

# Analisis Sentimen Evaluasi Reaksi *E-Learning* Menggunakan Algoritma Naïve Bayes Support Vector Machine dan Deep Learning

*Sentiment analysis of e-learning reaction evaluations using Naïve Bayes, Support Vector Machine, and Deep Learning algorithms*

Nurul Firdausy<sup>1</sup>, Imam Yuadi<sup>2</sup>, Ira Puspitasari<sup>3</sup>

<sup>1,2,3</sup>Program Studi Magister Pengembangan Sumber Daya Manusia, Sekolah Pascasarjana, Universitas Airlangga

<sup>2</sup>Departemen Ilmu Informasi dan Perpustakaan, Universitas Airlangga

<sup>3</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga

E-mail: <sup>1</sup>nurul.firdausy-2022@pasca.unair.ac.id, <sup>2</sup>imam.yuadi@fisip.unair.ac.id, <sup>3</sup>irap@fst.unair.ac.id

## Abstrak

Evaluasi reaksi atau evaluasi kepuasan merupakan bentuk evaluasi paling umum digunakan dalam pelatihan karena kemudahan dan sifatnya yang lekat dengan pelatihan. Meskipun mengandung wawasan yang dapat bernilai dalam peningkatan kualitas penyelenggaraan pelatihan, namun penelitian terkait reaksi peserta masih sangat terbatas. Penelitian ini bertujuan melakukan analisis sentimen terhadap evaluasi reaksi peserta e-learning menggunakan algoritma Naïve Bayes, Support Vector Machine dan Deep Learning. Reaksi peserta berupa komentar diklasifikasikan ke dalam kategori apresiasi, saran dan kritik. Hasil penelitian menunjukkan model Naïve Bayes memiliki kinerja yang lebih baik dibandingkan SVM dan Deep Learning dalam prediksi sentimen komentar peserta dengan tingkat akurasi, presisi dan recall masing-masing sebesar 82,54%, 68,08% dan 69,81%. Prediksi sentimen reaksi peserta menggunakan model Naïve Bayes diperoleh hasil 70% berupa apresiasi, 16% berupa saran dan 14% merupakan kritik. Penelitian ini memberikan kontribusi praktis analisis evaluasi reaksi pelatihan dan menambah literatur implementasi *text mining* pada domain *human resource analytics*.

Kata kunci: Analisis Sentimen, Naïve Bayes, Support Vector Machine, Deep Learning, Evaluasi Pelatihan

## Abstract

*Reaction evaluation or satisfaction evaluation is the most commonly used form of evaluation in training because of its ease and inherent nature in training. Although it contains insights that can be valuable in improving the quality of training provision, research related to participant reactions remains very limited. This study aims to conduct a sentiment analysis of participant reaction evaluations in e-learning using the Naïve Bayes, Support Vector Machine, and Deep Learning algorithms. Participants' reactions in the form of comments are classified into the categories of appreciation, suggestions, and criticism. The results show that the Naïve Bayes model performs better than SVM and Deep Learning in predicting participant comment sentiments, with accuracy, precision, and recall rates of 82.54%, 68.08%, and 69.81%, respectively. The prediction of participant reaction sentiments using the Naïve Bayes model yielded results of 70% in the form of appreciation, 16% in the form of suggestions, and 14% in the form of criticism. This study makes a practical contribution to the analysis of training reaction evaluations and adds to the literature on text mining implementation in the domain of human resource analytics.*

Keywords: Sentiment Analysis, Naïve Bayes, Support Vector Machine, Deep Learning, Training Evaluation

## 1. PENDAHULUAN

Pelatihan merupakan salah satu predictor keberhasilan organisasi [1], yang dibuktikan oleh hasil penelitian yang mengukur dampak pelatihan terhadap peningkatan kinerja karyawan [2], kemampuan melayani [3], dan *employee engagement* [4] baik pada domain perusahaan swasta maupun organisasi public. Tidak hanya terhadap individu dan organisasi, literatur review yang dilakukan [5] mengungkapkan pelatihan juga memberikan manfaat kepada masyarakat. Model evaluasi pelatihan yang banyak diadopsi praktisi dalam mengukur keberhasilan suatu pelatihan mencapai tujuan yang diharapkan perusahaan yaitu model evaluasi training 4 (empat) level Kirkpatrick [6], [7]. Model ini melakukan evaluasi terhadap pelatihan melalui 4 level evaluasi yang dilakukan secara bertahap, meliputi evaluasi reaksi peserta (*reactions*), pengetahuan dan/atau keterampilan (*learning*), transfer pengetahuan dan keterampilan ke tempat kerja (*behaviour*), dan evaluasi kinerja individu [8] dan organisasi (*result*) seperti penghematan biaya, peningkatan kualitas, atau peningkatan kuantitas.

Reaksi peserta adalah penilaian evaluatif oleh peserta terhadap fitur-fitur pelatihan, termasuk konten (apa yang disampaikan) dan bagaimana cara penyampaian (*delivery*) [9]. Bentuk evaluasi ini menjadi ukuran yang paling umum digunakan perusahaan untuk mengevaluasi program pelatihan [8], [10]–[12], bahkan sering menjadi satu-satunya ukuran evaluasi pelatihan [13] karena kemudahan dalam pengumpulan [9] dan sifatnya yang dekat dengan pelatihan itu sendiri [10]. [14] menyatakan bahwa reaksi peserta dalam pelatihan memiliki nilai tambahan seperti pengambilan keputusan, umpan balik, dan pemasaran. Data reaksi peserta dapat memberikan nilai yang berguna bagi organisasi di berbagai hal [9]. Sayangnya walau sering digunakan sebagai bentuk evaluasi, namun penelitian terkait dengan reaksi peserta masih terbatas [8], [9]. Salah satu penyebabnya karena reaksi kualitatif peserta kurang dapat dipahami dengan baik sehingga menimbulkan kebingungan dalam menafsirkannya [9], dan memunculkan pertanyaan mengenai nilai serta penggunaannya yang optimal dalam pengambilan keputusan [9], [12]. Oleh karena itu eksplorasi terhadap reaksi kualitatif pelatihan menjadi area yang penting untuk menjadi perhatian peneliti [11].

Survei evaluasi pelatihan pada umumnya meminta peserta untuk memberikan peringkat kuantitatif atas item reaksi dan respon terhadap pertanyaan terbuka [11], [15] (komentar terbuka atau *open-ended question*). Kemudahan memberikan komentar terutama pada evaluasi pelatihan-pelatihan online semakin menarik perhatian terhadap reaksi kualitatif suatu pelatihan [11]. Penelitian yang melibatkan respon kualitatif peserta pelatihan diperoleh dari komentar terbuka [15] maupun wawancara [16], [17] saat ini masih menggunakan analisis konten sebagai metode analisisnya. Teknik konvensional ini memiliki banyak kelemahan terkait ketelitian dan prosedur pengkodean yang rumit pada data teks dalam jumlah besar [18]. Adapun di sisi lain prosedur analisis secara manual mengakibatkan ketidakkonsistenan dalam penafsirannya [19].

Perkembangan teknologi tidak hanya memberikan kemudahan dalam kehidupan manusia sehari-hari, namun juga membawa dampak positif pada bidang sains, termasuk pada analisis data teks tidak terstruktur yang dikenal dengan *text mining*. *Text mining* adalah proses menganalisis informasi teks untuk menemukan struktur dan makna yang "tersembunyi" dalam teks [20]. Penelitian mengenai umpan balik pelanggan dan *text mining* telah difokuskan terutama pada ekstraksi informasi yang terkait dengan opini atau sentimen [19], [21]. Dalam beberapa tahun terakhir, penggunaan *text mining* mengalami peningkatan yang signifikan, terutama ketika digabungkan dengan analisis *big data* [21]. Agar diperoleh wawasan terkait umpan balik pelanggan, maka kalimat umpan balik perlu diklasifikasikan ke dalam kelas-kelas semantic yang berguna seperti apresiasi, saran dan keluhan [22], ini berguna untuk mengidentifikasi daftar keinginan pelanggan, fitur produk yang bernilai dan kekurangan produk atau layanan. Salah satu pendekatan klasifikasi dalam *text mining* yaitu menggunakan *machine learning* [19].

Berdasarkan studi literatur yang dilakukan [23], secara umum teknik *machine learning* yang banyak digunakan pada penelitian analisis sentimen atau *opinion mining* adalah Naïve Bayes (NB), support vector machine (SVM), dan neural network (NN) karena memiliki performa yang baik. Di samping itu, algoritma Deep Learning (DL) sebagai variasi bentuk neural network

dengan lapisan yang lebih banyak turut mengalami peningkatan popularitas penggunaan pada penelitian analisis sentimen [23], [24]. Penelitian yang dilakukan [25] menggunakan algoritma NB, SVM dan Random Forest (RF) untuk klasifikasi sentimen review destinasi wisata menyimpulkan algoritma SVM memiliki tingkat akurasi lebih baik yaitu 67,97% dibandingkan algoritma NB (61,33%) dan RF (63,55%). Analisis sentimen lain yang dilakukan [26] terhadap opini publik mengenai vaksin Covid-19 di Indonesia diperoleh hasil serupa di mana tingkat akurasi tertinggi diperoleh dari algoritma SVM sebesar 85% sedangkan akurasi algoritma RF lebih kecil yaitu 81%. Hasil berbeda diperoleh [27] yang membandingkan kinerja algoritma SVM dan NB pada analisis sentimen e-sport untuk kurikulum edukasi di mana algoritma NB mampu menghasilkan akurasi lebih tinggi dibandingkan SVM yaitu sebesar 70,32%. Adapun penggunaan algoritma Deep Learning (DL) tipe Convolutional Neural Network (CNN) pada sentimen analisis [28] mampu menghasilkan tingkat akurasi di atas 60%, dan pada penelitian lain pemanfaatan algoritma kombinasi CNN dengan Long-Short Term Memory (LSTM) atau CLSTM menghasilkan tingkat akurasi sebesar 85,8% untuk klasifikasi komentar ke dalam 4 kelas sentimen [29].

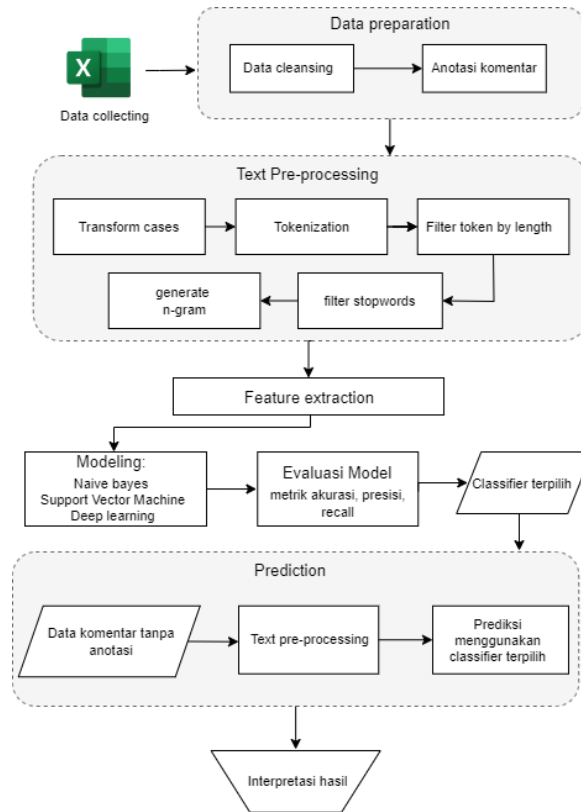
Penelitian ini bertujuan melakukan studi terkait reaksi kualitatif peserta e-learning dalam bentuk komentar terbuka menggunakan pendekatan *machine learning* dengan menerapkan algoritma NB, SVM dan DL untuk klasifikasi komentar menggunakan aplikasi RapidMiner. Data yang digunakan adalah hasil evaluasi *e-learning communication skill* tingkat dasar yang diselenggarakan Badan Pendidikan dan Pelatihan Keuangan (BPPK) pada tahun 2022. Mengacu pada [30] komentar peserta terhadap penyelenggaraan *e-learning* akan diklasifikasikan ke dalam 3 kelas yaitu apresiasi, saran dan kritik. Hasil penelitian ini dapat memberikan wawasan mengenai algoritma mana yang memiliki performa lebih baik dalam prediksi komentar peserta, serta memberikan gambaran bagaimana pendapat peserta secara keseluruhan terhadap penyelenggaraan e-learning. Penelitian ini diharapkan dapat memberikan kontribusi terhadap prosedur analisis evaluasi dan manajemen kualitas penyelenggaraan pelatihan serta menambah literatur implementasi teks mining pada bidang pengembangan sumber daya manusia atau *human resource analytics*.

## 2. METODE PENELITIAN

Tahapan proses pada penelitian ini merupakan adaptasi dari model Cross Industry Standard Process for Data Mining (CRISP-DM), yang merupakan *standard de facto* untuk pengembangan data mining dan proyek penemuan pengetahuan dari data berdasarkan banyak survey pengguna dan jejak pendapat [31] karena dapat mengurangi biaya dan waktu proses serta meminimalisir pengetahuan yang diperlukan pada proses data mining [32]. Alur proses penelitian sebagaimana disajikan pada Gambar 1 terdiri dari *data collecting* (pengumpulan data), *data preparation* (persiapan data), *feature extraction* (ekstraksi fitur), *modeling* (pemodelan), *evaluation* (evaluasi) dan *prediction* (prediksi).

### 2.1 Data collecting

Data komentar evaluasi *e-learning communication skill* tingkat dasar diperoleh dari aplikasi kediklatan BPPK yang diunduh secara bertahap mulai tanggal 1 April sampai dengan 2 Mei tahun 2023. Komentar peserta merupakan penilaian kualitatif e-learning selain penilaian kuantitatif persepsi pada 5 (lima) aspek *e-learning* yaitu materi pembelajaran, bahan ajar, kesesuaian metode pembelajaran, kesiapan panitia, menu dan fasilitas *e-learning* menggunakan skala likert 1 s.d 5 (tidak baik s.d. sangat baik).



Gambar 1. Alur proses penelitian

## 2.2 Data Preparation

Komentar teks merupakan data tidak terstruktur. Data preparation adalah tahap penting untuk mengubah teks menjadi data terstruktur yang dapat diproses oleh komputer. *Data preparation* melibatkan integrasi data, pembersihan data, pemberian anotasi atau label komentar pada data training dan *text pre-processing*. Integrasi data dilakukan dengan menggabungkan beberapa data *flat file* secara manual menjadi satu file kompilasi. Pembersihan data dilakukan dengan membersihkan data dari penggunaan simbol, komentar kosong dan komentar yang tidak relevan.

Pemberian anotasi/label dilakukan pada data yang akan digunakan sebagai data training untuk membangun model. Proses ini dilakukan oleh peneliti terhadap setiap komentar peserta dengan melabeli komentar ke dalam kelas sentimen apresiasi, saran atau kritik. Sebuah komentar berpeluang berisi lebih dari satu kelas sentimen. Mengacu pada tujuan evaluasi pelatihan adalah perbaikan kualitas, maka proses anotasi komentar menerapkan prioritas kategori sentimen dari tertinggi ke terendah yaitu kritik, saran, dan apresiasi. Kategori tertinggi dipilih jika dalam suatu komentar mengandung lebih dari satu kategori. Contoh pemberian anotasi disajikan Tabel 1.

Tabel 1. Contoh pemberian anotasi (label) pada komentar

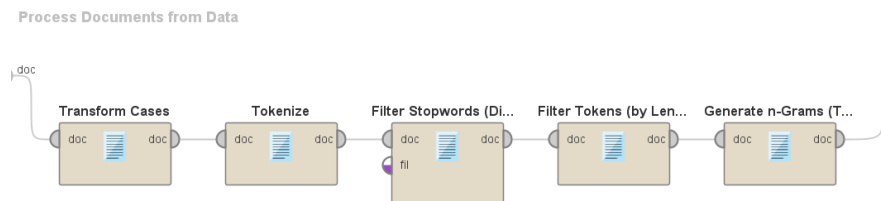
No	Komentar	Anotasi /label
1	Terima kasih.	apresiasi
2	Untuk di adakan tatap muka agar lebih memahami kembali	saran
3	Baik. Namun untuk Video dimana harus full attention, mengikuti Pointer step per step agak menyulitkan, smeoga bisa dikurangi.	saran
4	ada bahan yang tidak sinkron antara materi dengan ringkasan misal pada informasi jabatan. Terima kasih	kritik
5	sangat menarik, agar saat penjelasan materi volume backsoundnya bisa lebih dkecilkan lagi	saran

Text preprocessing melibatkan sejumlah langkah untuk membersihkan dan mengubah teks mentah menjadi bentuk yang lebih terstruktur dan siap untuk diproses lebih lanjut, proses ini

terdiri dari *transform cases*, *tokenization*, *filter token by length*, *filter stopwords* dan *n-grams* (Gambar 2).

- *transform cases*: penyeragaman bentuk huruf pada data komentar dengan mengubah huruf menjadi huruf kecil semua, atau besar semua.
- *tokenization*: memecah teks menjadi unit-unit yang lebih kecil, seperti kata-kata atau token. Ini adalah langkah awal untuk memahami struktur teks.
- *filter token by length*: menyaring hasil token yang dihasilkan dengan panjang minimal.
- *filter stopwords*: menghapus kata-kata umum yang sering muncul dan kurang memberikan makna pada analisis teks seperti "dan," "atau," "juga," dan sebagainya.
- *n-grams*: membantu dalam mempertahankan informasi urutan kata dalam teks dan dapat digunakan sebagai fitur dalam analisis teks untuk mengidentifikasi frasa atau pola tertentu.

Output dari rangkaian proses *data preparation* berupa representasi teks yang lebih terstruktur dan siap untuk diubah menjadi fitur numerik pada tahap selanjutnya.



Gambar 2. Operator text-preprocessing di dalam operator Process Document from Data pada RapidMiner

### 2.3 Feature extraction

Pada tahap ini teks yang telah dilakukan pre-proses dirubah menjadi representasi vector numerik yang berfungsi sebagai input pada model algoritma dalam analisis sentimen. Penelitian ini menggunakan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF). Vektor TF-IDF adalah representasi vektor dari teks yang menunjukkan seberapa penting suatu kata dalam sebuah dokumen dalam koleksi dokumen. *Term Frequency* (TF) menunjukkan berapa kali sebuah kata muncul dalam dokumen tersebut. Semakin sering kata tersebut muncul, semakin tinggi bobotnya dalam vektor TF. *Inverse Document Frequency* (IDF) mengukur seberapa penting sebuah kata dalam konteks koleksi dokumen secara keseluruhan. Kata-kata yang muncul di banyak dokumen mendapatkan nilai IDF yang lebih rendah, karena kemunculannya di banyak dokumen menunjukkan bahwa kata tersebut kurang spesifik dan kurang memberikan informasi unik. Penggabungan kedua komponen ini menghasilkan vektor TF-IDF sebagai representasi numerik bobot kata. Bobot kata yang lebih besar dianggap lebih informatif dan dipertahankan, adapun bobot yang lebih rendah dianggap kurang informatif dan dikecualikan [29]. Persamaan (1-3) digunakan untuk menghitung bobot vektor TF-IDF [27], [29].

$$TF_{(t,d)} = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{jumlah kata dalam dokumen } d} \quad (1)$$

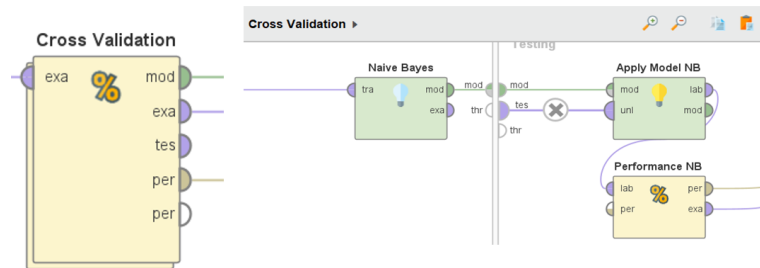
$$IDF_t = \log \frac{\text{jumlah dokumen}}{\text{jumlah dokumen yang mengandung kata } t} \quad (2)$$

$$\text{nilai TF - IDF} = \text{nilai TF} \times \text{nilai IDF} \quad (3)$$

### 2.4. Modeling

Pada tahap ini dilakukan pembentukan model prediksi sentimen menggunakan data training yang merupakan output dari tahap sebelumnya. Model prediksi sentiment kelak dapat mengklasifikasikan komentar peserta secara otomatis ke dalam kelas apresiasi, saran atau kritik. Model prediksi sentiment yang akan dibentuk menggunakan 3 (tiga) algoritma klasifikasi teks yaitu Naïve Bayes (NB), Support Vector Machine (SVM) dan Deep Learning (DL). Algoritma DL pada penelitian ini menggunakan operator *default* RapidMiner, algoritma deep learning H2O 3.30.0.1, yang dapat digunakan untuk membangun dan melatih jaringan syaraf dalam (*Deep Neural Network*). Proses *modeling* pada aplikasi RapidMiner dilakukan menggunakan operator

Cross Validation dengan memasukkan operator algoritma klasifikasi, operator Apply Model dan operator Performace ke dalamnya (Gambar 3). Proses ini sekaligus memberikan output hasil evaluasi terhadap model yang terbentuk.



Gambar 3. Operator pembentukan model di dalam operator Cross-Validation pada RapidMiner

### 2.5. Evaluation

Tahap ini melibatkan evaluasi kinerja model atau algoritma yang telah dikembangkan dalam melakukan klasifikasi sentimen. Metode pengujian pada umumnya menggunakan metrik akurasi, presisi, dan *recall* [26]. Akurasi mengevaluasi sejauh mana prediksi yang dihasilkan model sesuai dengan nilai aktualnya. Presisi mengukur seberapa banyak data positif yang diklasifikasikan dengan benar dibandingkan dengan jumlah data positif yang diprediksi. Sedangkan *recall* mengukur seberapa banyak data positif yang diklasifikasikan dengan benar dibandingkan dengan total jumlah data positif yang ada. Untuk mendapatkan nilai kinerja tersebut perlu dibangun *confusion matrix* yang umumnya berisi empat jenis nilai yaitu *true positif* (TP), *false positive* (FP), *true negative* (TN) dan *false negative* (FN). Persamaan (4-6) digunakan untuk menghitung kinerja model algoritma [26], [33].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

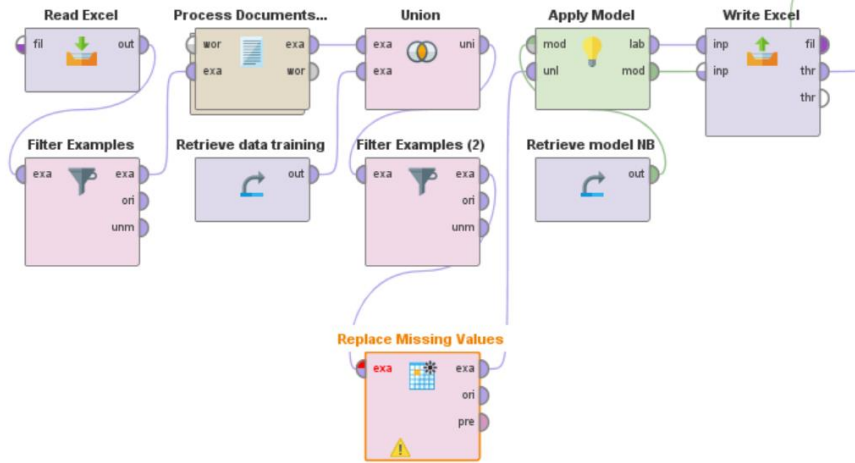
Evaluasi pada model menggunakan teknik *cross-validation* dengan 10 kali iterasi (*10-fold cross validation*). Teknik ini mengambil dataset yang tersedia dan membaginya menjadi 10 subset yang sama besar (*10-fold*), di mana salah satu subset menjadi data validasi dan 9 subset sisanya digunakan sebagai data training untuk melatih model. Proses tersebut diulangi 10 kali sehingga setiap subset menjadi data validasi secara bergantian. Akhirnya, 10 hasil evaluasi diambil rata-ratanya untuk menghasilkan nilai akhir performa model. Algoritma dengan metrik evaluasi terbaik akan dipilih menjadi *classifier* untuk melakukan prediksi terhadap data yang belum diberikan anotasi (data prediksi).

### 2.6 Prediction

Pada penelitian ini model klasifikasi terpilih digunakan untuk prediksi komentar reaksi peserta yang telah melalui proses *cleaning* namun belum memiliki anotasi/label. Hasil prediksi tersebut kemudian disimpulkan dan diinterpretasikan untuk memberikan wawasan atas pendapat peserta secara holistic mengenai penyelenggaraan *e-learning communication skill* tingkat dasar. Alur prediksi komentar peserta menggunakan *classifier* terpilih (Gambar 4), terdiri dari proses:

- Pemilihan data komentar yang akan diprediksi menggunakan operator Read Excel dan Filter Examples.
- Melakukan *Pre-processing text* menggunakan operator Process Document from Data dengan tahap pre-proses sebagaimana Gambar 2.
- Pencocokan dimensi vector data prediksi dengan dimensi vector data training menggunakan

- operator Union, Retrieve, Filter Examples dan Replace Missing values.
- Melakukan prediksi pada data prediksi menggunakan operator Apply Model dengan input *classifier* terpilih yang telah disimpan (operator Retrieve) dan kemudian mengunduh hasil prediksi menggunakan operator Write Excel.



Gambar 4. Proses prediksi komentar menggunakan *classifier* terpilih

### 3. HASIL DAN PEMBAHASAN

E-learning communication skill tingkat dasar yang diselenggarakan dalam 19 angkatan pada tahun 2022. Hasil pengumpulan data mentah diperoleh sebanyak 1.864 komentar. Proses pembersihan data mentah menghapus data komentar yang kosong dan tidak relevan, serta merubah kata “e-learning” menjadi “elearning”. Proses ini mereduksi data komentar menjadi 1.712 komentar. Data komentar yang telah dibersihkan kemudian dibagi menjadi dua yaitu data training sebanyak 859 komentar, dan data untuk prediksi 853 komentar. Proses anotasi/pelabelan dilakukan terhadap data training ke dalam kelas apresiasi, saran atau kritik. Hasil pelabelan pada data training menghasilkan 593 komentar apresiasi, 138 saran dan 122 kritik.

#### 1. Text Pre-processing

Pembahasan pada tahap ini menjelaskan proses menghasilkan dataset final yang siap dilakukan vektorisasi pada tahap ekstraksi fitur. Tahap pre-processing data teks komentar peserta dilakukan secara berurutan melalui proses *transform cases*, *tokenization*, *filter token by length*, *filter stopwords* dan *generate n-gram*. *Transform cases* merubah data teks komentar menjadi huruf kecil semua. *Tokenization* memecah kalimat komentar teks menjadi kata-kata tunggal. Pada aplikasi RapidMiner proses *tokenization* juga menghapus angka dan karakter khusus yang dianggap sebagai tanda baca. Contoh hasil kedua proses ini disajikan pada Tabel 2.

Tabel 2. Contoh hasil *transform cases* dan *tokenization*

Sebelum	Sesudah <i>transform cases</i>	Sesudah <i>tokenization</i>
Materi bagus	materi bagus	materi bagus
sangat bermanfaat bagi ASN, bahan ajar menarik	sangat bermanfaat bagi asn, bahan ajar menarik	sangat bermanfaat bagi asn bahan ajar menarik
sudah bagus, masih banyak istilah yang menggunakan B. Inggris, mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti. terimakasih	sudah bagus, masih banyak istilah yang menggunakan b. inggris, mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti. terimakasih	sudah bagus masih banyak istilah yang menggunakan b inggris mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti terimakasih

*Filter token by length* menghapus token yang dihasilkan tahap *tokenization* dengan panjang kurang dari 4 huruf. *Filter stopwords* diterapkan pada hasil output filter token menghilangkan kata-kata yang sifatnya tidak relevan dengan analisis seperti kata hubung atau kata tanya, sehingga token yang tersisa hanya berisi kata-kata kunci untuk dianalisis. Proses ini

menggunakan daftar *stopwords* bahasa Indonesia yang diunduh dari situs kaggle.com. Contoh hasil *filter token by length* dan *filter stopwords* disajikan pada Tabel 3.

Tabel 3. Contoh hasil *filter token by length* dan *filter stopwords*

Sebelum	Sesudah <i>filter token by length</i>	Sesudah <i>filter stopwords</i>
materi bagus	materi bagus	materi bagus
sangat bermanfaat bagi asn bahan ajar menarik	sangat bermanfaat bagi bahan ajar menarik	bermanfaat bahan ajar menarik
sudah bagus masih banyak istilah yang menggunakan b inggris mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti terimakasih	sudah bagus masih banyak istilah yang menggunakan inggris mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti terimakasih	bagus istilah inggris diterjemahkan peserta mudah mengerti terimakasih

Tahap *pre-processing text* terakhir yaitu *N-gram* menggunakan operator Generate N-gram pada aplikasi RapidMiner. Proses ini dilakukan untuk menghasilkan kombinasi n kata atau karakter yang berurutan dalam suatu teks sehingga dapat mempertahankan konteks kata-kata dalam teks seperti kata “terima” dan “kasih” yang berkaitan dengan makna apresiasi “terima kasih”. Jumlah kombinasi kata yang digunakan pada penelitian ini sebanyak 2 (bi-gram). Contoh hasil penerapan bi-gram disajikan pada Tabel 4.

Tabel 4. Contoh hasil *generate n-gram (bi-gram)*

Sebelum	Sesudah*
materi bagus	'materibagus'
bermanfaat bahan ajar menarik	'bermanfaat bahan' 'bahan ajar' 'ajar menarik'
bagus istilah inggris diterjemahkan peserta mudah mengerti terimakasih	'bagus istilah' 'istilah inggris' inggris diterjemahkan' 'diterjemahkan peserta' 'peserta mudah' 'mudah mengerti' 'mengerti terimakasih'

\*) tanda petik menggambarkan pemisahan atribut bukan merupakan bagian output proses

## 2. Feature extraction

Teks komentar yang telah melalui *pre-processing* dirubah menjadi representasi vector numerik agar dapat diinput untuk membangun model. Vektorisasi TF-IDF pada aplikasi RapidMiner menggunakan operator Process Document from Data dengan mengatur parameter pembuatan vektor menjadi TF-IDF. Contoh output tahap ini sebagaimana Tabel 5.

Tabel 5. Contoh hasil *feature extraction* vektor TF-IDF

Teks	ajar	ajar_menarik	bagus	bagus_istilah	...	peserta_mudah	terimakasih
Materi bagus	0,00	0,00	0,25	0,00	...	0,00	0,00
sangat bermanfaat bagi ASN, bahan ajar menarik	0,38	0,38	0,00	0,00	...	0,00	0,00
sudah bagus, masih banyak istilah yang menggunakan B. Inggris, mungkin bisa diterjemahkan supaya peserta lebih mudah mengerti. terimakasih	0,00	0,00	0,10	0,27	...	0,27	0,27

## 3. Modeling dan Model Evaluation

Pembangunan model klasifikasi sentiment menggunakan algoritma Naïve Bayes (NB), Support Vector Machine (SVM) dan Deep Learning (DL). Proses *modeling* pada aplikasi RapidMiner dilakukan sekaligus untuk menghasilkan evaluasi model menggunakan operator Cross Validation. Hasil evaluasi performansi dengan teknik cross validation 10 kali iterasi pada ketiga model yang terbentuk menghasilkan metrik akurasi, presisi dan *recall* yang memberikan gambaran kinerja model dalam melakukan prediksi prediksi sentimen pada data training (Tabel 6). Hasil tersebut menunjukkan bahwa model klasifikasi Naïve Bayes (NB) memiliki performa lebih baik dibandingkan dengan model klasifikasi Support Vector Machine (SVM) dan Deep Learning (DL). NB memiliki tingkat akurasi dan recall sebesar 82,54% dan 68,08% yang lebih tinggi dibandingkan SVM dan DL. Akan tetapi berdasarkan metrik presisi, algoritma SVM menunjukkan performa yang lebih baik dibandingkan NB dan DL dengan nilai 73,66%.

Hasil penelitian ini serupa dengan hasil analisis sentimen e-sport untuk kurikulum



edukasi [27] dimana algoritma NB mengungguli kinerja algoritma SVM dalam melakukan prediksi sentiment. Hasil ini berbeda dengan penelitian lain yang dilakukan [25], [26]. Algoritma SVM mengungguli algoritma NB dalam prediksi sentimen jika metrik keputusan yang digunakan adalah presisi. Penelitian ini sekaligus membuktikan algoritma DL pada analisis sentimen mampu mencapai akurasi cukup tinggi yaitu 79,95%, sebagaimana penelitian sebelumnya yang mampu mencapai akurasi di atas 60% [28]. Perbandingan ketiga model pada penelitian ini turut mendukung hasil studi literatur [23] di mana ketiga model memiliki kinerja yang bagus pada analisis sentimen sehingga banyak diimplementasikan pada penelitian *text mining*.

Tabel 6. Hasil evaluasi model klasifikasi NB, SVM dan DL

Algoritma	Accuracy	Recall	Precision
Naïve Bayes	<b>82,54</b>	<b>68,08</b>	69,81
SVM	78,67	54,18	<b>73,66</b>
Deep learning	79,95	58,76	68,42

Hasil *confusion matrix* (Tabel 7-9) memperlihatkan hasil prediksi sentimen ke dalam kelas apresiasi, saran dan kritik untuk tiap model yang terbentuk. Mengacu kepada tujuan evaluasi pelatihan adalah menentukan perbaikan untuk meningkatkan kualitas pelatihan, maka model yang dipilih adalah model yang mampu meminimalisir *false* apresiasi. Kondisi *false* apresiasi ialah salah memprediksi saran/kritik sebagai apresiasi sehingga mengakibatkan kesalahan dalam menyimpulkan kualitas pelatihan, hal ini berdampak hilangnya informasi perbaikan yang perlu dilakukan. *Confusion matrix* yang terbentuk menunjukkan algoritma NB mampu memprediksi saran dan kritik dengan lebih baik dibandingkan algoritma SVM dan DL, oleh karena itu algoritma NB dinilai sebagai *classifier* terbaik pada penelitian ini dan dapat digunakan untuk memprediksi data komentar baru.

Tabel 7. Confusion matrix model klasifikasi Naïve Bayes

	true apresiasi	true saran	true kritik
Pred. apresiasi	564	14	36
pred. saran	15	78	40
pred. kritik	14	30	62
class recall	95,11%	63,93%	44,93%

Tabel 8. *Confusion matrix* model klasifikasi Support Vector Machine

	true apresiasi	true saran	true kritik
Pred. apresiasi	590	61	86
pred. saran	2	43	14
pred. kritik	1	18	38
class recall	99,49%	35,25%	27,54%

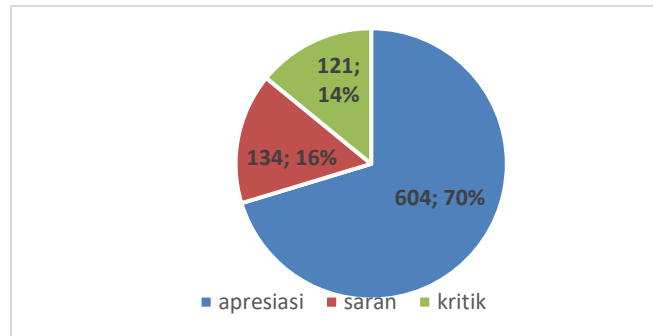
Tabel 9. *Confusion matrix model* klasifikasi Deep Learning

	true apresiasi	true saran	true kritik
Pred. apresiasi	579	44	55
pred. saran	4	44	24
pred. kritik	10	34	59
class recall	97,64%	36,07%	42,75%

#### 4. Prediction

Naïve Bayes yang merupakan *classifier* terpilih digunakan untuk memprediksi sentimen pada data yang belum diberikan label sebanyak 859 komentar. Hasil prediksi diperoleh 604 komentar adalah apresiasi (70%), 134 komentar merupakan saran (16%) dan sisanya sebanyak

121 komentar (14%) adalah kritik (Gambar 5). Adapun contoh hasil prediksi sentimen menggunakan model ini disajikan pada Tabel 10.



Gambar 5. Proporsi sentimen hasil prediksi model Naïve Bayes

Tabel 10. Contoh hasil prediksi komentar menggunakan model Naïve Bayes

Komentar peserta	Prediksi
Menarik,bagus,tp terlalu lama untuk 1 sesi elearning	kritik
sudah bagus, video dibuat menarik dan mudah diakses. mungkin video pembelajaran juga dapat didownload oleh peserta.	saran
sangat baik dan sangat membantu	apresiasi

Hasil tersebut memperlihatkan efektivitas *machine learning* dalam melakukan prediksi sentimen komentar teks jumlah banyak secara otomatis ke dalam kelas apresiasi, saran dan kritik. Pendekatan ini juga memiliki prosedur yang lebih sederhana dibandingkan teknik analisis konten konvensional yang membutuhkan prosedur pengkodean yang rumit pada data teks dalam jumlah besar. Dengan demikian penggunaan *machine learning* pada evaluasi pelatihan dapat mengurangi sumber daya yang dibutuhkan pihak penyelenggara dalam analisis untuk memperoleh informasi yang berkualitas dari data teks jumlah besar.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan penelitian sentimen reaksi peserta *e-learning communication skill* tingkat dasar dengan pendekatan *machine learning*, diperoleh kesimpulan model klasifikasi NB mampu memberikan hasil prediksi yang lebih baik dibandingkan model SVM dan DL dengan tingkat akurasi, *recall* dan presisi masing-masing sebesar 82,54%, 68,08% dan 69,81%. Prediksi terhadap 859 komentar peserta menggunakan model NB diperoleh wawasan bahwa 70% peserta merasa puas dengan memberikan apresiasi atas penyelenggaraan e-learning, 16% memberikan saran peningkatan terhadap perbaikan dan peningkatan kualitas e-learning dan 14% memberikan kritik atas pengalaman mengikuti e-learning.

Pemanfaatan *machine learning* pada hasil evaluasi pelatihan masih sangat terbatas, oleh karena itu eksplorasi teknik *machine learning* lainnya pada area ini masih terbuka luas. Penelitian ini mengklasifikasikan komentar peserta berdasarkan kalimat sentimen atau non topik, pendekatan klasifikasi berdasarkan topik dapat diterapkan pada penelitian lanjutan agar diperoleh informasi lebih terkait topik perbaikan kualitas pelatihan. Disamping itu, sentimen analisis dengan *machine learning* sangat bergantung terhadap jumlah dan kualitas data training yang digunakan, oleh karena itu penelitian sentimen dengan menyoroti penanganan data teks dari berbagai sumber sangat dianjurkan untuk dilakukan.

Pada penelitian ini tingkat akurasi yang menggambarkan kemampuan model memprediksi dengan tepat sentimen komentar peserta maksimal sebesar 82,54%. Penelitian [23] mengungkapkan pemanfaatan *ensemble learning* yaitu penggabungan beberapa model *machine learning* dapat meningkatkan akurasi model hingga mencapai akurasi sebesar 90%. Oleh karena

itu peneliti mendorong penggunaan algoritma lain seperti Random Forest, CNN atau *ensemble learning* seperti CLSTM pada analisis sentimen evaluasi reaksi pelatihan untuk menghasilkan model prediksi evaluasi reaksi peserta pelatihan yang lebih baik lagi.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Pendidikan dan Pelatihan Keuangan, Kementerian Keuangan, khususnya Pusdiklat Kepemimpinan dan Manajerial yang telah memberikan dukungan terhadap terlaksananya penelitian ini.

#### DAFTAR PUSTAKA

- [1] M. A. Memon, R. Sallaeh, M. N. R. Baharom, S. Md Nordin, and H. Ting, "The relationship between training satisfaction, organisational citizenship behaviour, and turnover intention: A PLS-SEM approach," *Journal of Organizational Effectiveness*, vol. 4, no. 3, pp. 267–290, Sep. 2017, doi: 10.1108/JOEPP-03-2017-0025.
- [2] N. P. Z. Thwin, Y. N. Soe, A. Bhaumik, and Midhunchakkarayarthi, "Impact of training and development on employee performance: a study of Myanmar internet industry," *Proceedings on Engineering Sciences*, vol. 4, no. 3, pp. 291–300, 2022, doi: 10.24874/PES04.03.007.
- [3] K. Nama, B. Daweti, M. Lourens, and T. Chikukwa, "The impact of training and development on employee performance and service delivery at a local municipality in South Africa," *Problems and Perspectives in Management*, vol. 20, no. 4, pp. 42–51, 2022, doi: 10.21511/ppm.20(4).2022.04.
- [4] S. Aishwarya, M. Ayishamillath, and K. M. Ali, "Analyzing the impact of learning and development on employee engagement in information technology sector," *International Journal of Advanced Science and Technology*, vol. 28, no. 19, pp. 1272–1282, 2019.
- [5] H. Aguinis and K. Kraiger, "Benefits of training and development for individuals and teams, organizations, and society," *Annual Review of Psychology*, vol. 60, pp. 451–474, Jan. 2009. doi: 10.1146/annurev.psych.60.110707.163505.
- [6] A. Alsalamah and C. Callinan, "The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020)," *Industrial and Commercial Training*, vol. 54, no. 1. Emerald Group Holdings Ltd., pp. 36–63, 2021. doi: 10.1108/ICT-12-2020-0115.
- [7] D. Eseryel, "Approaches to Evaluation of Training: Theory & Practice," *Educational Technology & Society*, vol. 5(2), pp. 93–98, 2002.
- [8] S. Kim, S. Park, J. Lavelle, M. Kim, and S. Chaudhuri, "Revisiting trainee reactions: A multilevel analysis of the nomological network," *Hum Resour Dev Q*, vol. 31, no. 2, pp. 173–191, Jun. 2020, doi: 10.1002/hrdq.21383.
- [9] K. G. Brown, "An examination of the structure and nomological network of trainee reactions: A closer look at 'smile sheets,'" *Journal of Applied Psychology*, vol. 90, no. 5, pp. 991–1001, Sep. 2005, doi: 10.1037/0021-9010.90.5.991.
- [10] W. Arthur, W. Bennett, P. S. Edens, and S. T. Bell, "Effectiveness of training in organizations: A meta-analysis of design and evaluation features," *Journal of Applied Psychology*, vol. 88, no. 2, pp. 234–245, Apr. 2003, doi: 10.1037/0021-9010.88.2.234.
- [11] R. P. Harman, J. Kemp Ellington, E. A. Surface, and L. F. Thompson, "Exploring qualitative training reactions: Individual and contextual influences on trainee commenting," *Journal of Applied Psychology*, vol. 100, no. 3, pp. 894–916, May 2015, doi: 10.1037/a0038380.
- [12] T. Sitzmann, K. G. Brown, W. J. Casper, K. Ely, and R. D. Zimmerman, "A Review and Meta-Analysis of the Nomological Network of Trainee Reactions," *Journal of Applied Psychology*, vol. 93, no. 2, pp. 280–295, Mar. 2008, doi: 10.1037/0021-9010.93.2.280.
- [13] A. Giangreco, A. Sebastiano, and R. Peccei, "Trainees' reactions to training: An analysis of the factors affecting overall satisfaction with training," *International Journal of Human*

- Resource Management*, vol. 20, no. 1, pp. 96–111, 2009, doi: 10.1080/09585190802528417.
- [14] K. Kraiger, “Decision-based evaluation,” 2002. [Online]. Available: <https://www.researchgate.net/publication/281601028>
- [15] D. R. Glerum, D. L. Joseph, A. F. McKenny, and B. A. Fritzsche, “The trainer matters: Cross-classified models of trainee reactions,” *Journal of Applied Psychology*, vol. 106, no. 2, pp. 281–299, 2021, doi: 10.1037/apl0000503.
- [16] A. Gegenfurtner, A. Zitt, and C. Ebner, “Evaluating webinar-based training: a mixed methods study of trainee reactions toward digital web conferencing,” *Int J Train Dev*, vol. 24, no. 1, pp. 5–21, 2020, doi: 10.1111/ijtd.12167.
- [17] K. Mehale, C. M. Govender, and C. M. Mabaso, “Maximising training evaluation for employee performance improvement,” 2021, doi: 10.4102/sajhrm.
- [18] F. Vencovsky, “Service Quality Evaluation Using Text Mining: A Systematic Literature Review,” in *Lecture Notes in Business Information Processing*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 159–173. doi: 10.1007/978-3-030-61140-8\_11.
- [19] F. V. Ordenes, B. Theodoulidis, J. Burton, T. Gruber, and M. Zaki, “Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach,” *J Serv Res*, vol. 17, no. 3, pp. 278–295, 2014, doi: 10.1177/1094670514524625.
- [20] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” 2021.
- [21] A. Gupta, V. Dengre, H. A. Kheruwala, and M. Shah, “Comprehensive review of text-mining applications in finance,” *Financial Innovation*, vol. 6, no. 1. Springer Science and Business Media Deutschland GmbH, Dec. 01, 2020. doi: 10.1186/s40854-020-00205-1.
- [22] S. Pawar, N. Ramrakhiyani, G. K. Palshikar, and S. Hingmire, *Natural Language Processing and Information Systems*, vol. 9103. in Lecture Notes in Computer Science, vol. 9103. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-19581-0.
- [23] L. P. Hung and S. Alias, “Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 1. Fuji Technology Press, pp. 84–95, Jan. 01, 2023. doi: 10.20965/jaciii.2023.p0084.
- [24] J. Liu, S. Hu, F. Mehraliyev, and H. Liu, “Text classification in tourism and hospitality – a deep learning perspective,” *International Journal of Contemporary Hospitality Management*. Emerald Publishing, 2023. doi: 10.1108/IJCHM-07-2022-0913.
- [25] N. A. K. M. Haris, S. Mutalib, A. M. A. Malik, S. Abdul-Rahman, and S. N. K. Kamarudin, “Sentiment classification from reviews for tourism analytics,” *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 1, pp. 108–120, Mar. 2023, doi: 10.26555/ijain.v9i1.1076.
- [26] D. A. Nurdeni, I. Budi, and A. B. Santoso, “Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer,” in *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 122–127. doi: 10.1109/EIConCIT50028.2021.9431852.
- [27] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. Septia Nugraha, and W. Gata, “Sentiment analysis on e-sport for education curriculum using Naive Bayes and Support Vector Machine,” *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)*, vol. 13, no. 2, pp. 109–122, 2020.
- [28] S. P. Kristanto, L. Hakim, D. Yusuf, E. S. Haq, and A. R. Asyari, “Classification of Public Opinion on Vaccine Administration Using Convolutional Neural Network,” in *2022 5th International Conference on Vocational Education and Electrical Engineering: The Future of Electrical Engineering, Informatics, and Educational Technology Through the Freedom*

- of Study in the Post-Pandemic Era, ICVEE 2022 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 65–71. doi: 10.1109/ICVEE57061.2022.9930412.
- [29] R. Haque, N. Islam, M. Tasneem, and A. K. Das, “Multi-class sentiment classification on Bengali social media comments using machine learning,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21–35, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.001.
- [30] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, “Mining product innovation ideas from online reviews,” *Inf Process Manag*, vol. 58, no. 1, Jan. 2021, doi: 10.1016/j.ipm.2020.102389.
- [31] 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, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- [32] W. Y. Ayele, “Adapting CRISP-DM for Idea Mining A Data Mining Process for Generating Ideas Using a Textual Dataset,” 2020. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [33] M. Farid Naufal, A. Fernando Susanto, C. Nathaneil Kansil, S. Huda, and K. kunci, “Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Potensi Hilangnya Nasabah Bank Application of Machine Learning to Predict Potential Loss of Bank Customer,” *Februari*, vol. 22, no. 1, pp. 1–11, 2023.