan *Fast Text* dan *Chi Square*

Riri Fajriah[1], Denni Kurniawan[2]
[1,2]Program Studi Magister Ilmu Komputer
Fakultas Teknologi Informasi, Universitas Budi Luhur, Indonesia
[1]2211600941@student.budiluhur.ac.id, [2]denni.kurniawan@budiluhur.ac.id

Article Info

Article history:

Received Jun 11, 2024
Revised Dec 08, 2024
Accepted Jan 29, 2025

Keywords:

Sentiment Analysis
CRISP-DM
Naive Bayes
Support Vector Machine
Fast Text
Chi Square

ABSTRACT

The effectiveness of programming education at the Faculty of Computer Science, Universitas Mercu Buana, is crucial for producing graduates with industry-relevant competencies. However, evaluations reveal that many graduates still lack adequate programming skills. This study aims to address this issue by analyzing the sentiments of students, lecturers, and alumni regarding the programming education process. Using an online questionnaire as the primary data source, sentiment analysis was conducted with the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. Two classification models, Naive Bayes and Support Vector Machine (SVM), were optimized with FastText for feature extraction and Chi-Square for feature selection to enhance performance. The results demonstrate significant improvements in accuracy, with Naive Bayes achieving 90.49% and SVM reaching 99.58% after optimization. This research highlights the importance of combining advanced feature extraction and selection techniques to improve sentiment analysis accuracy. The findings provide actionable insights for enhancing programming education strategies and aligning them with industry demands.

Corresponding Author:

Riri Fajriah,
Program Studi Magister Ilmu Komputer,
Universitas Budi Luhur,
Jl. Ciledug Raya No.99, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan,
Email: 2211600941@student.budiluhur.ac.id

1. PENDAHULUAN

Dalam era di mana teknologi menjadi pilar utama dalam menjalankan operasi bisnis dan mencapai keunggulan kompetitif, kemampuan untuk mengembangkan dan memelihara aplikasi bisnis yang inovatif dan efisien menjadi semakin penting [1]. Pemahaman mendalam tentang pemrograman aplikasi bisnis memungkinkan organisasi untuk mengotomatisasi proses, meningkatkan pengalaman pelanggan, dan mengelola data dengan lebih baik. Oleh karena itu, permintaan akan para profesional Teknologi Informasi yang terampil dalam pemrograman aplikasi bisnis terus tumbuh sejalan dengan transformasi digital yang terus berkembang.

Mensikapi dinamika yang terjadi saat ini, Fasilkom (Fakultas Ilmu Komputer) Universitas Mercu Buana perlu meningkatkan sasaran mutu pembelajaran dengan mempersiapkan lulusan yang memiliki kompetensi khususnya dalam bidang pemrograman sesuai kualifikasi kebutuhan industri kerja. Akan tetapi berdasarkan hasil sharing knowledge dengan alumni didapatkan informasi bahwa belum banyak lulusan Fasilkom UMB yang siap bekerja di industri dan memiliki kemampuan pemrograman sesuai dengan kebutuhan pengembangan aplikasi bisnis saat ini. Kondisi ini kontradiktif dengan hasil penelitian sebelumnya dimana perubahan transformasi digital pada bisnis saat ini akan berdampak pada meningkatnya kebutuhan tenaga professional TI di organisasi bisnis dan literasi kemampuan digital karyawan perlu ditingkatkan [2]. Penelitian ini berorientasi pada prediksi data mining dengan skema analisa sentimen pihak-pihak terkait dengan proses pembelajaran pemrograman seperti mahasiswa, dosen pengajar dan alumni terkait dengan kondisi

pembelajaran pemrograman yang berjalan saat ini di Fasilkom Universitas Mercu Buana. Sumber data yang akan diolah dalam penelitian bersumber pada kuesioner online yang didistribusikan kepada para mahasiswa, dosen pengajar dan alumni yang bertujuan untuk observasi pendapat terkait sistem pembelajaran pemrograman yang pernah mereka rasakan selama perkuliahan. Sumber dataset menggunakan teknik penyebaran kuesioner online mengacu pada sebuah penelitian terkait *Educational Data Mining* (EDM) dengan melakukan analisa prediksi data mining pada pengolahan sumber data bersumber dari kusioner online pada internal universitas yang membantu menganalisa data faktor-faktor yang mempengaruhi drop out mahasiswa di universitas menggunakan metode *Bayesian Profile Regression* [3]. Serta sifat pertanyaan kuesioner akan lebih berorientasi kepada narasi pendapat dalam bentuk tipe data text dari para responden yang nantinya dianalisa dalam penelitian analisis sentimen.

Pengolahan data *sentiment analysis* pada penelitian ini dengan komparasi pada dua metode *Naive Bayes* dengan konsep perhitungan probabilitas dari seberapa kemungkinan suatu instance data sentiment termasuk dalam kategori atau kelas tertentu untuk pengujian klasifikasi data [4] dan penerapan metode *Support Vector Machine* (SVM) yang merupakan algoritma klasifikasi yang bertujuan untuk mencari *hiperplane* terbaik yang memisahkan dua kelas data dalam ruang fitur dengan margin maksimal. *Hiperplane* ini berfungsi sebagai pemisah optimal antara kelas data, dan SVM mencari *hiperplane* ini dengan mengoptimalkan margin yang lebih besar [5]. Naive Bayes dan SVM memiliki kekuatan yang berbeda dalam tugas klasifikasi. Naive Bayes cenderung baik dalam mengatasi masalah teks dan data yang memiliki asumsi independensi fitur yang lebih mendekati. Di sisi lain, SVM cenderung lebih baik dalam menangani data yang kompleks dan tidak teratur seperti pada penelitian sentiment analysis berdasarkan data sosial media dan memanfaatkan kekuatan kedua algoritma ini dapat meningkatkan akurasi klasifikasi [6].

Dalam sebuah penelitian dengan penerapan algoritma SVM didapatkan pengukuran parameter mutu pelayanan bagi pelanggan agar tetap dapat menjaga loyalitas pelanggan dengan dynamic information yang tepat [6]. Berdasarkan pertimbangan hal-hal tersebut maka peneliti mengangkat topik penelitian dengan judul Analisis Sentimen Terhadap Penyelenggaraan Pembelajaran Pemrograman di Fakultas Ilmu Komputer Universitas Mercu Buana Menggunakan Metode *Naive Bayes* dan *Support Vector Machine*.

Penelitian ini berfokus pada optimisasi model analisis sentimen menggunakan metode *Naive Bayes* dan SVM, dengan dukungan FastText untuk *feature extraction* dan *Chi-Square* untuk feature selection. Kombinasi metode ini bertujuan untuk meningkatkan akurasi klasifikasi sentimen yang dapat memberikan wawasan lebih mendalam tentang persepsi mahasiswa dan alumni terhadap pembelajaran pemrograman. Dengan temuan ini, Fakultas Ilmu Komputer diharapkan dapat mengembangkan strategi pembelajaran yang lebih adaptif dan relevan dengan kebutuhan pasar kerja. Adapun kontribusi utama penelitian ini meliputi implementasi metode optimasi pada model klasifikasi untuk analisis sentimen di bidang pendidikan dan menyediakan pendekatan baru dalam memanfaatkan algoritma *machine learning* untuk memahami opini pemangku kepentingan serta memberikan rekomendasi strategis untuk meningkatkan mutu pembelajaran pemrograman berdasarkan hasil analisis sentimen.

2. METODE

2.1. Metode *Cross Industry Standard Process for Data Mining* (CRISP-DM)

Metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) adalah suatu metode standar yang digunakan dalam proses *data mining* atau penambangan data [7]. Model ini terdiri dari 6 (enam) tahap proses yaitu sebagai berikut :

1. ***Business Understanding***
   Pemahaman bisnis merupakan tahap awal yaitu pemahaman penelitian, penentuan tujuan dan rumusan masalah *data mining*.
2. ***Data Understanding***
   Pemahaman data dalam tahap ini dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan. Pada tahap ini data dikumpulkan dan dikarakterisasi. Pada tahapan ini juga akan dilakukan analisis terhadap data yang dimiliki untuk mengenali data lebih lanjut dan untuk mencari pengetahuan awal terhadap data yang dimiliki. Sumber data yang digunakan pada penelitian ini adalah *survey online* kepada para mahasiswa dan alumni terkait opini (*sentiment process*) terkait sistem pembelajaran pemrograman di Fakultas Ilmu Komputer Universitas Mercu Buana.
3. ***Data Preparation***
   Pada tahap ini adalah proses yang dilakukan terkait dengan pembersihan dan persiapan data analisis. Selain itu proses yang dilakukan pada tahap ini adalah pemfilteran, penggabungan, transformasi, dan pemilihan fitur.
4. ***Modeling***
   Pada tahap ini adalah proses yang dilakukan adalah memilih teknik pemodelan yang sesuai dan sesuaikan aturan model untuk hasil yang maksimal. Dapat kembali ke tahap pengolahan untuk menjadikan data ke

dalam bentuk yang sesuai dengan model tertentu. Pada fase pemodelan, berbagai teknik (misalnya *association rule*, *decision tree*, *logistic regression*, *k-means clustering*) dipilih dan diterapkan dan parameternya dikalibrasi ke nilai optimal. Model yang berbeda dibandingkan, dan mungkin dikombinasikan.

5. **Evaluasi**
   Mengevaluasi satu atau model yang digunakan dan menetapkan apakah terdapat model yang memenuhi tujuan pada tahap awal. Kemudian menentukan apakah ada permasalahan yang tidak dapat tertangani dengan baik serta mengambil keputusan hasil penelitian.
6. ***Deployment***
   Pada fase terakhir, *deployment* (penerapan) direncanakan, diimplementasikan, dan dipantau. Seluruh proyek biasanya didokumentasikan dan dirangkum dalam sebuah laporan. Menggunakan model yang dihasilkan seperti pembuatan laporan atau penerapan proses data mining pada departemen lain.
7. ***Monitoring***
   Pantau kinerja model dan hasilnya dalam situasi produksi. Pastikan bahwa model tetap efektif dan relevan terhadap perubahan dalam data dan lingkungan bisnis.

2.2. *Naive Bayes*

Naive Bayes adalah sebuah metode klasifikasi yang berdasarkan pada teorema *Bayes*. Metode ini mengasumsikan bahwa setiap fitur dalam dataset adalah independen satu sama lain, yang dikenal sebagai asumsi naif (*naive assumption*). Meskipun asumsi ini tidak selalu mencerminkan keadaan nyata, *Naive Bayes* sering memberikan hasil yang baik dalam praktiknya dan memiliki kecepatan komputasi yang tinggi. *Bayesian Classification* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naïve Bayes* merupakan sebuah metode untuk teknik pengklasifikasian dengan konsep probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [8]. Adapun tahapan pada Algoritma *Naive Bayes* [9] adalah sebagai berikut :

1. Perhitungan jumlah kelas / label
2. Perhitungan jumlah kasus per kelas
3. Perkalian semua variabel kelas
4. Perbandingan hasil per kelas

Berikut ini merupakan rumus untuk menghitung teorema bayes :

$$P\ (H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan:
X= Data yang kelasnya tidak diketahui.
H= Data hipotesis X pada sebuah kelas.
P(H|X) = Probabilitas dari hipotesis H berdasarkan dari kondisi X.
P(H) = Probabilitas dari hipotesis H.
P(X|H) = Probabilitas dari X berdasarkan dari kondisi dalam hipotesis H.
P(X) = Probabilitas X

2.3. *Support Vector Machine*

SVM (*Support Vector Machine*) adalah sebuah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane terbaik untuk memisahkan dua kelas dalam ruang fitur. *Hyperplane* ini dipilih sedemikian rupa sehingga jarak antara hyperplane dan instance terdekat dari masing-masing kelas (yang disebut vektor pendukung) adalah maksimum. Berikut adalah beberapa konsep kunci terkait *Support Vector Machine* [10]:

Hyperplane : SVM berfokus pada penemuan *hyperplane* yang memisahkan dua kelas dalam ruang fitur. *Hyperplane* ini dapat dianggap sebagai batas keputusan yang membagi data ke dalam kelas-kelas yang berbeda. Dalam kasus SVM linier, *hyperplane* dapat dijelaskan oleh persamaan umum :

1. $f(x) = \vec{w}.\vec{x} + b = 0$
2. *f (x)* adalah fungsi perintah
3. w adalah vektor bobot (*weight*)
4. x adalah vektor fitur input
5. b adalah bias

2.4. *Feature Extraction*

Dalam analisis sentimen, *feature extraction* adalah proses menemukan dan mengekstraksi fitur yang paling relevan atau informatif dari teks atau data teks yang digunakan untuk analisis sentimen. Tujuan utama dari *feature extraction* adalah mengonversi data teks menjadi representasi numerik yang dapat digunakan oleh algoritma pembelajar mesin untuk memahami dan memprediksi sentimen.

FastText adalah sebuah teknik dan library untuk pembuatan word embeddings yang dikembangkan oleh Facebook. Berbeda dengan Word2Vec atau GloVe yang memetakan setiap kata ke dalam satu vektor, *FastText* membagi setiap kata menjadi subkata atau "n-grams" (potongan kata) dan memetakan masing-masing subkata ke dalam vektor. Ini memungkinkan *FastText* memahami makna kata berdasarkan kontribusi subkatanya[11].

2.5. *Feature Selection*

Metode umum untuk meningkatkan akurasi klasifikasi dan mengurangi jumlah fitur ruang yang signifikan adalah *Feature Selection* (FS). Mengekstraksi fitur yang paling relevan, yang dapat meningkatkan kinerja pendeteksian, merupakan langkah terpenting dalam pembelajaran mesin. Semua fitur yang ditemukan tidak signifikan, tetapi dapat membantu dalam identifikasi yang tepat. Oleh karena itu, metode untuk memilih atribut penting, yang memiliki informasi yang relevan dan diperlukan, digunakan untuk memilih dataset. *Feature selection* adalah proses pemilihan subset fitur yang paling relevan dan informatif dari keseluruhan fitur yang tersedia dalam data [12]. Dalam konteks sentimen analisis, feature selection berperan penting karena dapat membantu meningkatkan kinerja model dan mengurangi *overfitting*. Uji *Chi-Square* dapat digunakan untuk memilih fitur-fitur penting dari kumpulan data berdimensi tinggi [13]. Dalam penelitian ini akan disajikan metode *feature selection Chi-Square* untuk meminimalkan data dan menghasilkan akurasi klasifikasi yang lebih tinggi.

2.6. *Literature Review*

Dalam sebuah penelitian dapat dianalisa mengenai pengembangan solusi analisis sentimen dapat diaplikasikan secara umum untuk konteks Pendidikan dengan memperhatikan definisi sentimen yang konsisten dan pendekatan yang sesuai untuk lingkungan pendidikan [14]. Penelitian lainnya dengan menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA), dimana pada penelitian ini peneliti mendapatkan pemahaman mengenai bahwa dengan memahami opini siswa tentang aspek-aspek tertentu dari pengajaran dan pembelajaran dengan teknik *sentiment analysis* [15], maka penelitian ini dapat membantu dalam meningkatkan kualitas pendidikan dengan memberikan wawasan yang lebih baik tentang bagaimana siswa merespons pengajaran.

Selain itu pada penelitian dengan menggunakan metode SENSE (*Student pErformance quaNtifier using SEntiment analysis*), maka esensi penelitian yang dipahami oleh peneliti melalui penelitian ini adalah bahwa teknik analisis sentimen memungkinkan untuk diimplementasikan pada penelitian di sektor pendidikan untuk memperoleh pemahaman yang lebih mendalam tentang sikap dan persepsi siswa terhadap proses pembelajaran. Hal ini dapat membantu dalam mengidentifikasi faktor-faktor yang memengaruhi kinerja akademik siswa [16].

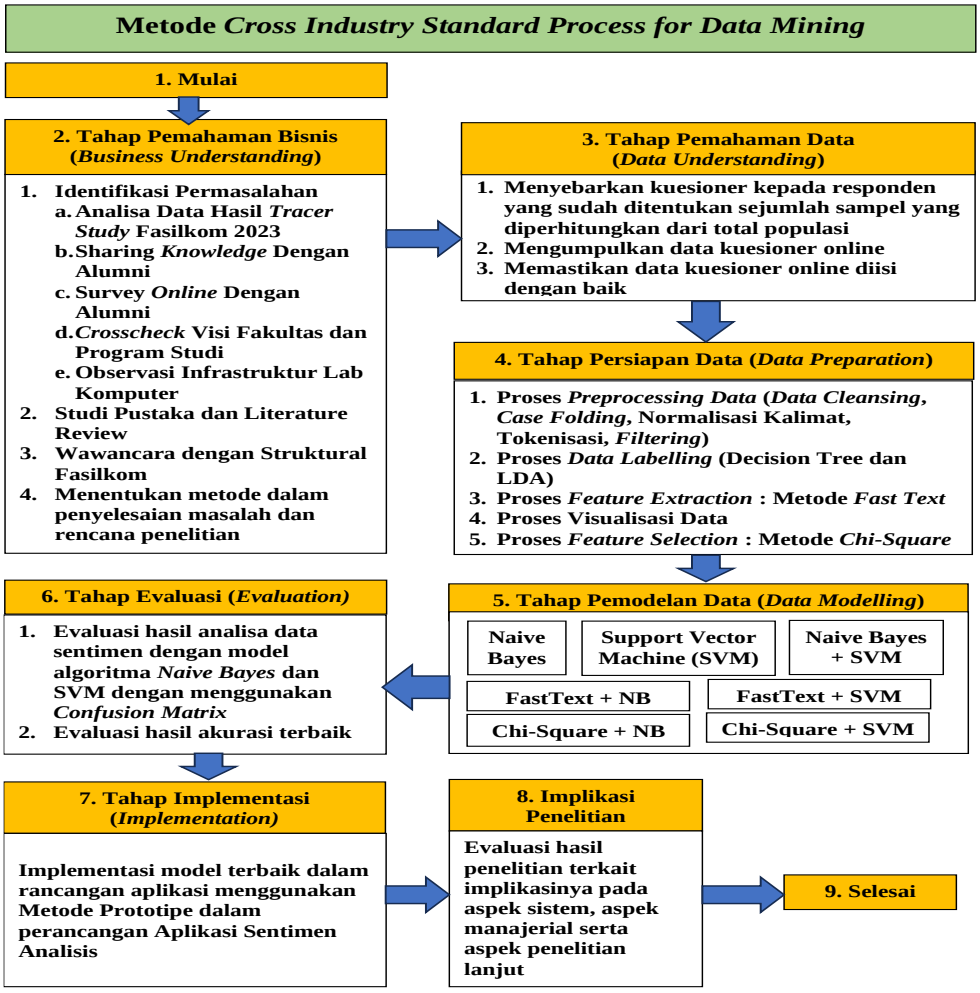
Selanjutnya pada penelitian dengan menggunakan metode *Multinomial Naive Bayes* (MNB) dan *Penerapan Particle Swarm Optimization* (PSO) untuk mengoptimalkan akurasi *Naive Bayes Classifier* (NBC) membantu peneliti untuk memahami mengenai bahwa penerapan metode optimasi seperti PSO dapat meningkatkan akurasi model NBC dalam analisis sentiment [17]. Oleh karena itu, penelitian selanjutnya dapat mengeksplorasi penggunaan metode optimasi lainnya untuk meningkatkan kinerja model analisis sentimen. Misalnya, penggunaan kombinasi metode *Naive Bayes* dan *Support Vector Machine* (SVM) atau *Naive Bayes* dan *Random Forest*. Penelitian lainnya yang menerapkan Metode *Literature Review* pada penelitian terkait analisis sentimen memberikan gambaran pemahaman bagi peneliti terkait dengan metode-metode *machine learning* seperti SVM, Naive Bayes, dan *neural networks* memiliki akurasi yang tinggi dan dapat dianggap sebagai metode pembelajaran dasar, sementara metode berbasis leksikon juga sangat efektif dalam mengolah data pada model algoritma yang digunakan pada penelitian terkait dengan *sentiment analysis* [18].

Penelitian lainnya dengan menggunakan Metode *Naïve Bayes* memberikan kontribusi pemahaman bagi peneliti terkait penelitian yang akan dilakukan yaitu penelitian ini menunjukkan bahwa pendekatan hibrida yang menggunakan kamus sentimen dan algoritma *Naïve Bayes* (NB) mampu mengklasifikasikan sentimen multi-dimensi dari data text seperti pesan danmaku dengan tingkat akurasi yang tinggi [19]. Hal ini dapat menjadi dasar untuk pengembangan metode analisis sentimen yang akan dilakukan dengan metode yang sama. Selain itu penelitian lain dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* memberikan pengetahuan bagi peneliti yaitu bahwa penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memberikan hasil yang lebih baik dalam menganalisis sentimen pengguna terhadap *e-wallet* dibandingkan dengan algoritma *Naïve Bayes*. Namun, penelitian ini juga menyarankan untuk menggunakan teknik seleksi fitur untuk meningkatkan kinerja algoritma [20].

Oleh karena itu pada penelitian ini, pengujian model klasifikasi dengan Algoritma *Naive Bayes* dan Algoritma *Support Vector Machine* akan dioptimalkan dengan implementasi *Feature Extraction Fast Text* dan *Feature Selection Chi Square*.

2.6. Kerangka Penelitian

Adapun langkah-langkah penelitian yang dilakukan adalah sebagai berikut :



Gambar 1. Kerangka Penelitian

Berdasarkan penggambaran kerangka penelitian pada Gambar 1 dapat dijelaskan sebagai berikut :

1. Langkah-langkah penelitian mengikuti tahapan pada Metode *Cross Industry Standard Process for Data Mining*
2. Pada tahap pemahaman bisnis (*business understanding*) peneliti melakukan observasi langsung pada kondisi infrastruktur, sarana dan prasarana ruang laboratorium komputer di Kampus Meruya Universitas Mercu Buana saat penyelenggaraan praktikum pemrograman. Lalu Peneliti melakukan studi pustaka dari buku-buku materi yang mendukung pemahaman atas konsep dan teori serta metode yang akan digunakan dalam penelitian.
3. Tahap Pemahaman Data (*Business Understanding*)
   Pada tahapan ini peneliti berupaya untuk membuat kuesioner yang menjadi alat bantu dalam pengumpulan data terkait penelitian analisa sentimen.
4. Tahap Persiapan Data (*Data Preparation*)
   Aktivitas utama dalam tahapan persiapan data adalah mempersiapkan dataset yang telah dikumpulkan untuk siap digunakan dalam proses penelitian analisis sentimen dengan melakukan hal-hal berikut ini :
   a. Proses *Preprocessing Data*
   b. Proses Data *Labelling*
   c. Proses *Feature Extraction* : Metode *Fast Text*
   d. Proses Visualisasi Data
   e. Proses *Feature Selection* : Metode *Chi-Square*

5. Tahap Pemodelan Data (*Data Modelling*)
   Pada tahapan ini peneliti akan melakukan pengujian data dengan model algoritma klasifikasi yang sudah ditentukan yaitu :
   a. Pertama, pengujian data analisa sentimen hanya dengan menggunakan model algoritma *Naive Bayes* saja.
   b. Kedua, pengujian data analisa sentimen hanya dengan menggunakan model algoritma *Support Vector Machine* saja.
   c. Ketiga, pengujian data analisa sentimen dengan menggabungkan model algoritma *Naive Bayes* dan *Support Vector Machine* dengan menggunakan metode *voting classifier*. *Voting Classifier* adalah salah satu bentuk penggabungan model (*ensemble learning*) dalam *machine learning. Voting Classifier* menggabungkan beberapa model pembelajaran mesin yang berbeda dan menggabungkan hasil prediksi mereka untuk menghasilkan hasil akhir [21].
   d. Keempat, pengujian data dengan menggunakan metode *Feature Extraction* dengan *Fast Text* dan pemodelan *Naive Bayes* dengan optimalisasi metode *Feature Selection* dengan *Chi Square*.
   e. Kelima, pengujian data dengan menggunakan metode *Feature Extraction* dengan *Fast Text* dan pemodelan *Support Vector Machine* dengan optimalisasi metode *Feature Selection* dengan *Chi Square*.
6. Tahap Evaluasi (*Evaluation*)
   Pada tahapan ini aktivitas yang dilakukan dalam penelitian adalah melihat *performance model* dengan menggunakan evaluasi model dalam bentuk *Confusion Matrix*.
7. Tahap Implementasi
   Pada tahapan ini dilakukan implementasi Prototipe Aplikasi Analisa Data Sentimen untuk menampilkan hasil visualisasi pengujian yaitu proses upload dataset, proses *data preprocessing*, pengujian model yaitu *Naive Bayes*, *Support Vector Machine*, NB dengan SVM, NB dengan *Fast Text*, NB dengan *Chi Square*, SVM dengan *Fast Text* dan SVM dengan *Chi Square*
8. Tahap Implikasi Penelitian
   Pada tahap ini peneliti mencoba memberikan evaluasi hasil penelitian terkait implikasinya pada aspek aspek sistem, aspek manajerial serta aspek penelitian lanjut.

Adapun keterbatasan penelitian dalam beberapa hal, yaitu :

1. Jumlah Responden : Data hanya diperoleh dari jumlah responden terbatas, sehingga generalisasi hasil mungkin kurang representatif.
2. Bias Responden : Ada kemungkinan bias dalam jawaban kuesioner yang dapat memengaruhi hasil analisis sentimen.
3. Keterbatasan Model : Metode yang digunakan (*Naive Bayes* dan SVM) memiliki keterbatasan dalam menangani data teks yang sangat kompleks dibandingkan dengan metode berbasis deep learning.

3. HASIL DAN PEMBAHASAN

3.1. *Preprocessing Data*

Berikut ini adalah beberapa tahapan dalam *preprocessing data* untuk *sentiment analysis* :

1. *Data Cleansing* : berfungsi untuk membersihkan kata dari karakter-karakter seperti tanda baca dan karakter khusus ("!\"#$%&()*+-./:;<=>?@[\]^_`{|}~\n,). Tujuan dari proses *data cleansing* ini adalah mengurangi noise dengan dilakukan pembersihan dari karakter-karakter yang tidak diperlukan.
2. *Tokenization* : Proses memecah teks menjadi unit-unit lebih kecil yang disebut token. Token bisa berupa kata, frasa, atau entitas lainnya. Contoh : "Saya suka film ini" dapat dipecah menjadi token "Saya", "suka", "film", "ini".
3. *Casefolding* : Mengubah semua huruf dalam teks menjadi huruf kecil. Ini membantu memastikan konsistensi dalam analisis karena "Saya" dan "saya" dianggap sama. Contoh: "Saya suka film ini" diubah menjadi "saya suka film ini". Tahapan *Case Folding* adalah proses penyeragaman bentuk huruf dalam dokumen, huruf kapital akan dirubah menjadi huruf kecil dan diseregamkan dari A-Z, selain huruf akan dihilangkan karena dianggap delimeter.
4. Normalisasi Kalimat : Tujuan normalisasi kalimat adalah untuk mengubah kalimat gaul (*slang word*) menjadi kalimat normal, termasuk menghilangkan kata-kata yang disingkat.
5. *Filtering* : mengambil kata-kata penting dari hasil tokenisasi, proses ini dapat dilakukan dengan algoritma stoplist. Adapun mekanisme filtering dengan menggunakan *stopword removal* adalah proses menghapus kata-kata umum yang tidak memberikan kontribusi besar terhadap pemahaman atau analisis teks. Dalam penelitian ini, fungsi *stop word removal* adalah untuk mengurangi noise dalam data serta mempengaruhi tingkat akurasi model.

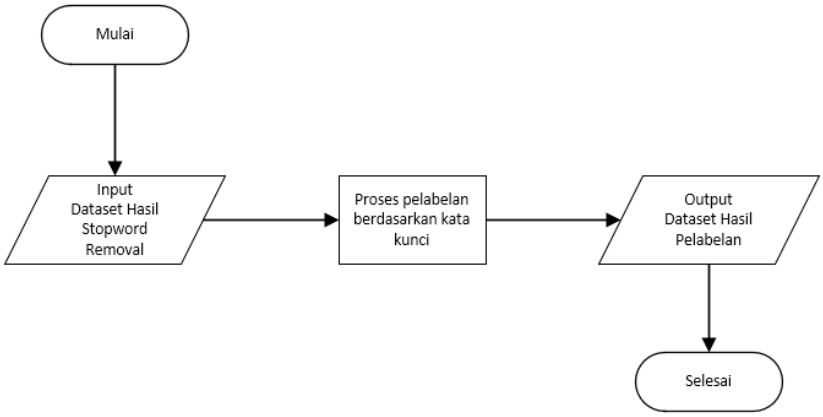
Adapun hasil proses *stopword removal* sesuai dengan mekanisme proses pada Gambar 2 didapatkan hasilnya pada Tabel 1. sebagai berikut :

Tabel 1. Hasil *Filtering* Dengan *Stopword Removal* Pada Data Sentimen

Text Setelah Tokenisasi	Text Hasil Proses *Stopword Removal*
['fasilitas', 'dan', 'sarana', 'dari', 'universitas', 'mercu', 'buana', 'sudah', 'bagus', 'namun', 'yang', 'saya', 'ingat', 'di', 'tahun', 'saya', 'belajar', '2015', '2019', 'dosen', 'yang', 'mengajari', 'saya', 'kebanyakan', 'kurang', 'mendalam', 'pengetahuannya', 'entang', 'apa', 'yang', 'sedang', 'di', 'hadapi', 'oleh', 'dunia', 'industri', 'saat', 'ini', 'sehingga', 'yang', 'diajarkan', 'hanya', 'modul', 'lama', 'yang', 'kurang', 'relevan', 'dengan', 'industri', 'saat', 'ini', 'sehingga', 'ketika', 'lulus', 'pun', 'mereka', 'tidak', 'punya', 'kapasitas', 'yang', 'cukup', 'untuk', 'dapat', 'diterima', 'oleh', 'perusahaan', 'mereka', 'mesti', 'belajar', 'lagi', 'di', 'tempat', 'pelatihan', 'terpadu', 'dan', 'lain-lain', 'jadi', 'kuliah', 'hanya', 'mengejar', 's1', 'untuk', 'mengejar', 'kompetensi', 'mereka', 'tidak', 'dapatkan', 'di', 'kuliah', 'tetapi', 'di', 'pelatihan', 'terpadu']	fasilitas sarana universitas mercu buana bagus belajar 2015 2019 dosen mengajari kebanyakan mendalam pengetahuannya entang hadapi dunia industri diajarkan modul relevan industri lulus kapasitas diterima perusahaan mesti belajar pelatihan terpadu lain-lain kuliah mengejar s1 mengejar kompetensi dapatkan kuliah pelatihan terpadu

3.2. *Data Labelling*

Adapun mekanisme pelabelan dengan kata kunci ditampilkan pada *flowchart* berikut ini :



Gambar 2. *Flowchart* Proses Pelabelan Data Berdasarkan Kata Kunci

Adapun hasil proses pelabelan dengan kata kunci adalah sebagai berikut :

Tabel 2. Hasil Pelabelan Dengan Kata Kunci

Text Hasil *Stopword Removal* Yang Dilabeli	Label
fasilitas sarana universitas mercu buana bagus belajar 2015 2019 dosen mengajari kebanyakan mendalam pengetahuannya entang hadapi dunia industri diajarkan modul **relevan** industri lulus kapasitas diterima perusahaan mesti belajar pelatihan terpadu lain-lain kuliah mengejar s1 mengejar kompetensi dapatkan kuliah pelatihan terpadu	positif
metode pembelajaran kebanyakan mata kuliah fakultas ilmu komputer metode bersaing kebutuhan industri materi dasar berumur dipertahankan mahasiswa **memahami situasi** terkini teknologi kebutuhan industri	positif
tugas diberika lumayan **menguras** energi tugas 1 tugas 2 pengerjaan nya 2 minggu mahasiswa jarang selesai	negatif
standar metode pembelajaran diajarkan	netral
pengajaran matakuliah pemograman berorientasi objek pengalaman mahasiswa masyarakat industri solusi aplikasi bahasa java **relevan** praktikum contoh kasus-kasus tugas industri masyarakat solusi ujian tugas dirumah dipresentasikan	positif
tingkatkan praktikum mata kuliah nya	netral

3.3. *Data Modelling*

Pada *data modelling* sesuai dengan hasil pengujian dari beberapa data modelling yang sudah dilakukan dengan menerapkan algoritma klasifikasi *Naive Bayes* dan *Support Vector Machine* serta optimalisasi algoritma menggunakan *feature extraction* menggunakan *fast text* dan *feature selection* menggunakan *chi square* didapatkan capaian akurasi tertinggi sebagai berikut :

Tabel 3. Hasil Akurasi Tertinggi Dari Pengujian *Data Modelling*

No	*Data Modelling* Dengan Pengujian Algoritma	Dataset Dengan Metode Pelabelan	Optimalisasi Algoritma	Kombinasi *Data Training* : *Data Testing*	Capaian Akurasi
1	*Naive Bayes*	Kata Kunci	-	20 : 80	67.68%
		Lexicon	-	20 : 80	63.94%
2	SVM	Kata Kunci	-	90 : 10	90.72%
		Lexicon	-	80 : 20	89.43%
3	*Naive Bayes* + SVM	Kata Kunci	-	80 : 20	90.72%
		Lexicon	-	80 : 20	89.95%
4	*Naive Bayes*	Kata Kunci	*Fast Text*	30 : 70	90.49%
		Lexicon	*Fast Text*	20 : 80	85.35%
5	*Naive Bayes*	Kata Kunci	*Chi Square*	70 : 30	71.65%
		Lexicon	*Chi Square*	70 : 30	73.20%
6	SVM	Kata Kunci	*Fast Text*	10 : 90	98.82%
		Lexicon	*Fast Text*	80 : 20	98.77%
7	SVM	Kata Kunci	*Chi Square*	90 : 10	99.58%
		Lexicon	*Chi Square*	90 : 10	99.37%

3.4. Evaluasi Hasil

Berdasarkan hasil uji coba perbandingan komposisi antara *data training* dan *data testing*, apabila tidak terfokus kepada jenis pelabelan dataset yang digunakan dapat diketahui keseluruhan hasil pengujian untuk capaian akurasi terbaik adalah :

Tabel 4. Keseluruhan Capaian Hasil Pengujian *Data Modelling*

Kombinasi *Data Training* : *Data Testing*	*NB*	*SVM*	*NB* + SVM	NB + *Fast Text*	*NB* + *Chi Square*	*SVM* + *Fast Text*	*SVM* + *Chi Square*
Dataset	Kata Kunci	Kata Kunci	Kata Kunci	Kata Kunci	*Lexicon*	Kata Kunci	Kata Kunci
90 : 10	64.95%	90.72%	90.21%	80.93%	71.65%	98.32%	99.58%
80 : 20	65.21%	90.46%	90.72%	82.22%	71.39%	98.43%	99.31%
70 : 30	64.26%	88.14%	86.60%	78.69%	73.20%	98.56%	99.05%
60 : 40	64.77%	87.23%	86.06%	83.35%	72.39%	98.63%	99.08%
50 : 50	64.91%	85.24%	83.90%	80.80%	70.49%	98.63%	98.92%
40 : 60	64.83%	83.32%	82.12%	86.84%	70.85%	98.70%	98.61%
30 : 70	65.56%	80.53%	80.53%	90.49%	70.35%	98.72%	97.88%
20 : 80	67.68%	75.81%	77.55%	90.06%	67.55%	98.74%	96.36%
10 : 90	62:21%	69:21%	69.50%	89.74%	60.61%	98.82%	92.65%

Berdasarkan informasi data pada Tabel 4. dapat di analisis beberapa hal yaitu :

1. Capaian nilai akurasi terbaik untuk pengolahan data analisis data sentimen menggunakan Algoritma Support Vector Machine yang dioptimalkan dengan penggunaan *Feature Selection Chi Square* dengan capaian nilai akurasi sebesar 99.58%.
2. Capaian nilai akurasi dengan pengujian model klasifikasi pada analisis data sentimen menggunakan Algoritma Naive Bayes tidak begitu mencapai hasil akurasi yang baik hanya sebatas nilai akurasi tertinggi sebesar 67.68%. Akan tetapi berdasarkan penelitian dapat dibuktikan bahwa penggunaan Algoritma Naive Bayes dapat ditingkatkan capaian akurasinya menggunakan optimalisasi algoritma feature extraction menggunakan fast text. Dengan kombinasi data training dan data testing yang sama 20:80, capaian nilai akurasi pengujian data modelling menggunakan *Algoritma Naive Bayes* dapat meningkat dari 67.68% ke 90.49%, apabila dioptimalkan dengan *feature extraction fast text*. Akan tetapi penggabungan pengujian model *Algoritma Naive Bayes* dengan *fast text* mencapai akurasi terbaik

sebesar 90.49% dengan komposisi antara *data training* dengan *data testing* sebesar 30:70 dengan menggunakan dataset dengan pelabelan kata kunci.

3. *Algoritma Support Vector Machine* dianggap memiliki kinerja lebih baik daripada *Naive Bayes* dalam pengujian *data modelling*, dimana dengan komposisi data training dengan data testing yang sama sebesar 90:10 maka hasil capaian akurasi maksimal menggunakan *Algoritma Support Vector Machine* adalah sebesar 90.72% dengan menggunakan dataset pelabelan kata kunci.
4. Penggabungan model klasifikasi data analisis sentimen antara *Algoritma Naive Bayes* dengan Algoritma *Support Vector Machine*, berdasarkan hasil penelitian memiliki capaian kinerja akurasi data lebih baik dengan nilai maksimal adalah sebesar 90.72% dengan kombinasi antara data training dengan data testing adalah sebesar 80:20 dengan menggunakan dataset pelabelan kata kunci.
5. Optimalisasi *Algoritma Naive Bayes* dengan *Feature Selection Chi Square* juga dapat meningkatkan kinerja data modelling dengan capaian akurasi maksimal di 73.20% meningkat lebih baik daripada hanya menggunakan Algoritma *Naive Bayes* saja sebesar 67.68% dengan kombinasi antara data training dan data testing yaitu 70:30 dengan menggunakan dataset pelabelan lexicon.
6. Kinerja Algoritma *Support Vector Machine* meningkat lebih baik dalam pengujian data analisis sentimen, apabila dioptimalkan dengan menggunakan *Feature Extraction Fast Text* mencapai maksimal sebesar 98.82% dengan kombinasi antara data training dengan data testing 10:90.
7. Berdasarkan hasil penelitian maka data modelling yang menghasilkan capaian akurasi terbaik apabila menerapkan :
   a. Pengujian Algoritma *Support Vector Machine* yang dioptimalisasi menggunakan *Feature Selection Chi Square* dengan kombinasi *data training* dan *data testing* adalah 90:10 menghasilkan nilai akurasi sebesar 99.58% menggunakan dataset pelabelan kata kunci.
   b. Pengujian dengan menggabungkan Algoritma *Naive Bayes* dengan *Support Vector Machine* dengan kombinasi *data training* dan *data testing* adalah 80:20 mencapai nilai akurasi 90.72% menggunakan dataset dengan pelabelan kata kunci.
   c. Pengujian dengan menggabungkan Algoritma *Naive Bayes* dengan optimalisasi algoritma menggunakan *Feature Extraction Fast Text* dengan kombinasi *data training* dan *data testing* adalah 30:70 mencapai nilai akurasi 90.49% menggunakan dataset dengan pelabelan kata kunci.
   d. Pengujian dengan menggabungkan Algoritma *Support Vector Machine* dengan optimalisasi algoritma menggunakan *Feature Extraction Fast Text* dan *Feature Selection Chi Square* dapat meningkatkan kinerja model untuk mencapai nilai akurasi yang baik. Dimana capaian nilai akurasi maksimal penggabungan Algoritma *Support Vector Machine* dengan *Fast Text* adalah sebesar 98.82% dengan kombinasi *data training* dan *data testing* adalah 80:20. Serta capaian nilai akurasi maksimal penggabungan Algoritma *Support Vector Machine* dengan *Chi Square* adalah sebesar 99.58% dengan kombinasi *data training* dan *data testing* adalah 90:10.

Tabel 5. Hasil Pengujian Terbaik

Algoritma	Klasifikasi	Percobaan Perbandingan Data *Weighted Average*				
		90:10	**80:20**	**70:30**	**30:70**	**10:90**
SVM	*Precision*	0.91	0.91	0.88	0.81	0.70
	Recall	0.91	0.90	0.88	0.81	0.69
	F1-Score	0.91	0.91	0.88	0.80	0.67
NB + SVM	*Precision*	0.90	0.91	0.87	0.81	0.70
	Recall	0.90	0.91	0.87	0.81	0.68
	F1-Score	0.90	0.91	0.87	0.81	0.64
NB + Fast Text	*Precision*	0.84	0.84	0.81	0.92	0.92
	Recall	0.81	0.82	0.79	0.90	0.90
	F1-Score	0.80	0.81	0.76	0.88	0.86
SVM + Fast Text	*Precision*	**0.99**	**0.99**	**0.99**	0.99	0.99
	Recall	**0.99**	**0.99**	**0.99**	0.99	0.99
	F1-Score	**0.99**	**0.99**	**0.99**	0.99	0.99
SVM + Chi Square	*Precision*	**1.00**	**0.99**	**0.99**	0.98	0.93
	Recall	**1.00**	**0.99**	**0.99**	0.98	0.93
	F1-Score	**1.00**	**0.99**	**0.99**	0.98	0.93

Penggunaan Naive Bayes dan SVM dengan optimasi FastText dan Chi-Square secara langsung relevan dengan permasalahan yang dihadapi :

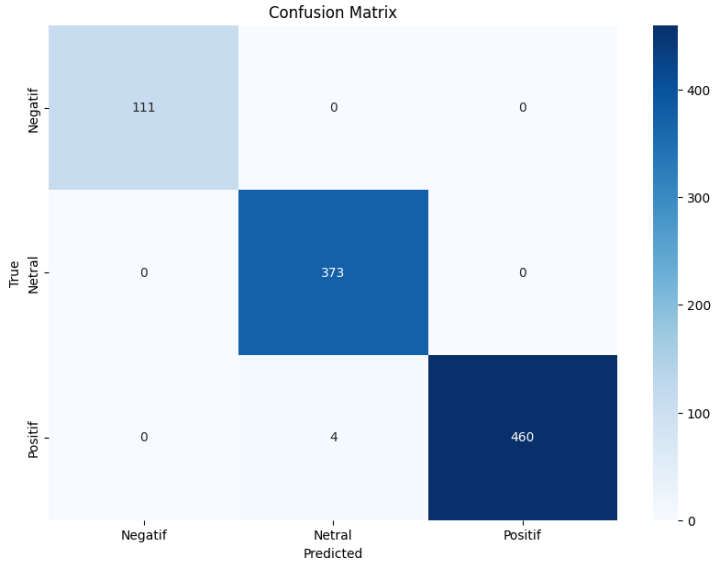
1. Kompetensi Pemrograman yang Rendah: Model optimasi menghasilkan analisis yang lebih akurat dalam mengidentifikasi opini positif, negatif, dan netral dari responden, sehingga membantu dalam mengevaluasi kelemahan pembelajaran pemrograman.
2. Kebutuhan Data yang Relevan: FastText memungkinkan representasi data teks yang lebih detail dengan mempertimbangkan sub-kata, sementara Chi-Square membantu menyaring fitur yang paling relevan untuk klasifikasi

Hasil penelitian juga memberikan wawasan tentang persepsi responden terhadap pembelajaran pemrograman :

1. Opini Positif: Sebagian besar responden mengapresiasi fasilitas dan infrastruktur kampus, namun hal ini belum cukup untuk menutupi kelemahan pada kurikulum dan metode pengajaran.
2. Opini Negatif: Responden menyoroti kurangnya relevansi materi dengan kebutuhan industri, yang menjadi sumber utama kesenjangan kompetensi.
3. Opini Netral: Menunjukkan perlunya perbaikan dalam metode pengajaran praktis agar lebih relevan dengan tantangan industri.

3.5. Prototipe Visualisasi Hasil

Dalam penelitian ini peneliti menggunakan *Software Python* untuk analisis sentimen karena memiliki beragam pustaka dan alat yang mendukung pengolahan NLP (*Natural Language Processing*) dan analisis teks yang dijalankan pada *platform Google Colabs*. Berikut ini adalah tampilan *Confusin Matrix* untuk capaian akurasi terbaik dengan penggabungan algoritma *Support Vector Machine* dengan *Chi Square* dengan komposisi data training dan data testing 90 : 10 mencapai nilai akurasi model SVM sebesar : 99.58% yaitu :



Gambar 3. Tampilan *Confusion Matrix* Model SVM + *Chi Square Keyword*

Dari *confusion matriks* pada Gambar 3, peneliti dapat menyimpulkan beberapa hal yaitu :

- **Tingkat Akurasi** : Model menunjukkan tingkat akurasi yang tinggi dengan sebagian besar sampel diklasifikasikan dengan benar.
- **Kesalahan Klasifikasi** : Ada sedikit kesalahan klasifikasi, terutama pada kelas Positif yang diprediksi sebagai Netral (4 sampel).
- **Keseimbangan Data** : Tampaknya tidak ada kesalahan prediksi untuk sampel Negatif dan Netral yang diprediksi sebagai Positif, menunjukkan model dapat membedakan dengan baik antara kelas Negatif, Netral, dan Positif.
- **Performa Model** : Model ini tampaknya sangat efektif dalam mengklasifikasikan kelas Negatif dan Positif, dengan kesalahan yang minimal pada kelas Netral yang diprediksi sebagai Positif.

Berikut ini hasil visualisasi data hasil pengujian model data analisis sentimen dengan menggunakan *world cloud*, yaitu sebagai berikut :

Gambar 4. Hasil Proses Visualisasi Data Dengan *World Cloud*

Penelitian ini memberikan rekomendasi praktis untuk mengatasi permasalahan yang telah diidentifikasi, yaitu :

1. **Revisi Kurikulum** : Fokus pada keterkaitan materi pembelajaran dengan tren industri terkini.
2. **Pelatihan Dosen** : Meningkatkan kompetensi pengajar melalui pelatihan berbasis industri.
3. **Evaluasi Berbasis Data** : Menggunakan model analisis sentimen untuk memantau secara berkala kepuasan dan kebutuhan mahasiswa.

4. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma Support Vector Machine (SVM) yang dioptimalkan dengan metode Feature Selection Chi Square menghasilkan tingkat akurasi tertinggi sebesar 99.58%. Hal ini menunjukkan bahwa kombinasi algoritma tersebut memiliki performa yang sangat baik dalam mengklasifikasikan sentimen mahasiswa, dosen, dan alumni terhadap pembelajaran pemrograman di Fakultas Ilmu Komputer Universitas Mercu Buana. Selain itu, optimalisasi Naive Bayes dengan FastText juga memberikan peningkatan akurasi yang signifikan, mencapai 90.49% dibandingkan penggunaan Naive Bayes tanpa optimasi. Hasil ini menegaskan bahwa pendekatan kombinasi atau ensemble dalam analisis sentimen dapat menghasilkan prediksi yang lebih akurat, mendukung pengambilan keputusan berbasis data untuk meningkatkan kualitas pembelajaran.

Dampak signifikan dari penelitian ini adalah memberikan dasar yang kuat bagi fakultas dalam memahami kebutuhan dan persepsi stakeholder terkait pembelajaran pemrograman. Dengan akurasi yang tinggi, hasil analisis dapat dimanfaatkan untuk mengidentifikasi kelemahan dalam sistem pembelajaran dan merumuskan strategi perbaikan yang lebih tepat sasaran. Sebagai implikasi, penelitian ini tidak hanya membantu meningkatkan relevansi pembelajaran dengan kebutuhan industri, tetapi juga memberikan kontribusi pada pengembangan model analisis sentimen di sektor pendidikan. Penelitian lebih lanjut dapat mengeksplorasi penggunaan teknik deep learning atau mengintegrasikan model lain untuk meningkatkan akurasi dan efisiensi analisis data sentimen.

UCAPAN TERIMAKASIH

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

DAFTAR PUSTAKA

[1] L. Alekseeva, J. Azar, M. Giné, S. Samila, and B. Taska, “The demand for AI skills in the labor market,” *Labour Econ.*, vol. 71, 2021, doi: 10.1016/j.labeco.2021.102002.

[2] H. W. Volberda, S. Khanagha, C. Baden-Fuller, O. R. Mihalache, and J. Birkinshaw, “Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms,” *Long Range Plann.*, vol. 54, no. 5, p. 102110, 2021, doi: 10.1016/j.lrp.2021.102110.

[3] E. M. Ikart, “Survey Questionnaire Survey Pretesting Method: An Evaluation of Survey Questionnaire via Expert Reviews Technique,” *Asian J. Soc. Sci. Stud.*, vol. 4, no. 2, p. 1, 2019, doi: 10.20849/ajsss.v4i2.565.

[4] B. T. Pham *et al.*, “A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling,” *Geocarto Int.*, vol. 34, no. 13, pp. 1385–1407, 2019, doi: 10.1080/10106049.2018.1489422.

[5] T. M. Ma, K. Yamamori, and A. Thida, “A Comparative Approach to Naïve Bayes Classifier and Support Vector Machine for Email Spam Classification,” *2020 IEEE 9th Glob. Conf. Consum. Electron. GCCE 2020*, pp. 324–

326, 2020, doi: 10.1109/GCCE50665.2020.9291921.
[6] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes-Support Vector Machine (NBSVM) Classifier," *Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019*, vol. 1, pp. 199–205, 2019, doi: 10.1109/ICOMITEE.2019.8920923.
[7] 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.
[8] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, no. xxxx, p. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.
[9] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.
[10] J. Gu and S. Lu, "An effective intrusion detection approach using SVM with naïve Bayes feature embedding," *Comput. Secur.*, vol. 103, p. 102158, 2021, doi: 10.1016/j.cose.2020.102158.
[11] T. W. B, S. Datko, and H. Maciejewski, *Feature Extraction in Subject*. Springer International Publishing, 2018. doi: 10.1007/978-3-319-91262-2.
[12] 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.
[13] I. S. Thaseen, C. A. Kumar, and A. Ahmad, "Integrated Intrusion Detection Model Using Chi-Square Feature Selection and Ensemble of Classifiers," *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3357–3368, 2019, doi: 10.1007/s13369-018-3507-5.
[14] 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.
[15] 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.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 2, no. January, pp. 40–46, 2021, doi: 10.1016/j.ijcce.2021.01.001.