
Analisis Sentimen Aplikasi Tiktok Tokopedia Seller Center dengan Pendekatan Machine learning: SVM, CNN, Naive Bayes

Neha Poetri Setiawati^{1*}, Nurmalitasari², Vihi Atina³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Duta Bangsa Surakarta

Email: ¹202020853@mhs.udb.ac.id, ²nurmalitasari@udb.ac.id, ³vihi_atina@udb.ac.id

(Naskah masuk: 18 Juli 2024, diterima untuk diterbitkan: 10 Januari 2025)

Abstrak: Tiktok Tokopedia Seller Center merupakan aplikasi yang diperuntukan untuk para pelaku usaha atau para penjual mengenai pengelolaan dan meningkatkan penjualan mereka pada platform Tiktok (Tiktok shop). Penelitian ini memiliki tujuan untuk melakukan analisis sentimen negatif dan positif dari pengguna mengenai fitur dan kinerja dari aplikasi Tiktok Tokopedia Seller Center dengan menggunakan 3 algoritma Machine learning, diantaranya SVM, CNN, Naive Bayes. Dikarenakan ketiga algoritma tersebut dinilai efektif dalam menangani kasus analisis sentimen. Data untuk penelitian ini didapatkan dari ulasan aplikasi Tiktok Tokopedia Seller Center yang terdapat pada Google Play Store. Hasil penelitian menunjukkan bahwa SVM memiliki kinerja dan performa yang lebih unggul dibandingkan CNN, dan Naive Bayes, karena SVM didapatkan akurasi akhir sebesar 0.90, disusul CNN dengan akurasi sebesar 0.88 dan Naive Bayes dengan akurasi sebesar 0.84. Dari penelitian ini dapat memberikan gambaran atau wawasan bagi pengembang aplikasi Tiktok Tokopedia Seller Center dalam melakukan perbaikan dan meningkatkan kualitas pelayanan.

Kata Kunci – SVM; CNN; Naive Bayes; Machine learning; Analisis Sentimen

Sentiment Analysis of Tiktok Tokopedia Seller Center with Machine learning Approach: SVM, CNN, Naive Bayes

Abstract: Tiktok Tokopedia Seller Center is an application intended for business people or sellers regarding management and increasing their sales on the Tiktok platform (Tiktok shop). This study aims to analyze negative and positive sentiments from users regarding the features and performance of the Tiktok Tokopedia Seller Center application using 3 Machine learning algorithms, including SVM, CNN, Naive Bayes. Because these three algorithms are considered effective in handling sentiment analysis cases. The data for this study were obtained from reviews of the Tiktok Tokopedia Seller Center application found on the Google Play Store. The results showed that SVM has superior performance and performance compared to CNN, and Naive Bayes, because SVM obtained a final accuracy of 0.90, followed by CNN with an accuracy of 0.88 and Naive Bayes with an accuracy of 0.84. From this research, it can provide an overview or insight for the Tiktok Tokopedia Seller Center application developer in making improvements and improving service quality.

Keywords – SVM; CNN; Naive Bayes; Machine learning; Sentiment Analysis

1. PENDAHULUAN

Tiktok Tokopedia Seller Center merupakan aplikasi yang bergerak pada bidang sosial e-commerce di Indonesia. Tiktok Tokopedia Seller Center merupakan inovasi terbaru dari Aplikasi Tiktok Seller Center, yang dilatar belakangi karena adanya kerjasama antara Tiktok dan Tokopedia pada tanggal 11 Desember 2023 yang notabennya adalah perusahaan e-commerce dalam negeri. Tiktok menggandeng Tokopedia sebagai mitra bisnisnya dikarenakan kedua e-commerce tersebut memiliki visi dan misi yang sama, yaitu untuk memajukan UMKM lokal. Majemen PT GoTo Gojek Tokopedia Tbk dan pihak Tiktok menyimpulkan bahwa adanya kerja sama yang terjalin tersebut mendukung lebih dari 90% pelaku UMKM di Indonesia melalui program yang telah di buat oleh pihak Tiktok dan PT GoTo Gojek Tokopedia [1], [2].

Tiktok Tokopedia Seller Center adalah aplikasi yang dirancang untuk membantu penjual mengelola dan meningkatkan penjualan mereka. Tiktok Tokopedia Seller Center memiliki fitur yang cukup lengkap bagi seorang merchant atau pelaku usaha untuk menjalankan bisnis e-commerce mereka di Tiktok shop. Aplikasi ini menyiapkan solusi lengkap yang mencakup manajemen pengoperasian toko, berbagai fitur transaksi pembayaran, iklan, laporan penjualan, fitur pemantauan, perkembangan penjualan, layanan dan engagement pelanggan, analisis kinerja, pembaruan kebijakan dan lain sebagainya. Dengan aplikasi ini pelaku bisnis dapat meningkatkan penjualan dengan cara mengoptimalkan strategi marketing atau pemasaran. Tiktok Tokopedia Seller Center pada Google Play store telah didownload 10 juta kali, dan memiliki sebanyak 220 ribu ulasan pada tahun 2024. Ulasan yang diberikan pengguna memberikan pengaruh untuk aplikasi Tiktok Tokopedia Seller Center dalam melakukan perbaikan aplikasi [3].

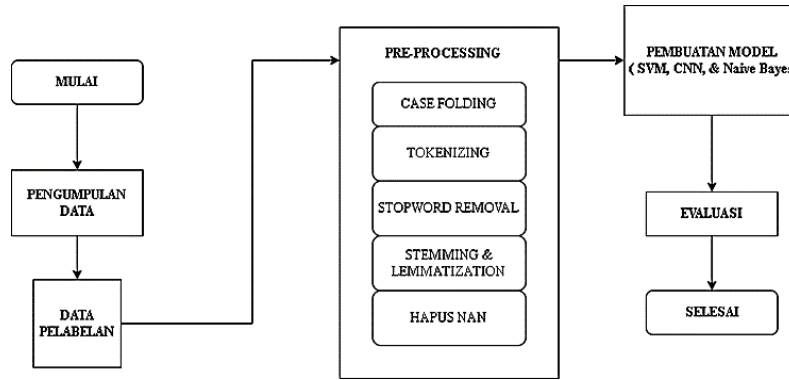
Pada Google Play Store, banyaknya ulasan aplikasi dari pengguna yang tidak terstruktur membuatnya sulit untuk dipahami oleh calon pengguna. Ulasan tersebut berisi sentimen mengenai kepuasan dari pengguna terhadap aplikasi Tiktok Tokopedia Seller Center. Karena sebab itu penelitian ini menggunakan analisis sentimen. Analisis sentimen merupakan pembelajaran komputasi yang bertujuan untuk mengenali dan mengekspresikan sebuah pandangan, sentimen, evaluasi, sikap, emosional, subjektifitas yang terdapat pada suatu teks [4] Dalam melakukan analisis sentimen terhadap Aplikasi Tiktok Tokopedia Seller Center, penulis menggunakan metode pendekatan Machine learning dengan algoritma SVM, CNN, dan Naïve Bayes. Machine learning memiliki kemampuan dalam mengidentifikasi pola yang tidak linear dan pola yang berbeda antar kelas dalam data[5]

Support Vector Machine (SVM) merupakan salah satu algoritma Machine learning yang dapat menunjukkan kinerja terbaik untuk klasifikasi sentimennya. Kelebihan dari SVM adalah algoritma yang mampu mengatasi kasus klasifikasi teks bebas linear. Seperti pada penelitian yang dilakukan oleh Rizky Wahyudi & Gilang Kusumawardana[6] dalam penelitiannya menjelaskan bahwa penggunaan metode SVM dalam penelitiannya untuk mendapatkan hasil klasifikasi teks review yang lebih baik dan akurat, terbukti dalam penelitian ini mendapatkan akurasi yang bagus sebesar 85,54%. *Convolutional Neural Network* (CNN) merupakan jaringan syaraf tiruan berlapis dengan propagasi maju kemudian dihubungkan dengan fully connected layer seperti multilayer neural network. CNN dapat memanfaatkan operasi konvolusi untuk mengekstrak fitur dari data visual, sehingga mampu menganalisis dan memahami pola dan bentuk tersebut[7] Algoritma CNN juga mampu mengidentifikasi fitur relevan maupun tidak relevan dalam sebuah data, sehingga dapat mengurangi kemungkinan overfitting. Seperti pada penelitian yang dilakukan oleh Faiz Adil Khatami[7] model CNN dilakukan modifikasi epoch untuk meningkatkan prediksi tepat analisis, dengan merubah learning, partisi data, batch size, dan epoch, mendapatkan hasil akurasi sebesar 85%, ini dapat menggambarkan bahwa algoritma CNN dapat memberikan kinerja yang baik dalam analisis sentimen. Naive Bayes merupakan algoritma Machine learning dengan model sederhana untuk kasus analisis sentimen, selain itu algoritma ini juga populer dikarenakan sering digunakan dalam klasifikasi teks berdasarkan teorema bayes. Seperti dalam penelitian yang dilakukan oleh Eko Kurnianto, dan Dimas Febriawan[8] penelitiannya tersebut terdapat kesimpulan algoritma naïve bayes terbukti menjadi pilihan yang tepat untuk analisis sentimen karena memperoleh hasil akurasi mencapai 86,97% menunjukkan bahwa kinerja model naïve bayes yang baik.

Dengan demikian, penelitian ini dilakukan untuk membandingkan hasil analisis sentimen yang diperoleh dari masing-masing algoritma Machine learning diatas, dan dapat juga memberikan wawasan mengenai algoritma yang cocok dalam kasus analisis sentimen.

2. METODE PENELITIAN

Pada penelitian ini menggunakan metode kuantitatif, yaitu salah satu metode yang digunakan untuk menyelesaikan kasus penelitian yang memiliki keterkaitan dengan data berupa angka dan statistik [7] Berikut kami sajikan kerangka berpikir atau tahapan dari penelitian ini seperti pada Gambar 1.



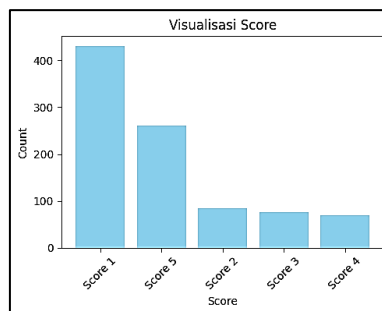
Gambar 1. Metodologi Penelitian

Tahapan analisis sentimen aplikasi Tiktok Tokopedia Seller Center dengan Pendekatan Machine learning: SVM, CNN, Naive Bayes disajikan pada gambar 1. Proses pengumpulan data menggunakan teknik scrapping dengan menggunakan pustaka pemrograman Python untuk mengambil data ulasan pengguna aplikasi Tiktok Tokopedia Seller Center pada *Google-Play-Store*. Setelah data diperoleh maka diperoleh dataset lalu dilakukan proses pelabelan data untuk menentukan sentimen negatif dan positif dari ulasan pengguna. Selanjutnya dilakukan pre-processing yaitu melakukan penyaringan teks ke dalam proses case folding, tokenizig, stopwords removal stemming & lemmatization, dan menghapus teks NaN. Pre-processing ini mempunyai peran penting dalam menghasilkan akurasi dan performa model algoritma yang telah ditentukan [9]. Data yang sudah dilakukan pre-processing siap untuk diimplementasikan ke dalam algoritma SVM, CNN, dan Naive Bayes dan didapatkan hasil analisis sentimennya. Pada bagian evaluasi dilakukan pengujian terhadap model yang bertujuan untuk mengukur kemampuan kinerja dan permorma dari model yang telah di bentuk melalui parameter accuracy, presicion, recall, f1-score, dan support sehingga dapat diperoleh kesimpulan dari penelitian tersebut.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan teknik scrapping dengan menggunakan pustaka *Google-Play-scrapper* pemrograman Python. Data yang digunakan ini merupakan data ulasan pengguna Aplikasi Tik Tok Tokopedia Seller Center dari *Google-Play-Store*. Dengan teknik scraping ini, peneliti berhasil mengumpulkan 995 menjadi 921 dikarenakan data yang diambil hanya data dari tahun 2023 sampai dengan pertengahan 2024 (data terbaru). Untuk Analisis sentimen pada penelitian ini hanya menggunakan 2 atribut saja yaitu content dan score dari pengguna. Berikut kami sajikan diagram komposisi Score satu sampai dengan 5.

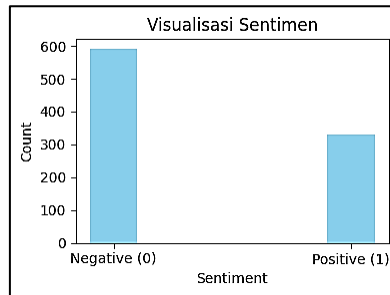


Gambar 2. Visualisasi Komposisi Score

Dapat dilihat dari Gambar 2. bahwa score yang medominasi adalah score 1 yang jumlahnya diatas 400 ulasan, dan score 4 merupakan score paling rendah.

3.2. Pelabelan Data

Proses Pelabelan data ini berfungsi untuk memberikan label sentimen pada ulasan agar memudahkan pengolahan data nantinya. Label sentimen pada data ini kami melabelkan 0 sebagai kateogri sentimen negatif, dan 1 sebagai kategori sentimen positif. Pemberian lebel negatif dan positif tersebut didasarkan pada pemberia *score* pengguan. Apabila *score* dibawah 3 akan dikategorikan sentimen negatif (0) dan *score* diatas 3 masuk ke dalam skategori sentimen positif. Berikut diagram komposisi sentimen



Gambar 3. Diagram Komposisi Score

Dapat dilihat dari gambar di atas komposisi sentimen kelas negatif (0) lebih dominan daripada kelas positif (1).

3.3. Pre-Processing

Pre-processing data adalah teknik awal untuk mengubah data mentah menjadi data siap olah[10]. Pre-processing ini mempunyai peran penting dalam menghasilkan akurasi dan performa model algoritma yang telah ditentukan [9]. Urutan pre-processing dimulai dari case folding, tokenizing, stopword removal, stemming & lemmatization, dan menghapus teks NaN.

3.3.1. Case Folding

Tahap pertama adalah case folding, dimana pada proses ini mengubah semua huruf pada teks menjadi huruf kecil, yang bertujuan untuk mengurangi banyak variasi pada teks karena hal tersebut dapat mempengaruhi pada saat pemrosesan dan analisis.

Tabel 1. Sampel Data Case Folding

Teks_Input	Teks_Ouput
Pelanggaran Pelanggaran dan Pelanggaran Padahal nilai produk saya termasuk bagus malah produk di bekukan gara ² penilaian negatif Harusnya admin tiktok bisa	pelanggaran pelanggaran dan pelanggaran padahal nilai produk saya termasuk bagus malah produk di bekukan gara ² penilaian negatif Harusnya admin tiktok bisa
aplikasi yang sangat menyesengsarakan penjual, mau untungnya saja, sudah iklan mahal, putaran duit sangat lama, sering kena point. aplikasi ga jelas. sangat kecewa	aplikasi yang sangat menyesengsarakan penjual, mau untungnya saja, sudah iklan mahal, putaran duit sangat lama, sering kena point. aplikasi ga jelas. sangat kecewa

3.3.2. Tokenizing

Proses dimana teks dipotong menjadi bagian yang lebih kecil atau biasa disebut dengan token[10].

Tabel 2. Sampel Data *Tokenizing*

Teks_Input	Teks_Ouput
pelanggaran pelanggaran dan pelanggaran padahal nilai produk saya termasuk bagus malah produk di bekukan garaa ² penilaian negatif Harusnya admin tiktok bisa	['pelanggaran', 'pelanggaran', 'dan', 'pelanggaran', 'padahal', 'nilai', 'produk', 'saya', 'termasuk', 'bagus', 'malah', 'produk', 'di', 'bekukan', 'garaa ² ', 'penilaian', 'negatif', 'Harusnya', 'admin', 'tiktok', 'bisa']
aplikasi yang sangat menyenesarakan penjual, mau untungnya saja, sudah iklan mahal, putaran duit sangat lama, sering kena point. aplikasi ga jelas. sangat kecewa	['aplikasi', 'yang', 'sangat', 'menyenesarakan', 'penjual', 'mau', 'untungnya', 'saja', 'sudah', 'iklan', 'mahal', 'putaran', 'duit', 'sangat', 'lama', 'sering', 'kena', 'point', 'aplikasi', 'ga', 'jelas', 'sangat', 'kecewa']

3.3.3. Stopword Removal

Stopword Removal atau penghapusan kata umum yang sering muncul pada teks dan tidak memberikan nilai signifikan dalam analisis, seperti "atau", "di", "dari", "yang", dan sejenisnya. Tujuan dari proses ini yaitu untuk membantu mengurangi dimensi data serta meningkatkan fokus dalam kata penting yang lebih bermakna dalam proses analisis.

Tabel 3. Sampel Data *Stopword Removal*

Teks_Input	Teks_Ouput
['pelanggaran', 'pelanggaran', 'dan', 'pelanggaran', 'padahal', 'nilai', 'produk', 'saya', 'termasuk', 'bagus', 'malah', 'produk', 'di', 'bekukan', 'garaa ² ', 'penilaian', 'negatif', 'Harusnya', 'admin', 'tiktok', 'bisa']	['pelanggaran', 'pelanggaran', 'pelanggaran', 'padahal', 'nilai', 'produk', 'saya', 'termasuk', 'bagus', 'malah', 'produk', 'bekukan', 'garaa ² ', 'penilaian', 'negatif', 'Harusnya', 'admin', 'tiktok', 'bisa']
['aplikasi', 'yang', 'sangat', 'menyenesarakan', 'penjual', 'mau', 'untungnya', 'saja', 'sudah', 'iklan', 'mahal', 'putaran', 'duit', 'sangat', 'lama', 'sering', 'kena', 'point', 'aplikasi', 'ga', 'jelas', 'sangat', 'kecewa']	['aplikasi', 'sangat', 'menyenesarakan', 'penjual', 'mau', 'untungnya', 'saja', 'sudah', 'iklan', 'mahal', 'putaran', 'duit', 'sangat', 'lama', 'sering', 'kena', 'point', 'aplikasi', 'ga', 'jelas', 'sangat', 'kecewa']

3.3.4. Stemming & Lemmatization

Tabel 4. Sampel Data *Stopword Removal*

Teks_Input	Teks_Ouput
['pelanggaran', 'pelanggaran', 'pelanggaran', 'padahal', 'nilai', 'produk', 'saya', 'termasuk', 'bagus', 'malah', 'produk', 'bekukan', 'garaa ² ', 'penilaian', 'negatif', 'Harusnya', 'admin', 'tiktok', 'bisa']	['pelanggaran', 'pelanggaran', 'pelanggaran', 'padahal', 'nilai', 'produk', 'saya', 'termasuk', 'bagus', 'malah', 'produk', 'beku', 'gara', 'nilai', 'negatif', 'Harus', 'admin', 'tiktok', 'bisa']
['aplikasi', 'sangat', 'menyenesarakan', 'penjual', 'mau', 'untungnya', 'saja', 'sudah', 'iklan', 'mahal', 'putaran', 'duit', 'sangat', 'lama', 'sering', 'kena', 'point', 'aplikasi', 'ga', 'jelas', 'sangat', 'kecewa']	['aplikasi', 'sangat', 'sengsara', 'penjual', 'mau', 'untung', 'saja', 'sudah', 'iklan', 'mahal', 'putaran', 'duit', 'sangat', 'lama', 'sering', 'kena', 'point', 'aplikasi', 'ga', 'jelas', 'sangat', 'kecewa']

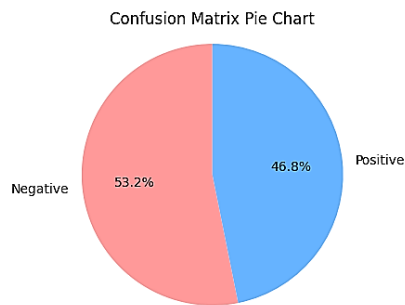
Stemming yaitu tahapan ini terdapat proses mengganti kata ke bentuk dasarnya dan menghilangkan imbuhan. Begitupun dengan Lemmatization juga berfungsi untuk mereduksi setiap kata ke dalam bentuk dasar, akan tetapi lebih mempertahankan arti kata dan mempertimbangkan

konteks. Tujuan dari tahapan ini agar model algoritma dapat mudah mengenali pola dan makna suatu teks. Dan yang terakhir menghapus teks NaN, dan pada data ini tidak ada teks yang NaN

3.4. Pembuatan Model

3.4.1 Support Vector Machine (SVM)

Pembuatan model SVM diawali dengan melakukan TF-Idf pada dataset yang telah disiapkan, lalu dilakukan pembuatan Model SVM dengan teknik SMOTE[11].Pustaka 'imblearn' digunakan untuk melakukan oversampling dengan teknik SMOTE, SMOTE di berikan nilai dengan 'random_state=42 yang berfungsi untuk memastikan hasil yang tetap atau konsisten. Setelah itu dilakukan resampling dengan SMOTE, setelah itu dilakukan pembagian dataset yaitu dibagi menjadi data latih dan data uji dengan perbandingan 80 : 20 selanjutnya melakukan Augmentasi data dengan SMOTE untuk menghindari bocornya data. Setelah itu membuat Grid Search untuk mencari parameter terbaik pada model SVM Untuk mencari parameter terbaik SVM dengan melakukan validasi silang sebanyak 5 kali. Selanjutnya menggunakan evaluasi model dengan menggunakan model SVM terbaik yang telah ditemukan dari Grid search. Setelah itu baru didapatkan matriks evaluasi seperti Accuracy, Precision, Recall, F1-support. Berikut hasil analisis sentimen dengan model SVM.



Gambar 4. Distribusi Prediksi Sentimen Model SVM

Dari Gambar 4 menunjukkan bahwa hasil prediksi analisis sentimen aplikasi Tiktok Tokopedia Seller Center dengan 921 data ulasan dan di olah menggunakan model *Support Vector Machine* dengan teknik SMOTE menghasilkan bahwa sentimen negatif (0) lebih banyak dibandingkan dengan sentimen positif (1), dihasilkan perbandingan negatif 53.2 % dan 46.8 %.

3.4.2 Convolutional Neural Network (CNN)

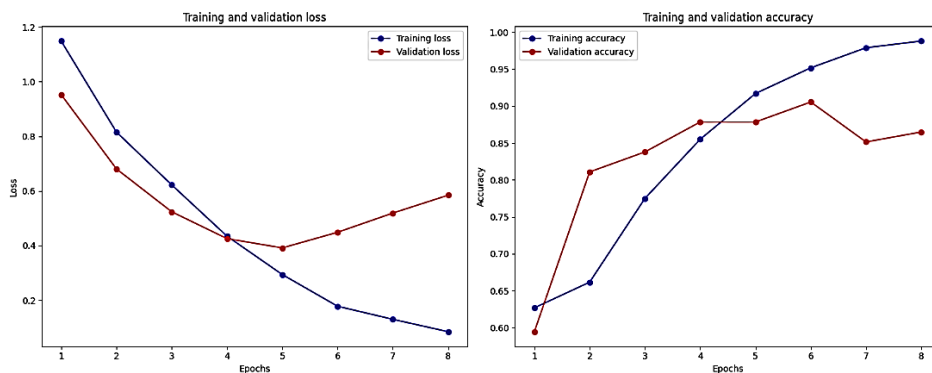
Layer (type)	Output Shape	Param #
embedding_5 (Embedding)	(None, 50, 128)	640000
conv1d_10 (conv1D)	(None, 48, 64)	24640
max_pooling1d_10 (MaxPool1d)	(None, 24, 64)	0
conv1d_11 (conv1D)	(None, 22, 128)	24704
max_pooling1d_11 (MaxPool1d)	(None, 11, 128)	0
flatten_5 (Flatten)	(None, 1408)	0
dense_10 (Dense)	(None, 32)	45088
dropout_5 (Dropout)	(None, 32)	0
dense_11 (Dense)	(None, 2)	66

Total params: 734498 (2.80 MB)
 Trainable params: 734498 (2.80 MB)
 Non-trainable params: 0 (0.00 Byte)

Gambar 5. Arsitektur CNN

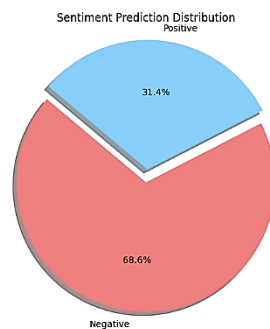
Arsitektur CNN dengan lapisan yang terbentuk dapat dilihat pada Gambar 5.

Memasukkan data untuk di evaluasi pada CNN dinamakan sebagai *input layer*. Tahapan dimulai dengan mengubah setiap kata pada data diubah menjadi vektor lalu di input ke ke lapisan *Convolutional*. *ConvID* ini dipakai dalam lapisan *Convolutional* dengan menggunakan penyaring atau filter sebanyak 64, dengan ukuran kernel 3. Setelah itu hasil dari ini akan diteruskan ke lapisan *MaxPollingID* yang memiliki tujuan untuk pengurangan beban pada komputasi, setelah itu meneruskan ke lapisan *Dropout* yang berfungsi untuk mengurangi atau menghindari *overfitting*[16].Selanjutnya meneruskan ke lapisan *Dense* yang berfungsi untuk melanjutkan atau meneruskan fungsi dari *Dropout* ke *Flatten*, dalam lapisan *Flatten* dimensi data multi array diubah menjadi vektor. Dan yang terakhir yaitu lapisan *Dense* yang berfungsi untuk melakukan prediksi akhir model[7]. Selanjutnya melakukan kompilas model dengan cara menentukan *loss function*, *optimizer*, dan *matrix* untuk digunakan evaluasi selama proses pelatihan. Proses pelatihan ini bertujuan untuk memonitoring *loss validation* dan *validation accuracy*. Berikut grafik hasil proses pelatihan



Gambar 6. Grafik Hasil Pelatihan Model CNN

Proses pelatihan menggunakan data latih X_{train_padded} , y_{train} . Untuk mencegah overfitting dan mengoptimalkan proses pelatihan dapat digunakan *callback EarlyStopping*. Dari gambar gravik diatas menunjukkan bahwa *validation loss* tertinggi saat dilakukan pelatihan urutan pertama yang memiliki *validation loss* 0.9528 namun memiliki *validation accuracy* paling redah dibawah 0.5946, dan *validation loss* terendah terletak pada pelatihan urutan ke lima dengan *validation loss* 0.3916 namun memiliki *validation akurasi* senilai 0.8784, dan *validation accuracy* tertinggi terletak pada pelatihan nomor enam sebesar 0.9054 dan *validation loss* 0.4488, setelah dilakukan *early stopping* dilakukan prediksi model untuk membuat prediksi pada data uji dengan menggunakan model yang telah dilatih sebelumnya. Setelah itu baru dilakukan analisis hasil untuk menunjukkan kinerja model melalui matriks evaluasi (*accuracy*, *precision*, *recall*, *f1-score*, dan *support*). Berikut hasil analisis sentimen dengan menggunakan model CNN.

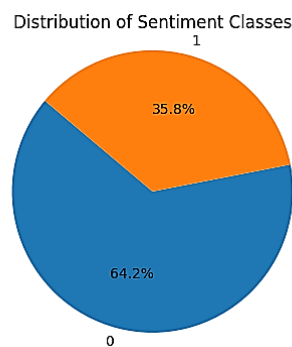


Gambar 7. Distribusi Prediksi Sentimen Model CNN

Dari gambar 7 dapat menunjukkan bahwa hasil prediksi analisis sentimen aplikasi Tiktok Tokopedia Seller Center dengan 921 data ulasan dan di olah menggunakan model *Convolutional Neural Network*, menghasilkan sentimen negatif (0) lebih besar sejumlah 68.6% dibandingkan sentimen positif (1) 31.4%.

3.4.3 Naive Bayes

Algoritma Naive Bayes ini merupakan metode klasifikasi berdasarkan teorema Bayes dengan memiliki asumsi yang kuat bahwa setiap fitur adalah independen. Pembuatan model Naive bayes diawali dengan melakukan TF-IDF lalu melakukan pembagian data yang terdiri dari data latih dan data uji dengan perbandingan 80:20 , dan dilanjutkan dengan validasi silang sebanyak 5 kali. Tujuan dari validasi silang yaitu untuk mengevaluasi kinerja lebih baik dengan memastikan tidak hanya pada satu pembagian data tetapi jika dilakukan pada variasi pembagian data yang berbeda dengan hasil yang tetap konsisten[12][13]selanjutnya menggunakan teknik SMOTE untuk mengatasi ketidak seimbangan data[14]Setelah itu dilakukan evaluasi model naive bayes yang telah di optimalkan dengan hyperparamete terbaik, hasil akhir dari kinerja dan performa model naive bayes melalui matriks evaluasi seperti *Accuracy, Precision, Recall, F1-support*



Gambar 8. Distribusi Prediksi Sentimen Model Naive Bayes

Dari gambar 9 dapat menunjukkan bahwa hasil prediksi analisis sentimen aplikasi Tiktok Tokopedia Seller Center dengan 921 data ulasan dan di olah menggunakan model Naive Bayes, menghasilkan sentimen negatif (0) lebih besar sejumlah 64.2% dibandingkan sentimen positif (1) 35.8%.

3.5. Hasil Analisis

Hasil analisis sentimen dari 3 algoritma Machine learning, menunjukkan bahwa SVM dengan teknik SMOTE mampu menghasilkan analisis yang seimbang dengan perbandingan sentimen positif 53,2% dan sentimen negatif 46,8%. Kemampuan SVM dengan teknik SMOTE untuk mengatasi masalah ketidak seimbangan data dibuktikan juga dalam penelitian yang dilakukan oleh Meishita Inelza Putri & Iqbal Khairudin[15] Selain itu diukung dengan penelitian yang dilakukan oleh Rizky Wahyudi & Gilang Kusumawardana dalam penelitiannya menjelaskan bahawa SVM merupakan algoritma yang mampu mengatasi kasus klasifikasi teks bebas linear. Dalam penelitiannya menjelaskan bahawa penggunaan metode SVM dalam penelitiannya untuk mendapatkan hasil klasifikasi teks review yang lebih baik dan akurat, terbukti dalam penelitiannya mendapatkan akurasi yang bagus sebesar 85,54% [6]Dan juga di dukung dengan penelitian yang dilakukan oleh Thalita Meisya Pertama Aulia, dkk [11],dalam penelitiannya menjelaskan bahawa algoritma SVM adalah algoritma yang optimal untuk kasus klasifikasi sentimen, hal tersebut di tunjukkan hasil penelitian dengan mendapat akurasi paling bagus yaitu 0.87.

3.6. Evaluasi

Tahapan terakhir yaitu evaluasi, dimana pada bagian ini kami akan mengevaluasi kinerja model yang telah dibangun menggunakan Confusion Matrix digunakan untuk mengukur performa

klasifikasi dari model yang telah dibuat seperti Accuracy, Precision, Recall, F1-Score. Berikut kami sajikan rumus *confusion matrix* untuk menyajikan Precision, Recall, F1-Score, dan Accuracy[4]:

$$Accuracy = \frac{(TP) + (TN)}{(TP) + (TN) + (FN) + (TN)} \quad (1)$$

$$Precision = \frac{(TP)}{(TP) + (FP)} \quad (2)$$

$$Recall = \frac{(TP)}{(TP) + (FN)} \quad (3)$$

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + precision} \quad (4)$$

3.6.1. Hasil Evaluasi SVM, CNN, dan Naive Bayes

Semua tabel dan gambar yang anda masukkan dalam dokumen harus disesuaikan dengan urutan 1 kolom atau ukuran penuh satu kertas, agar memudahkan bagi reviewer untuk mencermati makna gambar. Gambar dan tabel yang dimuat harus dirujuk dan jika memungkinkan bisa dijelaskan didalam naskah.

Tabel 5. Perbandingan Accuracy SVM, CNN, dan Naive Bayes

Algoritma	Accuracy
SVM	0.90
CNN	0.88
Naive Bayes	0.84

Dari Tabel 5 dapat dilihat bahwa akurasi tertinggi dimiliki oleh SVM dengan akurasi sebesar 0.90 ini menunjukkan SVM dapat menangani kasus analisis sentimen dengan baik. Selanjutnya diikuti oleh CNN dengan akurasi sebesar 0.88 hampir sama dengan model SVM, dan yang terakhir adalah Naive Bayes dengan akurasi 0.84

Tabel 5. Perbandingan precision, recall, f1-score dan support (0)

Algoritma	Presicion	Recall	F1-Score	Support
SVM	0.86	0.95	0.90	114
CNN	0.87	0.96	0.91	115
Naive Bayes	0.84	0.92	0.88	119

Dari Tabel 5 dapat dilihat perbandingan antara precision, recall, f1-score, dan support untuk klasifikasi sentimen negatif (0):

- 1) Naive bayes memiliki precision lebih rendah sebesar 0.84, sedangkan CNN dan SVM memiliki precision sebesar 0.87 dan 0.86 .Hal ini menunjukkan bahwa model CNN dan SVM mempunyai kinerja lebih baik dalam memberi label false positive (false negatif) dibandingkan dengan model naive bayes.
- 2) Pada recall model CNN memiliki recall yang lebih tinggi sebesar 0.96 disusul dengan model SVM 0.95 dan CNN dan Naive bayes dengan recall 0.92.Hal tersebut menunjukkan bahwa CNN dan SVM memiliki kinerja yang bagus dalam mengumpulkan kelas *True Negatif*.
- 3) Untuk f1-scroe Model SVM dan CNN memiliki nilai yang sama sebesar 0.91, sedangkan naive bayes memiliki f1-score sebesar 0.90. f1-score ini merupakan matrix yang mempunyai peran dalam menggabungkan precision dan recall, sehingga menunjukkan bahwa SVM dan CNN memiliki Keseimbangan dalam precision dan recall untuk prediksi kelas sentimen negatif (0).

Tabel 7. Perbandingan precision, recall, f1-score dan support (1)

Algoritma	Precision	Recall	F1-Score	Support
SVM	0.95	0.85	0.90	123
CNN	0.91	0.76	0.83	70
Naive Bayes	0.82	0.70	0.75	66

Dari Tabel 5 dapat dilihat perbandingan antara precision, recall, f1-score, dan support untuk klasifikasi sentimen positif (1):

- 1) SVM memiliki precision paling tinggi sebesar 0.96, diikuti dengan CNN dengan precision sebesar 0.91. Hal tersebut menunjukkan bahwa model SVM adalah model dengan kinerja terbaik untuk menghindari kesalahan false Positif, dimana prediksi positif sebenarnya negatif.
- 2) Dalam recall SVM memiliki nilai recall yang paling tinggi sebesar 0.85, disusul dengan CNN dengan recall sebesar 0.76 dan yang terakhir Naive Bayes dengan recall 70. Hal ini menunjukkan bahwa model SVM memiliki kinerja yang lebih bagus dalam menangkap kelas true positif dibandingkan CNN dan Naive Bayes
- 3) Nilai F1-score paling tinggi dimiliki oleh model SVM sebesar 0.90, CNN memiliki f1-score 0.83 dan yang terakhir adalah Naive Bayes dengan f1-score sebesar 0.75. F1-score yang memiliki nilai paling tinggi adalah SVM berarti SVM merupakan model yang memiliki keseimbangan antara precision dan recall yang lebih baik dalam prediksi kelas sentimen positif (1).

4. KESIMPULAN

Berdasarkan penelitian kasus Analisis Sentimen Aplikasi Tiktok Seller Center dengan Pendekatan Machine learning: SVM, CNN, Naive Bayes, didapatkan dataset sebanyak 921 data ulasan terbaru (2023-2024) dengan atribut content dan score. Hasil penelitian ini menunjukkan bahwa:

- 1) Hasil dari pengolahan data dengan 3 algoritma machine learning menjelaskan bahwa prediksi sentimen negatif (0) lebih tinggi daripada sentimen positif (1) pada aplikasi ini.
- 2) Model SVM memiliki kinerja paling bagus dalam prediksi sentimen negatif (0) maupun positif (1) dengan memiliki akurasi sebesar 0.90, disusul model CNN dengan akurasi sebesar 0.88, dan kemudian model Naive Bayes dengan Akurasi sebesar 0.84,

DAFTAR PUSTAKA

- [1] Wulandari, D. A., Storm, B., Djie, F., Nugroho, A. A., Raya, J. R. F., Labu, P., & Cilandak, K. (n.d.). Media Hukum Indonesia (MHI) Published by Yayasan Daarul Huda Krueng Mane Peluang Social Commerce Melalui Akuisisi Tokopedia Oleh Tiktok Shop Ditinjau Dari Aspek Legalitas. 2(3), 445.
- [2] Alfayed, E., Ramadeli, L., Agnestasia, R., Amalina, V., Octavia Swid, Z. H., & Riofita, H. (2023). ANALISIS STRATEGI PEMASARAN DAN PENJUALAN E-COMMERCE PADA TIKTOK SHOP. Jurnal Ekonomi Manajemen Dan Bisnis, Vol. 1, No.2.
- [3] Sarina, Adam M Tanniewa. (2023) Implementasi Algoritma Support Vector Learning Terhadap Analisis Sentimen Penggunaan Aplikasi Tiktok Shop Seller Center.
- [4] Satya Marga, N., Rahman Isnain, A., & Alita, D. (2021). Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA). Abstrak, 453(4), 453-463.
- [5] Cindo, M., & Rini, D. P. (2019). Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review: Metode Klasifikasi Pada Sentimen Analisis. In Januari.
- [6] Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. JURNAL INFORMATIKA, 8(2).

- [7] Y, R. A., & L, K. M. (2023). Perbandingan Algoritma Cnn Dan Svm Untuk Analisis Sentimen Mengenai Kenaikan Harga Bahan Bakar Minyak. *EProceedings of Engineering*, Vol.10, No.6.
- [8] Khatami, F. A., Irawan, B., Si, S., & Setianingsih, C. (n.d.). Analisis Sentimen Terhadap Review Aplikasi Layanan E-Commerce Menggunakan Metode *Convolutional Neural Network* Sentiment Analysis Of E-Commerce Application Reviews Using The *Convolutional Neural Network* Method
- [9] Kurnianto, E., & Febriawan, D. (2023). Analisis Sentimen Perbedaan Pendapat Netizen Indonesia Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naïve Bayes. *Jurnal Sistem Komputer Dan Informatika (JSON)* Hal: 404–, 414(2).
- [10] Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine learning. *JBASE - Journal of Business and Audit Information Systems*, 4(2).
- [11] Zamsuri, A., Asril, E., Sadar, M., & Turnandes, Y. (2023). Analisis sentimen opini vaksinasi Covid-19.... *ZONAsi: Jurnal Sistem Informasi*, 5(1).
- [12] Meisya, T., Aulia, P., Arifin, N., & Mayasari, R. (n.d.). Perbandingan Kernel *Support Vector Machine* (Svm) Dalam Penerapan Analisis Sentimen VAKSINISASI COVID-19.
- [13] Farah Zhafira, D., Rahayudi, B., & Korespondensi, P. (2021). *Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes Dan Pembobotan Tf-Idf Berdasarkan Komentar Pada Youtube* (Vol. 2, Issue 1).
- [14] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1-7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- [15] Putri, M. I., & Kharisudin, I. (2022). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia. *PRISMA, Prosiding Seminar Nasional Matematika*, 5, 759-766. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [16] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 11765-11788, 2021, doi: [10.1007/s11042-020-10183-2](https://doi.org/10.1007/s11042-020-10183-2). Institusi. Nomor standar. Judul. Tempat Publikasi. Penerbit. Tahun Publikasi.