Volume 5 No 1 Maret 2025
e-ISSN 2776-9690
DOI: https://doi.org/10.53514/jco.v5i1.631

Studi Sentimen Masyarakat Terhadap Layanan FLIX Cinemas di Twitter Dengan Pendekatan Naïve Bayes dan SVM

Ahmad Ridlan[1],
Daniel Eliazar Latumaerissa[2*],
Muhaimin Hasanudin[3],
Derisma[4]
Muhamad Fadli[3],

[1]*Information System, Institut Teknologi dan Bisnis Stikom*
[2] *Information System, universita Pembangunan Nasional Veteran Yogyakarta*
[3]*Faculty of Computer Science, Mercu Buana University*
[4]*Departemen Informatika Fakultas Teknologi Informasi Universitas Andalas, Padang, Indonesia*

*Corresponding author email: *danieleliazar.latumaerissa@upnyk.ac.id*

ABSTRAK

Penelitian ini melakukan analisis sentimen opini publik terhadap FLIX Cinemas dengan platform Twitter menggunakan metode Naïve Bayes dan Support Vector Machine (SVM). Tujuan utama dari penelitian adalah klasifikasi sentimen yang akurat - baik positif maupun negatif terkait layanan FLIX untuk memperoleh denyut respons publik terhadap layanan. Metodologi yang digunakan adalah pengumpulan twitter, praproses, pelabelan manual dan otomatis, pelabelan yang dibobot dengan TF-IDF. Klasifikasi menggunakan Naïve Bayes dan SVM dengan tiga skenario berbagi data pada 90:10, 80:20, dan 70:30. Hasilnya menyatakan bahwa SVM memiliki akurasi maksimum sebesar 92% untuk pelabelan otomatis dan Naïve Bayes sebesar 87% dalam pelabelan manual. Dari hasil penelitian ini menyatakan bahwa SVM lebih efektif untuk dataset besar sedangkan Naïve Bayes berlaku untuk data kecil sehingga dapat menjadi salah satu acuan bagi FLIX dalam upayanya meningkatkan pelayanan sesuai dengan hasil analisis sentimen.

***Kata Kunci**: Analisis Sentimen, Naïve Bayes, Support Vector Machine, FLIX Cinemas*

I. Pendahuluan

Di era digital saat ini, media sosial seperti Twitter menjadi platform yang banyak digunakan oleh masyarakat untuk menyampaikan pendapat dan opini. Salah satu topik yang sering dibahas adalah pengalaman menonton film di bioskop. FLIX Cinemas menawarkan konsep baru dalam menonton film dan berhasil mengurangi kesan monopoli di industri bioskop Indonesia. Namun, sebagai bagian dari industri jasa, FLIX perlu memahami bagaimana opini publik terhadap layanan mereka. Analisis sentimen menjadi salah satu cara untuk mengetahui tanggapan masyarakat terhadap FLIX melalui media sosial khususnya Twitter. Dengan menganalisis sentimen dapat diketahui apakah opini publik cenderung positif atau negatif terhadap layanan FLIX Cinemas.

Beberapa penelitian sebelumnya telah menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) (Azi et al., 2024)untuk analisis sentimen. Misalnya, penelitian oleh (Gunawan et al., 2024; Winanto & Budihartanti, 2022)menunjukkan bahwa Naïve Bayes mampu mencapai akurasi tinggi dalam mengklasifikasikan teks pengaduan masyarakat. Sementara itu, penelitian oleh (Akter et al., 2024) menemukan bahwa SVM memberikan performa yang baik dalam klasifikasi data dengan akurasi mencapai 90,9%. Kombinasi kedua metode ini juga telah diuji dalam penelitian lain seperti yang dilakukan oleh (Styawati et al., 2021) yang menunjukkan bahwa Naïve Bayes dapat mencapai akurasi hingga 94% dalam prediksi kelulusan mahasiswa. Dengan demikian, kedua metode ini telah terbukti efektif dalam berbagai konteks analisis sentimen.

Media sosial seperti Twitter menjadi platform yang banyak digunakan untuk menyampaikan pendapat, termasuk tanggapan masyarakat terhadap layanan bioskop. FLIX Cinemas, sebagai salah satu jaringan bioskop di Indonesia, perlu memahami opini publik untuk meningkatkan kualitas layanannya. Analisis sentimen menjadi solusi untuk mengklasifikasikan tanggapan tersebut menjadi positif atau negatif. Penelitian sebelumnya oleh (Fitri et al., 2019) menunjukkan bahwa metode Naïve Bayes efektif dalam analisis sentimen terkait kampanye sosial di Twitter, dengan akurasi mencapai 85%. Sementara itu, penelitian (Winanto & Budihartanti, 2022) menemukan bahwa Support Vector Machine (SVM) memberikan akurasi lebih tinggi, yaitu 90,9%, dalam klasifikasi sentimen terkait kebijakan publik. Kombinasi kedua metode ini juga diuji oleh (Styawati et al., 2021) yang menyimpulkan bahwa Naïve Bayes mencapai akurasi 94% dalam prediksi kelulusan mahasiswa, sedangkan SVM unggul dalam menangani data tidak seimbang. Selain itu, penelitian (Mandloi & Patel, 2020) menekankan pentingnya preprocessing data untuk mengurangi noise dalam analisis sentimen berbasis Twitter. Berdasarkan temuan tersebut, penelitian ini bertujuan menganalisis sentimen

DOI: https://doi.org/10.53514/jco.v5i1.631

masyarakat terhadap FLIX Cinemas dengan membandingkan performa Naïve Bayes dan SVM. Data diambil dari tweet yang terkait dengan FLIX kemudian dilakukan preprocessing untuk membersihkan data dari noise. Setelah itu, data diberi label positif atau negatif menggunakan kamus kata yang telah disiapkan. Proses pembobotan kata dilakukan dengan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menentukan seberapa penting suatu kata dalam dokumen. Selanjutnya, data diklasifikasikan menggunakan Naïve Bayes dan SVM dengan rasio pembagian data 90:10, 80:20, dan 70:30 untuk pelatihan dan pengujian. Hasilnya diharapkan dapat memberikan rekomendasi bagi FLIX Cinemas dalam meningkatkan layanan serta kontribusi metodologis bagi pengembangan analisis sentimen di masa depan.

II. Landasan Teori

Bagian landasan teori akan menjelaskan beberapa pengertian singkat FLIX Cinemas, TF-IDF, Algoritma Naïve Bayes dan Blacbox Testing:

A. FLIX Cinemas

FLIX Cinemas merupakan salah satu jaringan bioskop di Indonesia yang menawarkan konsep baru untuk memberikan pengalaman yang berbeda saat menonton film. FLIX Cinemas membuka jaringan bioskop pertamanya di Paris Van Java Mall, Bandung. Kehadiran FLIX Cinemas menghilangkan kesan monopoli yang terjadi dalam jaringan bisnis bioskop di Indonesia karena sebelumnya didominasi oleh Cinema 21 yang telah lebih dahulu sukses dalam pasar sinema di Indonesia.

B. TF-IDF

Untuk menilai seberapa signifikan suatu kata mencerminkan sebuah kalimat, dilakukan proses pembobotan atau perhitungan. Dalam skema TF-IDF, penilaian ini bergantung pada seberapa sering kata tersebut muncul dalam dokumen tersebut, dengan frekuensi kata menjadi dasar penilaian.

C. Naïve Bayes

Naïve Bayes Classifier (NBC) adalah metode klasifikasi yang didefinisikan berdasarkan teorema Bayes(Dwiasnati & Devianto, 2019; Fajriah & Kurniawan, 2025). Metode klasifikasi dengan menggunakan metode reliabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang masa depan berdasarkan pengalaman sebelumnya, dikenal dengan Teorema Bayes. Menurut teorema Bayes, Naïve Bayes adalah Algoritma yang menggabungkan probabilitas sebelumnya dengan probabilitas bersyarat dalam rumus yang digunakan untuk menghitung probabilitas setiap klasifikasi dalam probabilitas. Naïve Bayes Classifier dalam mengklasifikasikan terdapat dua proses penting yaitu learning (pelatihan) dan testing(Golpour et al., 2020; Hasanudin et al., 2024; Mahar et al., 2023).

$$Py = PxP(x)P(y) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Keterangan:

Y = data dengan kelas yang belum diketahui.
X = hipotesis data y merupakan suatu kelas spesifik
P(x|y) = probabilitas hipotesis x berdasarkan kondisi y.
P(x) = probabilitas hipotesis x.
P(y|x) = probabilitas y berdasarkan kondisi pada hipotesis x.
P(y) = probabilitas dari y.

D. Support Vector Machine

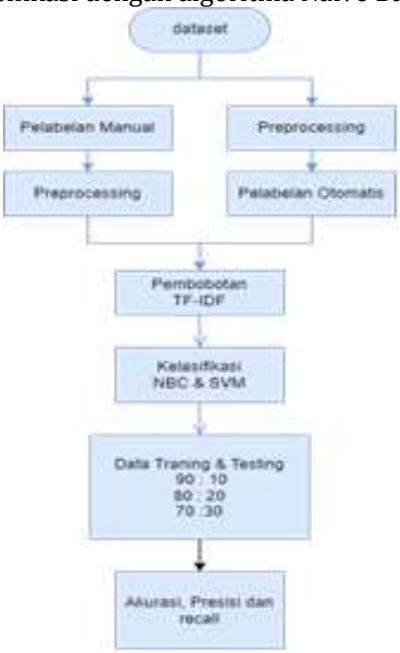
Support Vector Machine (SVM) merupakan teknik yang relatif baru untuk membuat prediksi, baik dari segi klasifikasi maupun regresi(Sheykhmousa et al., 2020; Winanto & Budihartanti, 2022). Penelitian ini diklasifikasikan ke dalam klasifikasi biner yang terdiri dari dua kelas. Kelas itu positif atau negatif. Konsep klasifikasi dengan SVM adalah mencari hyperplane terbaik yang berfungsi sebagai pemisah dua kelas data. Algoritma Support Vector Machine bekerja dengan membagi data secara optimal menggunakan pengukuran jarak hyperplane dari titik terdekat daripada mencari titik maksimum untuk memaksimalkan jarak antar label kelas berdasarkan pengukuran batas akuisisi kelas(Sheykhmousa et al., 2020; Styawati et al., 2021).

 DOI: https://doi.org/10.53514/jco.v5i1.631

III. Metode Penelitian

A. Tahapan Penelitian

Bab ini menjelaskan metodologi penelitian yang digunakan, meliputi pengumpulan data, preprocessing, pelabelan, dan klasifikasi dengan algoritma Naïve Bayes dan SVM.



Gambar 1. Tahapan Penelitian

Gambar 1 Tahapan Penelitian menggambarkan alur penelitian yang terdiri dari beberapa tahap. Dimulai dengan Pengumpulan Data dari Twitter, kemudian dilakukan Preprocessing Data untuk membersihkan dan mempersiapkan data. Selanjutnya, data diberi label (Labeling) secara manual atau otomatis. Setelah itu, dilakukan Pembobotan TF-IDF untuk menilai signifikansi kata. Tahap terakhir adalah Klasifikasi menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) dengan pembagian data training dan testing.

B. Data

Pada penelitian ini, data diperoleh dari tweet pengguna Twitter dengan menggunakan Google Colab yang terintegrasi dengan Twitter Cookies untuk mendukung proses pengumpulan data. Fokus pengambilan data adalah pada tweet yang terkait dengan topik dan studi kasus mengenai Bioskop FLIX. Kata kunci yang relevan untuk penelitian ini mencakup FLIX, Bioskop FLIX, dan Benefit FLIX dari Studi Kasus FLIX. Prosedur pengumpulan informasi dilakukan dengan mengetikkan kata kunci tersebut di dalam Google Colab, sehingga data yang diambil mencakup tweet yang berkaitan dengan topik yang tercakup dalam studi kasus. Seluruh proses pengumpulan informasi, manipulasi data, dan analisis dilakukan menggunakan Google Colab untuk memastikan konsistensi dan kemudahan akses pada seluruh dataset. Totalnya, terdapat 1239 studi kasus FLIX yang membentuk keseluruhan kumpulan data penelitian.

C. Preprocessing Data

Preprocessing data adalah membersihkan dataset dari data mentah menjadi data yang siap digunakan, memudahkan proses klasifikasi menjadi positif dan negatif(Ridlan et al., 2025). Preprocessing data merupakan teknik penambangan data yang melibatkan transformasi data mentah menjadi format yang mudah dipahami. Dalam penelitian ini, digunakan lima langkah dalam tahap preprocessing. Setelah dilakukan tahap preprocessing, data yang telah diproses menjadi 1224 untuk data studi kasus FLIX.

Pada fase preprocessing data dilakukan untuk mentransformasi data mentah dari tweet menjadi format yang siap untuk analisis. Pertama, dilakukan tokenisasi untuk memecah teks menjadi unit-unit kata atau token. Selanjutnya, semua karakter non-alfanumerik, tanda baca, serta simbol dihapus melalui proses cleansing. Kemudian, teks dinormalisasi dengan mengubah semua huruf menjadi

DOI: https://doi.org/10.53514/jco.v5i1.631

lowercase untuk memastikan konsistensi. Stop words atau kata-kata umum yang tidak memiliki nilai informasional tinggi dihilangkan menggunakan daftar stop words bahasa Indonesia. Tahap berikutnya adalah stemming, di mana kata-kata direduksi ke bentuk dasarnya menggunakan algoritma Porter Stemmer atau Nazief-Adriani untuk bahasa Indonesia. Selain itu, dilakukan penyaringan terhadap kata-kata yang terlalu pendek atau panjang serta kata-kata yang tidak bermakna. Hasil akhir dari preprocessing adalah kumpulan teks yang telah dibersihkan dan distandarisasi, siap untuk diproses lebih lanjut dalam tahap pembobotan TF-IDF atau klasifikasi. Proses ini sangat penting untuk mengurangi noise dan meningkatkan akurasi model dalam mengidentifikasi sentimen.

D. Labeling

Pada fase ini, data yang telah bersih selama proses preprocessing akan diberi label sebagai kategori data. Pelabelan kategori data dibagi menjadi dua, yaitu kategori positif dan kategori negatif(Akter et al., 2024; Mandloi & Patel, 2020; Zahoor & Rohilla, 2020). Proses pelabelan dilakukan menggunakan bahasa R melalui platform Google Colab. Proses pelabelan dengan bahasa R melibatkan penggunaan kamus bahasa Indonesia yang terhubung dengan Google Drive. Kamus bahasa Indonesia yang digunakan terdiri dari dua bagian, yakni kamus untuk kata-kata positif dan kamus untuk kata-kata negatif.

Pada tahap pelabelan, dilakukan penskoran untuk setiap kalimat. Setiap kata yang terdeteksi diberi skor untuk menilai kelas sentimen(Afiyati et al., 2020; Fitri et al., 2019; Zahoor & Rohilla, 2020). Untuk kata positif diberi nilai 1 sedangkan untuk kata negatif diberi nilai -1. Jika ada kata yang tidak ada dalam kamus positif atau negatif, maka diberikan nilai 0. Penilaian dilakukan dengan cara menghitung jumlah nilai setiap kata dalam satu kalimat. Jika nilainya >=0 maka diberi label sebagai tweet sentimen positif, sebaliknya jika bernilai <0 maka diberi label sebagai tweet sentimen negatif. Kelas data positif akan diberi nilai 1, sedangkan kelas data Negatif akan diberi nilai 0.

Tabel 1. Nilai Klasifikasi Otomatis

No	Sentimen	Positif	Negatif
1	FLIX	487	737

Dalam penelitian ini, eksperimen juga dilakukan dengan menggunakan pelabelan manual pada dataset. Label positif adalah tweet yang memiliki pujian sedangkan tweet negatif adalah kata-kata kasar dan sarkastik.

Tabel 2. Nilai Klasifikasi Manual

No	Sentimen	Positif	Negatif
1	FLIX	1035	189

IV. Hasil Dan Pembahasan

A. Hasil

Penelitian ini terbagi menjadi tiga percobaan, dengan rasio data 90%:10%, 80%:20%, dan 70%:30%, yang melibatkan penggunaan labeling otomatis dan manual. Pada percobaan pertama dengan rasio 90%:10% dan menggunakan labeling otomatis, diperoleh nilai akurasi sebesar 89% untuk algoritma SVM dan 81% untuk algoritma Naïve Bayes.

```
0.924812030075188
[[69  1]
 [ 9 54]]
              precision    recall  f1-score   support

           0       0.88      0.99      0.93        70
           1       0.98      0.86      0.92        63

    accuracy                           0.92       133
   macro avg       0.93      0.92      0.92       133
weighted avg       0.93      0.92      0.92       133
```

Gambar 4. SVM 90:10 Label Otomatis

```
0.849624060150376
[[64  6]
 [14 49]]
              precision    recall  f1-score   support

           0       0.82      0.91      0.86        70
           1       0.89      0.78      0.83        63

    accuracy                           0.85       133
   macro avg       0.86      0.85      0.85       133
weighted avg       0.85      0.85      0.85       133
```

Gambar 5. Naïve Bayes 90:10 Label Otomatis

 DOI: https://doi.org/10.53514/jco.v5i1.631

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 80% : 20% bahwa studi kasus FLIX memiliki nilai akurasi 84% untuk algoritma SVM dan 83% untuk algoritma Naïve Bayes.

0.9132075471698113
[[146 4]
[19 96]]

	precision	recall	f1-score	support
0	0.88	0.97	0.93	150
1	0.96	0.83	0.89	115
accuracy			0.91	265
macro avg	0.92	0.90	0.91	265
weighted avg	0.92	0.91	0.91	265

Gambar 6. SVM 80:20 Label Otomatis

0.8716981132075472
[[142 8]
[26 89]]

	precision	recall	f1-score	support
0	0.85	0.95	0.89	150
1	0.92	0.77	0.84	115
accuracy			0.87	265
macro avg	0.88	0.86	0.87	265
weighted avg	0.88	0.87	0.87	265

Gambar 7. Naïve Bayes 80:20 Label Otomatis

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 70% : 30% bahwa untuk studi kasus Indihome memiliki nilai akurasi 83% untuk algoritma SVM dan 83% untuk algoritma Naïve Bayes.

0.8816120906801007
[[205 12]
[35 145]]

	precision	recall	f1-score	support
0	0.85	0.94	0.90	217
1	0.92	0.81	0.86	180
accuracy			0.88	397
macro avg	0.89	0.88	0.88	397
weighted avg	0.89	0.88	0.88	397

Gambar 8. SVM 70:30 Labeling Otomatis

0.853904282115869
[[203 14]
[44 136]]

	precision	recall	f1-score	support
0	0.82	0.94	0.88	217
1	0.91	0.76	0.82	180
accuracy			0.85	397
macro avg	0.86	0.85	0.85	397
weighted avg	0.86	0.85	0.85	397

Gambar 9. Naïve Bayes 70:30 Labeling Otomatis

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 90% : 10% bahwa studi kasus FLIX memiliki nilai akurasi 77% untuk algoritma SVM dan 76% untuk algoritma Naïve Bayes.

0.8288288288288288
[[2 19]
[0 90]]

	precision	recall	f1-score	support
0	1.00	0.10	0.17	21
1	0.83	1.00	0.90	90
accuracy			0.83	111
macro avg	0.91	0.55	0.54	111
weighted avg	0.86	0.83	0.77	111

Gambar 10. SVM 90:10 Label Manual

0.8738738738738738
[[3 10]
[4 94]]

	precision	recall	f1-score	support
0	0.43	0.23	0.30	13
1	0.90	0.96	0.93	98
accuracy			0.87	111
macro avg	0.67	0.59	0.62	111
weighted avg	0.85	0.87	0.86	111

Gambar 11. Naïve Bayes 90:10 Label Manual

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 80% : 20% bahwa studi kasus FLIX memiliki nilai akurasi 77% untuk algoritma SVM dan 76% untuk algoritma Naïve Bayes.

0.8099547511312217
[[2 41]
[1 177]]

	precision	recall	f1-score	support
0	0.67	0.05	0.09	43
1	0.81	0.99	0.89	178
accuracy			0.81	221
macro avg	0.74	0.52	0.49	221
weighted avg	0.78	0.81	0.74	221

Gambar 12. SVM 80:20 Label Manual

0.8461538461538461
[[7 30]
[4 180]]

	precision	recall	f1-score	support
0	0.64	0.19	0.29	37
1	0.86	0.98	0.91	184
accuracy			0.85	221
macro avg	0.75	0.58	0.60	221
weighted avg	0.82	0.85	0.81	221

Gambar 13. Naïve Bayes 80:20 Label Manual

 DOI: https://doi.org/10.53514/jco.v5i1.631

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 70% : 30% bahwa studi kasus FLIX memiliki nilai akurasi 70% untuk algoritma SVM dan 70% untuk algoritma Naïve Bayes.

0.8403614457831325
[[3 53]
[0 276]]

	precision	recall	f1-score	support
0	1.00	0.05	0.10	56
1	0.84	1.00	0.91	276
accuracy			0.84	332
macro avg	0.92	0.53	0.51	332
weighted avg	0.87	0.84	0.78	332

Gambar 14. SVM 70:30 Label Manual

0.8554216867469879
[[10 40]
[8 274]]

	precision	recall	f1-score	support
0	0.56	0.20	0.29	50
1	0.87	0.97	0.92	282
accuracy			0.86	332
macro avg	0.71	0.59	0.61	332
weighted avg	0.82	0.86	0.83	332

Gambar 15. Naïve Bayes 70:30 Label Manual

Pembagian rasio data dalam eksperimen ini (90%:10%, 80%:20%, dan 70%:30%) dilakukan untuk mengevaluasi ketahanan model Naïve Bayes dan SVM terhadap variasi ukuran data pelatihan dan pengujian. Rasio 90%:10% menunjukkan performa terbaik untuk SVM dengan akurasi 89% karena jumlah data pelatihan yang besar memungkinkan model mempelajari pola secara lebih mendalam. Namun, rasio ini berisiko overfitting jika data pelatihan terlalu dominan. Pada rasio 80%:20%, akurasi SVM sedikit menurun menjadi 84%, tetapi tetap lebih tinggi daripada Naïve Bayes (83%) menunjukkan bahwa SVM lebih stabil dalam menangani distribusi data yang lebih seimbang. Sementara itu, rasio 70%:30% menghasilkan akurasi yang sama (83%) untuk kedua model, mengindikasikan bahwa dengan peningkatan data pengujian, perbedaan performa antara SVM dan Naïve Bayes menjadi kurang jelas. Hal ini menegaskan bahwa SVM lebih unggul dalam skenario data pelatihan besar, sedangkan Naïve Bayes cenderung konsisten pada data yang lebih kecil. Ketidakseimbangan data terutama dominasi sentimen positif dalam pelabelan manual memengaruhi hasil, di mana model lebih bias terhadap kelas mayoritas. Penelitian ini melibatkan ekstraksi fitur menggunakan TF-IDF untuk mengukur bobot kata, diikuti klasifikasi dengan Naïve Bayes yang mengandalkan probabilitas bersyarat dan SVM yang memaksimalkan margin pemisah antar kelas. Hasil ini memperkuat temuan bahwa pilihan rasio data dan metode pelabelan secara langsung memengaruhi akurasi klasifikasi sentimen.

B. Pembahasan

Dari hasil eksperimen terlihat bahwa SVM cenderung memberikan akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes dalam berbagai skenario pembagian data terutama ketika menggunakan pelabelan otomatis. Pada skenario pembagian data 90:10, SVM mencapai akurasi sebesar 89% sedangkan Naïve Bayes hanya mencapai 81%. Hal ini menunjukkan bahwa SVM lebih efektif dalam menangani data dengan jumlah yang relatif besar (Hidayat et al., 2022). Namun, ketika menggunakan pelabelan manual, perbedaan akurasi antara kedua metode tidak terlalu besar. Misalnya, pada skenario 70:30, baik SVM maupun Naïve Bayes mencapai akurasi sekitar 83%. Selain itu, penggunaan pelabelan manual juga memberikan hasil yang cukup baik dengan akurasi tertinggi mencapai 87% untuk Naïve Bayes pada pembagian data 90:10.

Salah satu permasalahan yang muncul dalam penelitian ini adalah ketergantungan pada kualitas data yang digunakan. Data yang diambil dari Twitter seringkali mengandung noise seperti kata-kata tidak baku, singkatan, bahasa gaul sehingga memengaruhi hasil klasifikasi (Fitri et al., 2019). Meskipun telah dilakukan preprocessing untuk membersihkan data tetap ada beberapa data tidak terklasifikasi dengan benar. Selain itu, pelabelan manual yang dilakukan untuk memastikan keakuratan data memerlukan waktu dan usaha yang tidak sedikit. Permasalahan lain adalah ketidakseimbangan jumlah data antara sentimen positif dan negatif. Dalam beberapa kasus, jumlah tweet dengan sentimen positif jauh lebih banyak dibandingkan dengan yang negatif sehingga model cenderung lebih bias terhadap sentimen positif. Hal ini dapat mengurangi kemampuan model dalam mengidentifikasi sentimen negatif dengan akurat (Abdullah et al., 2022).

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode SVM lebih efektif dalam mengklasifikasikan sentimen opini publik terhadap FLIX Cinemas di Twitter dibandingkan dengan Naïve Bayes. Namun, Naïve Bayes juga menunjukkan performa yang cukup baik terutama pada pembagian data yang lebih kecil. Hasil penelitian ini dapat menjadi referensi bagi FLIX Cinemas untuk

meningkatkan layanan mereka berdasarkan tanggapan masyarakat di media sosial. Selain itu, penelitian ini juga memberikan gambaran tentang efektivitas metode Naïve Bayes dan SVM dalam analisis sentimen yang dapat digunakan dalam penelitian serupa di masa depan. Secara keseluruhan, penelitian ini berhasil menunjukkan bahwa metode SVM dan Naïve Bayes dapat digunakan untuk analisis sentimen di media sosial khususnya Twitter. Namun, perlu dilakukan perbaikan dalam hal kualitas data dan penanganan ketidakseimbangan data untuk meningkatkan akurasi dan keandalan model. Selain itu, penggunaan kamus kata yang lebih lengkap dan sesuai dengan konteks bahasa sehari-hari juga dapat membantu meningkatkan performa klasifikasi (Veny et al., 2019).

V. Kesimpulan

Penelitian ini menganalisis sentimen masyarakat terhadap FLIX Cinemas di Twitter menggunakan metode Naïve Bayes dan Support Vector Machine (SVM). Hasil menunjukkan bahwa SVM memberikan akurasi lebih tinggi terutama dengan pelabelan otomatis mencapai 89% pada skenario 90:10. Namun, saat menggunakan pelabelan manual, perbedaan akurasi antara kedua metode tidak terlalu besar dengan kedua metode mencapai sekitar 70% pada skenario 70:30. Penelitian mengidentifikasi masalah seperti ketergantungan pada kualitas data dan ketidakseimbangan data. Tujuan penelitian memberikan informasi tentang opini publik terhadap FLIX Cinemas dan mengevaluasi efektivitas metode klasifikasi yang digunakan. Hasil diharapkan menjadi masukan bagi FLIX untuk meningkatkan pelayanan. Untuk penelitian selanjutnya, disarankan meningkatkan kualitas data, memperbaiki ketidakseimbangan data dan menggunakan kamus kata yang lebih lengkap. Selain itu, eksplorasi metode klasifikasi lain dan teknik preprocessing yang lebih baik dapat dipertimbangkan untuk meningkatkan akurasi.

Daftar Rujukan

Afiyati, A., Azhari, A., Sari, A. K., & Karim, A. (2020). Challenges of Sarcasm Detection for Social Network : A Literature Review. *JUITA: Jurnal Informatika*, *8*(2), 169. https://doi.org/10.30595/juita.v8i2.8709

Akter, M. L., Moon, I. T., Akter, M. A., & Shafa, Z. T. (2024). Classification of Malicious Financial Applications Using Naive Bayes and K-Nearest Neighbor Algorithms Based on Android Permission. *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering, ICAEEE 2024*, *Icaeee*, 1–6. https://doi.org/10.1109/ICAEEE62219.2024.10561780

Azi, A., Saleh, R. F., & Ardana, W. M. (2024). *Comparative Study of Iconnet Jabodetabek and Banten Using Linear Regression Regression and Support Vector Regression*. *6*(3), 119–127.

Dwiasnati, S., & Devianto, Y. (2019). Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms. *Journal of Physics: Conference Series*, *1179*(1). https://doi.org/10.1088/1742-6596/1179/1/012023

Fajriah, R., & Kurniawan, D. (2025). *Optimalisasi Model Klasifikasi Naive Bayes dan Support Vector Machine Dengan Fast Text dan Chi Square*. *17*(4), 334–345. https://doi.org/10.30998/faktorexacta.v17i4.24751

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, *161*, 765–772. https://doi.org/10.1016/j.procs.2019.11.181

Golpour, P., Ghayour-Mobarhan, M., Saki, A., Esmaily, H., Taghipour, A., Tajfard, M., Ghazizadeh, H., Moohebati, M., & Ferns, G. A. (2020). Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography. *International Journal of Environmental Research and Public Health*, *17*(18), 1–9. https://doi.org/10.3390/ijerph17186449

Gunawan, W., Hasanudin, M., & Ridwan, W. (2024). *Pembelajaran Algoritma Naïve Bayes dan Apriori untuk Siswa SMK Media Informatika Jakarta*. *3*(1), 90–98.

Hasanudin, M., Devianto, Y., Gunawan, W., Dwiasnati, S., & Prihandi, I. (2024). *Isometric Contraction ankle joint in Cerebral Palsy using Naive Bayes*. *12*(3), 78–81.

Mahar, N. M. A., Vihi Atina, & Nugroho Arif Sudibyo. (2023). Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes Di Uniba. *Jurnal Manajemen Informatika Dan Sistem Informasi*, *6*(2), 148–158. https://doi.org/10.36595/misi.v6i2.875

Mandloi, L., & Patel, R. (2020). Twitter sentiments analysis using machine learninig methods. *2020 International Conference for Emerging Technology, INCET 2020*, 1–5. https://doi.org/10.1109/INCET49848.2020.9154183

Ridlan, A., Hasanudin, M., Cizela, O., & Latumaerissa, D. E. (2025). *Early Autism Disorder Prediction Using Machine Learning*. *13*(1).

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Member, S. (2020).

Support Vector Machine Versus Random Forest for Remote Sensing Image Classification : A Meta-Analysis and Systematic Review. *13*, 6308–6325.

Styawati, S., Isnain, A. R., Hendrastuty, N., & Andraini, L. (2021). Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(1), 56–60. https://doi.org/10.30591/jpit.v6i1.3245

Winanto, A., & Budihartanti, C. (2022). Comparison of the Accuracy of Sentiment Analysis on the Twitter of the DKI Jakarta Provincial Government during the COVID-19 Vaccine Time. *Journal of Computer Science and Engineering (JCSE)*, *3*(1), 14–27. https://doi.org/10.36596/jcse.v3i1.249

Zahoor, S., & Rohilla, R. (2020). Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study. *Proceedings - 2020 International Conference on Advances in Computing, Communication and Materials, ICACCM 2020*, 194–199. https://doi.org/10.1109/ICACCM50413.2020.9213011