ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

PENERAPAN ALGORITMA *NAÏVE BAYES CLASSIFIER* MENGENAI SENTIMEN TERHADAP APLIKASI STREAMING FILM ONLINE NETFLIX DI TWITTER

Agung Sanjaya[1*], Cindi Wulandari[2], Bunga Intan[3], Nolan Efranda[4]
[1,2,3]Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau
[4]Program Studi Informatika, Universitas Bina Insan, Lubuklinggau

e-mail: *[1]2002030015@mhs.unibninainsan.ac.id, [2]cindiwulandari@univbinainsan.ac.id, [3]bungaintan@univbinainsan.ac.id, [4]nolanefranda@univbinainsan.ac.id

Abstrak

Menonton merupakan salah satu hiburan yang paling digemari oleh masyarakat. Di era digital yang cepat dan fleksibel ini, banyak orang yang beralih dari televisi ke layanan TV digital. Pesatnya perkembangan internet di dunia menyebabkan banyak bisnis dan layanan beralih dari mode offline ke online, salah satunya adalah layanan streaming. Salah satu penyedia layanan media streaming untuk menonton online adalah Netflix, kehadiran Netflix merupakan berkah bagi para pecinta filmDalam penelitian ini penulis menggunakan metode kuantitatif, yaitu Proses pencarian pengetahuan dengan menggunakan data yang dikumpulkan sebagai alat untuk menganalisis informasi tentang apa yang ingin kita ketahui. Dalam penelitian ini, penulis menggunakan data yang diperoleh dari jejaring sosial Twitter untuk mengetahui opini masyarakat terhadap aplikasi streaming film Netflix. Data penelitian diperoleh melalui pengumpulan menggunakan google collab, dengan total 2520 sentimen dari tahun 2023 hingga desember 2023. Hasil labelling menunjukkan untuk sentimen positif sebanyak 885, sedangkan untuk sentimen negatif sebanyak 773 sehingga total data mentah sebanyak 1608. Hasil klasifikasi dengan menggunakan 10 *k-folds cross validation* pada algoritma *naïve bayes* menghasilkan algoritma *naïve bayes* didapatkan hasil *accuracy* sebesar 75,25%. *Precision* positif 77.58%, *Precision* negatif 72,41% dan *Recall* positif 77,40%, dan *Recall* negatif 72,61%.

Kata kunci : Netflix, Twitter, Analisis Sentimen, *Naïve Bayes*.

Abstract

Watching is one of the most popular entertainment among people. In this fast and flexible digital era, many people are switching from television to digital TV services. The rapid development of the internet in the world has caused many businesses and services to switch from offline to online mode, one of which is streaming services. One provider of streaming media services for watching online is Netflix, the presence of Netflix is a blessing for film lovers. In this research the author uses a quantitative method, namely the process of searching for knowledge using the data collected as a tool to analyze information about what we want to know. In this research, the author uses data obtained from the social network Twitter to determine public opinion regarding the Netflix film streaming application. Research data was obtained through collection using Google Collab, with a total of 2520 sentiments from 2023 to December 2023. The labeling results show that there are 885 positive sentiments, while there are 773 negative sentiments, so the total raw data is 1608. Classification results using 10 k-folds cross validation on the naïve Bayes algorithm resulted in the naïve Bayes algorithm obtaining an accuracy of 75.25%. Positive Precision 77.58%, negative Precision 72.41% and positive Recall 77.40%, and negative Recall 72.61%.

Keywords : *Netflix, Twitter, Sentiment Analysis, Naïve Bayes*

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

I. PENDAHULUAN

Menonton merupakan salah satu hiburan yang paling digemari oleh masyarakat, Pesatnya perkembangan internet di dunia menyebabkan banyak bisnis dan layanan beralih dari mode *offline* ke *online*, salah satunya adalah layanan streaming. Salah satu penyedia layanan media streaming untuk menonton online adalah Netflix. Banyak cara untuk memanfaatkan perkembangan teknologi ini salah satunya dengan menggunakan komputer sebagai media input, pengolahan, penyimpanan, dan penyajian data ataupun informasi kepada pengguna(Irvai & Efranda, 2022). Salah satu contohnya adalah data text yang diambil dari twitter. Twitter adalah layanan jejaring sosial dan microblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks(Ratnawati, 2018). Opini adalah pendapat, ide atau pikiran untuk menjelaskan kecenderungan atau preferensi tertentu terhadap perspektif dan ideologi akan tetapi bersifat tidak objektif karena belum mendapatkan pemastian atau pengujian, Hal tersebut menjadikannya memiliki berbagai macam topik yang informasi bisa digali kembali(Ratnawati, 2018).

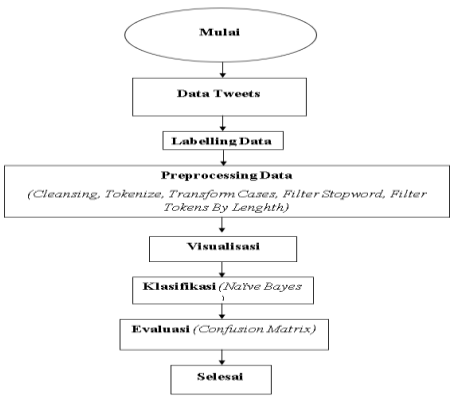
Penelitian terdahulu yang dilakukan Ratnawati, Fajar. Dengan judul Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter Berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90 % dengan rincian nilai precission 92%, recall 90% dan f-measure 90%(Ratnawati, 2018). Husna, Rizqa El Dengan Judul Analisis Klasifikasi Sentimen Pada Twitter Mengenai Netflix Yang Diblokir Oleh Telkom Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Jurnal Ilmiah menghasilkan akurasi klasifikasi yang didapat yaitu sebesar 75.47%. Sedangkan metode Support Vector Machine menggunakan Kernel Radial Basic Function (RBF) diperoleh akurasi sebesar 85.92%. Berdasarkan nilai akurasi, metode Support Vector Machine menggunakan Kernel RBF memiliki performa lebih baik dalam mengklasifikasi data sentimen(Husna, 2020).

Analisis yang digunakan dalam penelitian ini adalah analisis sentimen analisis sentimen adalah riset komputasional dari opini, sentimen dan emosi yang diekpresikan secara tekstual Untuk melakukan analisis sentimen ada beberapa algoritme yang dapat digunakan salah satunya adalah algoritme Naive Bayes. Naive Bayes classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya(Ratnawati, 2018). Sejak beroperasi di Indonesia pada tahun 2016 sampai Januari 2020 pengguna Netflix sudah mencapai 900 juta pelanggan, Netflix adalah layanan streaming yang menawarkan berbagai acara televisi, film, film dokumenter, dan anime yang diakses dengan perangkat yang terhubung ke internet. Pengguna Netflix bisa menonton sepuasnya, kapan pun, dimana pun, dengan media apa pun dengan biaya langganan per bulan(Husna, 2020). Pendapat dari setiap opini netflix di twitter dapat berpengaruh pada pengguna baru sebagai bahan pertimbangan penggunaan aplikasi, dan juga peneliti berharap dapat memberikan manfaat yang signifikasi bagi perkembangan aplikasi netflix.

Adapun tujuan umum dari adanya penelitian ini yaitu untuk melakukan analisa sentimen positif dan negatif pada aplikasi netflix menggunakan algoritma *Naive Bayes* dan
.

II. METODOLOGI PENELITIAN

Metode yang digunakan oleh peneliti ini ditampilkan pada gambar 1. Dimulai dari pengumpulan data, labelling data, preprocessing data, visualisai, klasifikasi algoritma *naive bayes*, evaluasi *confusion matrix*.



Gambar 1. Alur Penelitian

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

a. Data Tweets

Pengumpulan data dilakukan menggunakan pemrograman phyton/google collab.

b. Labelling Data

Dimana proses yang dilakukan dengan membaca satu persatu dari masing-masing ulasan dengan melihat kata-kata yang mengandung emosi akan dilabeli sentimen negatif begitu sebaliknya jika terdapat kata-kata pujian maka akan dilabeli sentimen positifapakah positif, negatif atau netral.

c. Preprocessing Data

Tahap *preprocessing* adalah mengubah data yang berhasil dikumpulkan menjadi format yang diperlukan. Adapun langakah-langkah tahap *preprocessing* yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. *Cleansing*

   proses *cleansing* yang digunakan pada beberapa operator untuk pengahapusan simbol, karakter, angka. Pada penghapusan simbol ini operator yang digunakan yaitu operator replace.

2. *Tokenize*

   Proses ini yaitu memecahkan dan pemotongan kata pada dokumen menjadi term-term berdasarkan spasi.

3. *Transform Cases*

   Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses tokenize(Saputra & Pribadi, 2023).

4. *Filter Stopword*

   Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan(Saputra & Pribadi, 2023).

5. *Filter Tokens By Lenght*

   Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti(Saputra & Pribadi, 2023).

d. Visualisasi

Visualiasi adalah cara untuk menyajikan informasi dan pola data melaluli grafik, diagram, atau peta, karna dapat memudahkan dalam memahami data dengan cepat dan lebih efektif

e. Klasifikasi

Klasifikasi merupakan suatu teknik yang dapat digunakan untuk memetakan masukan menjadi keluaran diskrit yang dinamakan dengan label dan atribut. Klasifikasi adalah teknik yang paling penting yang dipakai dalam data mining. pengkelompokan data mining adalah metode pengklasisfikasi yang mempelajari data latih untuk diklasifikasin Adapun beberapa algoritma klasifikasi, antara lain Bayesian Classification, K-Nearest Neighbor, Decision Tree Induction, Case-BasedReasoning, Genetic Algorithms, dan Support Vector Machine(Thaniket et al., 2019).

F. Evaluasi

Dalam penelitian ini metode pengujian dilakukan menggunakan library yang terdapat pada metode algoritma Naive Bayes yaitu confusion matrix 4 (empat) istilah yang mempresentasikan hasil dari proses klasifikasi.

1) True Positive (TP) Berarti seberapa banyak data yang aktual kelasnya positif, dan model juga memprediksi positif.

2) False Positive (FN) Berarti seberapa banyak data yang aktual kelasnya negatif, namun model memprediksi positif.

3) True Negative (TN) Berarti seberapa banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.

4) False Negative (FN) Berarti seberapa banyak data yang aktual kelasnya positif, namun model memprediksi negatif.

Akurasi, *precision*, dan*recall* dapat dihitung menggunakan informasi yang diberikan dalam confusion matrix.

1) Akurasi

Akurasi adalah jumlah proporsi prediksi yang benar. Akurasi digunakan sebagai tingkat ketepatan antara nilai actual dengan nilai prediksi. Adapun rumus penghitungan akurasi dapat dilihat pada persaman berikut.

$$ACCURACY = \frac{TP}{TP + TN + FP + FN}$$

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

2) Presisi

Precision adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks yang terpilih oleh sistem. Precision digunakan sebagai tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem Rumus precision dapat dilihat pada persamaan berikut.

$$PRECISION = \frac{TP}{TP + FP}$$

3) Recall

Recall adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks relevan yang ada pada koleksi. Nilai recall untuk kelas positif disebut juga dengan nilai sensitivity, sedangkan untuk kelas negatif disebut dengan nilai specificity. Recall digunakan sebagai ukuran keberhasialan sistem dalam menemukan kembali informasi recall dapat dilihat pada persamaan berikut.

$$RECALL = \frac{TP}{TP + FN}$$

III. HASIL DAN PEMBAHASAN

A.Hasil

1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan pemrograman phyton. Langkah pertama dalam pengumpulan data adalah menginstall *phyton package* dan Node.Js di *google colab*. Pada penelitian ini data yang dikumpulkan berjumlah 2520 data. Berikut gambar pengambilan data.

```
[ ]  #@title Twitter Auth Token

     twitter_auth_token = 'replace_with_your_twitter_auth_token_here'
```

Gambar 2. *Source Code Bagian 1*

Gambar 3. *Source Code Bagian 2*

Gambar 4. *Source Code Bagian 3*

```
[ ]  import pandas as pd

     # Specify the path to your CSV file
     file_path = f"tweets-data/{filename}"

     # Read the CSV file into a pandas DataFrame
     df = pd.read_csv(file_path, delimiter=";")

     # Display the DataFrame
     display(df)
```

Gambar 5. *Source Code Bagian 4*

```
[ ]  # Cek jumlah data yang didapatkan

     num_tweets = len(df)
     print(f"Jumlah tweet dalam dataframe adalah {num_tweets}.")
```

Gambar 6. *Source Code Bagian 5*

2. Labelling

Dimana proses yang dilakukan dengan membaca satu persatu dari masing-masing ulasan dengan melihat kata-kata yang mengandung emosi akan dilabeli sentimen negatif begitu sebaliknya jika terdapat kata-kata pujian maka akan dilabeli sentimen positif

Tabel 1. Komposisi Labelling Data

LABEL	JUMLAH
Positif	885
Negatif	723
	1.608

3. Preprocessing Data

Tahap P*reprocessing* adalah mengubah data yang berhasil dikumpulkan menjadi format yang diperlukan, dengan menggunakan beberapa *tools operator rapidminer.* Adapun langkah dari tahapan saat dilakukan *preprocessing* yang terdiri dari :

1. *Cleansing*
2. *Tokenize*
3. *Transform Cases*
4. *Filter Stopword*
5. *Filter Tokens By Length*

Gambar 7. *Operators Replace*

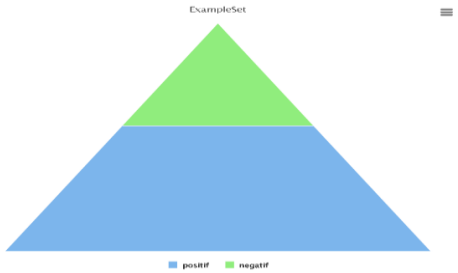
ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Gambar 8. *Cleansing, Tokenize, Transform Cases, Filter Stopword, Filter Tokens By Length*

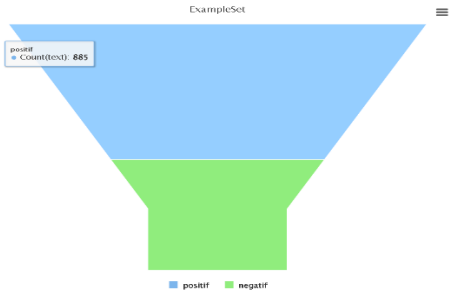
Tahapan *preprocessing* telah dilakukan bertujuan untuk menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data / menyeleksi data text agar menjadi lebih terstruktur lagi dengan melalui rangkai tahapan pada gambar 7.

4.Visualisasi

Visualisasi yang digunakan adalah visualisasi Pyramid dan Funnel



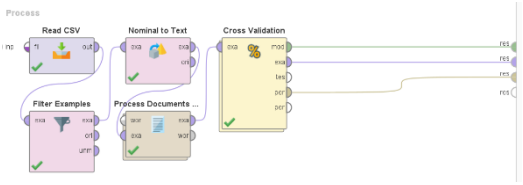
Gambar 9. Diagram Pie



Gambar 10. Diagram Funnel

5. Klasifikasi

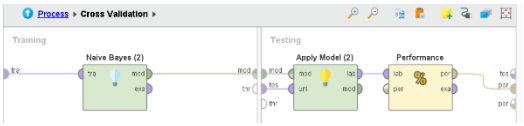
Dalam tahap pembuatan model ini, membuat model dengan klasifikasi dataset untuk sentimen yang telah melalui *preprocessing*. Dalam tahap penelitian dua algoritma ini didukung oleh *tools rapidminer*. Terdapat beberapa operator diantaranya "*read csv*" yang berfungsi menampung data, lalu dihubungkan dengan operator "*filter example*" yang berfungsi menyaring atau memfilter baris tertentu dalam dataset, kemudian dihubungkan dengan operator "*Nominal to Text*" untuk mengubah nominal ke text, lalu dihubungkan ke "*Process Document from Data*" yang didalamnya terdapat *subproses preprocessing* seperti *tokenize, transfrom case, filter stopword dan filter tokens by lenght.* lalu dihubungkan ke operator *cross validation*, Pada pengujian kedua menggunakan k-fold cross validation berguna untuk menghitung kinerja proses suatu algoritma yang membagi dataset jadi beberapa k-buah secara acak serta mengelompokkan data itu sebanyak k-fold. Pada penelitian ini memakai metode 10-fold cross validation. Cara metode 10-fold cross validation bekerja yakni membagi data jadi 10-fold model yang berukuran sama sehingga memiliki 10 subset data yang masing-masing digunakan 9 subset data latih dan 1 subset untuk data uji dengan 10 kali iterasi(Furqan et al., 2022).



Gambar 11. Model Klasifikasi

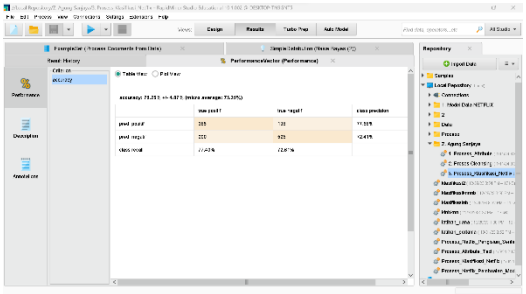
6. Evaluasi

Tahap ini dilakukan untuk menghitung evaluasi dari model naive bayes dengan *cross validation* yang didalamnya dihubungkan dengan *apply model* dan *performance* untuk dihitung terhadap hasil akurasi dari algoritma *naïve bayes.*



Gambar 12. Model Evaluasi

Setelah melalui proses validasi menggunakan algoritma *naïve bayes*, proses evaluasi akan menentukan akurasi untuk model tersebut. Berikut dibawah ini hasil dari akurasi model *naïve bayes.*



Gambar 13. Hasil Evaluasi

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Hasil klasifikasi terhadap model algoritma *naïve bayes* didapatkan hasil *accuracy* sebesar 75,25%. *Precision positif* 77.58%, *precision negatif* 72,41% dan *recall positif* 77,40%, dan *recall negatif 72*,61%.

2. Pembahasan

Masalah yang dibahas pada penelitian ini adalah bagaimana menganalisis sentimen pengguna twitter terhadap streaming film online aplikasi netflix menggunakan algoritma *naïve bayes*.

Langkah awal yaitu pengumpaulan data penelitian menggunakan *google collab*. Data yang berhasil dikumpulkan dari tahun 2023 sampai dengan bulan Desember 2023. Data penelitian berhasil didapatkan berjumlah 2520 Sentimen yang telah diberi oleh pengguna twitter. Proses selanjutnya yaitu *labelling* data. *Labelling* data dilakukan secara manual, dengan membaca isi sentimen satu persatu apakah sentimen tersebut positif, maupun negatif maupun netral. Hasil dari labelling data yaitu untuk sentimen positif 883, untuk sentimen negatif 723.

Yang kemudian dilakukan *preprocessing* pada data yang terah berhasil didapatkan dari 2520 data menjadi 1608 Setelah dilakukan *preprocessing* maka proses selanjutnya yaitu klasifikasi dan evaluasi menggunakan metode *naïve bayes* Klasifikasi diawali dengam membuat model klasifikasi pada operator-operator yang terdapat di *tools rapidminer* dimulai dari operator *read csv, filter example, Nominal to Text, Process Documment From Data, Cross Validation* dengan menggunakan 10 k-fold. Kemudian melakuan model evaluasi dengan menerapkan algoritma *naïve bayes.*

Pada hasil klasifikasi terhadap model algoritma *naïve bayes* didapatkan *accuracy* sebesar 75,25%. *Precision positif* 77.58%, *precision negatif* 72,41% dan *recall positif* 77,40%, dan *recall negatif 72*,61%.

Dari hasil tersebut dapat diketahui bahwa dengan algoritma *naïve bayes* menghasil *accuracy* sebesar 75,25%.

IV. KESIMPULAN

Adapun kesimpulan dari penelitian yang dilakukan mengenai analisis sentimen terhadap aplikasi streaming film online netflix menggunakan algoritman naïve adalah sebagai berikut :

1. Data penelitian diperoleh melalui pengumpulan menggunakan google collab, dengan total 2520 sentimen dari tahun 2023 hingga desember 2023. Hasil labelling menunjukkan untuk sentimen positif sebanyak 885, sedangkan untuk sentimen negaitif sebanyak 773.
2. Yang kemudian dilakukan *preprocessing* pada data yang terah berhasil didapatkan dari 2520 data menjadi 1608.
3. Hasil klasifikasi dengan menggunakan 10 *k-folds cross validation* pada algoritma naïve bayes menghasilkan algoritma *naïve bayes* didapatkan *accuracy* sebesar 75,25%. *Precision positif* 77.58%, *precision negatif* 72,41% dan *recall positif* 77,40%, dan *recall negatif 72*,61%.Jadi hasil penilaian akurasi cukup baik.

V. SARAN

Untuk meningkatkan hasil pengembangan pada penelitian ini selanjutnya, ada beberapa saran yang mungkin bisa dilakukan yaitu :

1. Memilih model klasifikasi atau algoritma yang berbeda
2. memilih topik yang berbeda untuk penelitian selanjutnya
3. menggunakan platform yang berbeda pada penelitian berikutnya

VI. DAFTAR PUSTAKA

Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, *21*(1), 51–60. https://doi.org/10.33633/tc.v21i1.5446

Husna, R. El. (2020). *Analisis Klasifikasi Sentimen Pada Twitter Mengenai Netflix Yang Diblokir Oleh Telkom Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Jurnal Ilmiah*. http://repository.unimus.ac.id

Irvai, M., & Efranda, N. (2022). Implementasi

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Algoritma SHA512 Pada Keamanan Berita Acara Hasil Sidang Berbasis Web. *Journal of Information Technology …*, *3*(3), 453–466. https://journal-computing.org/index.php/journal-ita/article/view/349

Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, *3*(1), 50. https://doi.org/10.35314/isi.v3i1.335

Saputra, D., & Pribadi, M. R. (2023). Analisis Sentimen Masyarakat terhadap Layanan Provider Internet di Indonesia menggunakan SVM. *MDP Student Conference*, *2*(1), 32–41. https://doi.org/10.35957/mdp-sc.v2i1.4554

Thaniket, R., Kusrini, & Luthf, E. T. (2019). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine. *JURNAL FATEKSA: Jurnal Teknologi Dan Rekayasa*, *13*(2), 69–83.