

IMPLEMENTASI *WEB SCRAPING* UNTUK ULASAN PADA *TWITTER* MENGGUNAKAN ASOSIASI TEKS (STUDI KASUS: *FILM KKN DI DESA PENARI*)

Implementation of Web Scraping for Reviews on Twitter Using Text Associations (Case Study: Film KKN di Desa Penari)

Frans Xavier Natalius Antoni¹, Neva Satyahadewi^{2*}, Hendra Perdana³

^{1,2,3} Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Kalimantan Barat, Indonesia

E-mail Corresponding Author: neva.satya@math.untan.ac.id

Abstrak: Pengguna *twitter* biasanya memberikan komentar yang berisi mengenai ulasan suatu *film* yang sedang tayang. Teknik yang dapat digunakan untuk mengambil komentar pada *twitter* yaitu *Web Scraping*. Penelitian ini bertujuan untuk mengimplementasikan *Web Scraping* dalam mengumpulkan data pada *twitter* dan mengimplementasikan Asosiasi Teks untuk mendapatkan informasi antar kata yang terbentuk. Penelitian ini menggunakan komentar yang berisi mengenai ulasan *Film KKN di Desa Penari* pada tanggal 30 April 2022. Komentar yang diperoleh tidak semua berisi ulasan, sehingga perlu dilakukan seleksi terhadap komentar tersebut. Hasil seleksi dari 866 komentar diperoleh sebanyak 116 ulasan positif dan 83 ulasan negatif. Data yang diperoleh dari komentar tidak bisa langsung dianalisis, sehingga perlu melalui tahap *text preprocessing*. Adapun tahap *text preprocessing* yaitu *cleansing data*, *case folding*, *spelling normalization*, *filtering*, dan *tokenizing*. Setelah melalui tahap *text preprocessing*, ulasan tersebut kemudian dianalisis untuk mendapatkan informasi yang penting dengan menggunakan Asosiasi Teks. Hasil Asosiasi Teks untuk ulasan positif diperoleh informasi bahwa penonton memberikan penilaian terhadap tokoh, akting dan sinematografi yang bagus, kemudian *film* yang ditayangkan juga sesuai dengan cerita *thread* pada *twitter*, dan sinematografi juga keren. Sedangkan untuk ulasan negatif penonton memberikan penilaian bahwa *Film KKN di Desa Penari*, *film* yang biasa dan hantu yang ditayangkan juga kurang seram.

Kata Kunci: Asosiasi Teks, *Text Preprocessing*, *Twitter*, Ulasan, *Web Scraping*

Abstract: *Twitter* users usually provide comments that contain reviews of a film that is currently airing. The technique that can be used to retrieve comments on *Twitter* is *Web Scraping*. This research aims to implement *Web Scraping* in collecting data on *Twitter* and implementing *Text Association* to get information between words formed. This research uses comments that contain reviews of films *KKN di Desa Penari* on April 30, 2022. The comments obtained do not all contain reviews, so it is necessary to select the comments. The selection results from 866 comments obtained as many as 116 positive reviews and 83 negative reviews. Data obtained from comments cannot be directly analyzed, so it needs to go through the *text preprocessing* stage. The *text preprocessing* stages are *data cleansing*, *case folding*, *spelling normalization*, *filtering*, and *tokenizing*. After going through the *text preprocessing* stage, the reviews are then analyzed to get important information using *Text Association*. The results of *Text Association* for Positive Reviews obtained information that the audience gave an assessment of the characters, acting, and good cinematography, then the film shown was also in accordance with the story thread on *Twitter*, and the cinematography was also cool. As for the negative reviews, the audience gave an assessment that the film *KKN di Desa Penari* was an ordinary film and the ghosts shown were also less scary.

Keywords: *Reviews*, *Text Association*, *Text Preprocessing*, *Twitter*, *Web Scraping*

1. PENDAHULUAN

Dewasa ini perkembangan di bidang teknologi informasi dan komunikasi, terutama internet sudah mengalami perkembangan yang sangat pesat. Internet kini sudah menjadi kebutuhan penting dalam kegiatan sehari-hari. Informasi yang disajikan dari penyedia layanan *web* salah satunya yaitu mengenai data teks. Salah satu data teks yang dapat diambil ialah data dari *twitter*. *Twitter* merupakan salah satu dari sekian banyak media sosial yang sangat familiar di masyarakat Indonesia. Banyaknya pemakai *twitter* di Indonesia pada awal 2022 berjumlah 18,45 juta [1].

Pengguna aktif di *twitter* akan menyampaikan informasi atau komentar terbaru mengenai apa yang ramai dibicarakan sehingga dapat menjadi *trending topic*. Masyarakat Indonesia sering menyampaikan pendapat atau opininya melalui *twitter*. Data tentang opini mudah dikumpulkan dikarenakan banyaknya pengguna aktif yang menyampaikan opininya melalui *twitter* [2]. Salah satu komentar yang disampaikan oleh pengguna *twitter* adalah ulasan mengenai *film* yang sedang tayang di bioskop. Hal tersebut dapat dimanfaatkan untuk mengklasifikasikan berbagai ulasan sehingga menghasilkan informasi yang baik [3].

Banyaknya komentar yang masuk pada *twitter* tentu sangat memakan waktu yang lama apabila menyalin secara manual. Teknik yang digunakan untuk mengambil data semi terstruktur dari situs atau *website* adalah *Web Scraping*. Data yang diperoleh dari *Web Scraping* tersebut kemudian akan digunakan untuk analisis [4]. Analisis data tersebut dilakukan dengan menggunakan *text mining*. *Text mining* ialah ekstraksi pola berupa informasi yang tidak terstruktur yang kemudian akan didapatkan pola dari data tersebut [5]. Metode *text mining* yang bisa digunakan salah satunya yaitu dengan mengimplementasikan Asosiasi Teks. Asosiasi Teks digunakan untuk mengetahui hubungan atau keterkaitan antar kata yang muncul bersamaan dalam sebuah ulasan [6].

Penelitian yang dilakukan oleh [7] menggunakan Asosiasi Teks untuk mengetahui sentimen pada aplikasi Tiktok di *Google Play*. Hasil asosiasi untuk ulasan positif menunjukkan bahwa pengguna menyukai Tiktok versi saat ini karena terdapat video lucu yang ada di beranda. Mereka juga tidak menyesal telah mendownload Tiktok serta mengajak pembaca untuk mendownload juga. Sedangkan asosiasi untuk ulasan negatif menunjukkan bahwa pengguna memberikan ulasan negatif karena tidak berhasil mendaftar atau masuk karena tidak memenuhi syarat tanggal lahir.

Penelitian lain juga dilakukan oleh [8] menggunakan asosiasi untuk menganalisis sentimen pada situs *tripadvisor*. Pengumpulan data dilakukan dengan menggunakan *Web Scraping*. Hasil asosiasi untuk ulasan positif menunjukkan bahwa pengunjung memberikan penilaian positif mengenai kamar hotel yang minimalis dan bergaya moderen, staf yang ramah serta dapat membantu, menu makanan yang bervariasi pada restoran, dan memiliki fasilitas lain berupa taman bermain, kolam renang, *wifi*, dan *jogging track*. Namun, asosiasi teks untuk ulasan negatif terdiri dari pelayan yang membersihkan kamar kurang cermat dalam membersihkan bagian sudut-sudut kamar.

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah pada penelitian ini adalah bagaimana implementasi *Web Scraping* untuk mengumpulkan data ulasan *film* dari *twitter*. Kemudian bagaimana implementasi dari Asosiasi Teks untuk mendapatkan informasi yang diperoleh dari asosiasi antar kata yang terbentuk. Tujuan dari penelitian ini ialah untuk mengimplementasikan *Web Scraping* dalam mengumpulkan data ulasan dari komentar *twitter*. Kemudian dari data yang diperoleh, mengimplementasikan Asosiasi Teks untuk mendapatkan informasi dari asosiasi antar kata yang terbentuk.

2. METODOLOGI

2.1. Deskripsi Data

Data yang digunakan diambil dari komentar yang ada di *twitter* pada tanggal 30 April 2022 yang berisi tentang ulasan *Film KKN* di Desa Penari. Ulasan yang digunakan merupakan ulasan yang menggunakan Bahasa Indonesia. Ulasan yang dipilih kemudian dipisahkan menjadi ulasan positif dan ulasan negatif. Ulasan positif merupakan tanggapan atau positif atau yang baik mengenai *film* yang diulas, sedangkan ulasan negatif merupakan tanggapan negatif atau yang buruk mengenai *film KKN* di Desa Penari.

2.2. Tahapan Penelitian

Proses *Web Scraping* komentar pada *twitter* dilakukan dengan bantuan *Jupyter Notebook* dan harus terkoneksi ke internet. Komentar yang diperoleh tidak semuanya berisi mengenai ulasan. Masih terdapat komentar yang berisi tentang informasi penayangan *film*, keinginan untuk menonton *film*, dan komentar yang hanya memberikan *rating* namun tidak memberikan tanggapan mengenai *film*. Proses pemilihan komentar yang berisi ulasan disebut juga *data selection*. *Data selection* dalam *data mining* merupakan proses yang digunakan dengan tujuan memilih data yang sesuai untuk dianalisis [9].

Setelah mendapatkan data yang digunakan, langkah selanjutnya adalah melakukan *text preprocessing*. *Text preprocessing* ialah tahapan awal yang dilakukan untuk melakukan analisis pada *text mining* agar menjadi data yang struktural [10]. Tahapan *text preprocessing* yang dilakukan adalah sebagai berikut:

- Cleansing Data*, dilakukan dengan tujuan untuk menghapus *link*, tanda baca dan karakter lain seperti *hashtag* dan *mention* yang ada pada teks [11].
- Case Folding*, yaitu langkah untuk mengganti seluruh huruf kapital menjadi huruf kecil [11].
- Spelling Normalization*, yaitu langkah untuk memperbaiki ejaan kata yang tidak standar menjadi ejaan kata yang standar atau baku [12].
- Filtering*, menghilangkan kata yang tidak penting misalnya kata ganti, kata depan, dan kata keterangan dengan menggunakan *stop word* [12].
- Tokenizing*, merupakan langkah untuk memisahkan teks menjadi potongan-potongan kata yang tidak berhubungan satu sama lain [13].

Setelah data dipisahkan menjadi potongan-potongan kata, langkah selanjutnya yang dilakukan adalah membuat *word cloud* dari masing-masing ulasan. *Word cloud* adalah cara yang populer untuk memvisualisasikan data teks ke dalam bentuk grafis [14]. Frekuensi sebuah kata menentukan ukuran kata pada *word cloud* [15].

Setelah memperoleh kata-kata yang digunakan, langkah selanjutnya adalah mencari nilai asosiasi teks pada masing-masing ulasan. Asosiasi teks dilakukan dengan tujuan untuk menemukan seberapa sering kombinasi elemen yang muncul bersamaan pada kumpulan data [16]. Istilah yang biasa digunakan untuk menyatakan hubungan yang terdiri dari dua variabel adalah korelasi. Korelasi digunakan untuk data yang bersifat kuantitatif, sedangkan istilah asosiasi dikaitkan dengan hubungan dua variabel yang bersifat kualitatif [7].

Penelitian ini menggunakan asosiasi *Rank Spearman*. Asosiasi *Rank Spearman* merupakan ukuran asosiasi dengan kedua variabel yang diukur dengan skala ordinal. Variabel yang akan diteliti dapat diberi peringkat dalam dua rangkaian secara berurutan. Rumus yang dapat digunakan untuk mendapatkan nilai koefisien Korelasi *Rank Spearman* yaitu sebagai berikut [17]:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

dimana $i = 1, 2, \dots, n$, dan d_i merupakan selisih antara *ranking* x_i dan y_i untuk masing-masing pengamatan. Besarnya kekuatan hubungan antar variabel atau kata yang didasarkan dari nilai koefisien korelasi dapat dilihat pada Tabel 1 [18]:

Tabel 1. Kekuatan Hubungan Antar Variabel

Nilai Koefisien	Kekuatan Hubungan
0,00 – 0,25	Lemah
0,26 – 0,50	Sedang
0,51 – 0,75	Kuat
0,76 – 1,00	Sangat Kuat

Kekuatan hubungan antar variabel tidak hanya diukur dari besaran nilai koefisien korelasi, namun juga perlu dilakukan pengujian hipotesis. Adapun hipotesis yang digunakan sebagai berikut:

H_0 : tidak terdapat hubungan yang signifikan antar dua variabel

H_1 : terdapat hubungan yang signifikan antar kedua variabel

Statistik uji koefisien dapat dihitung dengan menggunakan rumus berikut [19]:

$$t_{hitung} = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \tag{2}$$

Kriteria pengujian yang digunakan yaitu tolak H_0 apabila nilai $|t_{hitung}| > t_{\frac{\alpha}{2}; n-2}$. Kesimpulan yang dapat diambil apabila H_0 ditolak adalah bahwa terdapat hubungan yang signifikan antar kedua variabel tersebut.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Web Scraping Twitter

Kata kunci yang digunakan ketika melakukan pengambilan komentar pada *twitter* adalah “Film KKN di Desa Penari”. Jumlah komentar yang diperoleh sebanyak 866 komentar dan atribut yang diperoleh terdiri dari tanggal dan waktu ketika pengguna *twitter* memberikan komentar, *tweet ID*, isi komentar, dan *username twitter*. Hasil *Web Scraping* yang diperoleh ditunjukkan pada Gambar 1.

	Datetime	Tweet Id	Text	Username
0	2022-04-30 23:56:46+00:00	1520552772999217152	KKN di Desa Penari:\n- Ceritanya menarik\n- Ak...	TonyApbeaux
1	2022-04-30 23:51:48+00:00	1520551524216098816	KKN di desa penari tuh sampe kapan ya di XXI? ...	Jehaaa97
2	2022-04-30 23:49:14+00:00	1520550879711358976	abis lebaran kita gas dr strange dan kkn di de...	yaudahlahrid
3	2022-04-30 23:40:01+00:00	1520548559308804097	Review mvs KKN di Desa Penari dari sender, fil...	moviementes
4	2022-04-30 23:40:01+00:00	1520548557031612417	Nanti siang mau nyoba nonton horor 'KKN Desa P...	tyoksat
...
861	2022-04-30 00:30:45+00:00	1520198937285898240	Adek gua ngajakin gua nonton KKN Desa Penari u...	slowhare
862	2022-04-30 00:10:38+00:00	1520193874727440386	Gas ntar malem nonton KKN Di Desa Penari. 🍕	ziilm_
863	2022-04-30 00:07:06+00:00	1520192987908624384	Belum Tayang, Tiket Film KKN di Desa Penari Lu...	VIVAcoid
864	2022-04-30 00:03:17+00:00	1520192027677577216	Pgn nonton KKN di Desa Penari	aliothjake
865	2022-04-30 00:02:36+00:00	1520191853693644800	@diosefia Setuju banget hahaha,semoga 2910 MDP...	arwaaaann

Gambar 1. Komentar Hasil Web Scraping

3.2. Seleksi Data

Komentar yang diperoleh tidak semuanya berisi mengenai ulasan *film*. Penelitian ini menggunakan data yang berisi mengenai ulasan *film* KKN di Desa Penari. Komentar tersebut akan diseleksi dan dipilih komentar yang berisi mengenai ulasan. Setelah diseleksi terdapat 116 ulasan positif dan 83 ulasan negatif. Kemudian ulasan tersebut digunakan sebagai data penelitian ini. Contoh komentar yang berisi ulasan dan bukan ulasan ditampilkan pada Tabel 2.

Tabel 2. Contoh Komentar Ulasan dan Bukan Ulasan

No	Komentar	Keterangan
1	KKN di Desa Penari: - Ceritanya menarik - Akting Aktor dan Aktris sangat bagus - Di pertengahan <i>film</i> terdengar <i>error sound</i> - <i>Ending so sad</i> .	Ulasan
2.	Abis lebaran kita gas dr <i>strange</i> dan kkn di desa penari	Bukan Ulasan

3.3. Text Preprocessing

Data yang telah diseleksi kemudian akan melewati tahap *text preprocessing*. Hal ini dikarenakan data awal yang diperoleh masih terdapat banyak *noise* dan masih belum terstruktur dengan baik. Proses *text preprocessing* terdiri dari *cleansing data*, *case folding*, *spelling normalization*, *filtering*, dan *tokenizing*. Hasil dari *text preprocessing* ditampilkan pada Tabel 3.

Isi Ulasan	Proses
13. #kkndidesapenari Nonton KKN di Desa Penari di hari dan jam pertama <i>Film</i> yang pantas ditunggu, memang sebgas itu. Pesan moralnya dapet, horrornya dapet, dan tentu kearifan lokalnya juga dapet @SimpleM81378523https://t.co/rdF4hZHkTK	Ulasan Awal
Nonton di hari dan jam pertama <i>Film</i> yang pantas ditunggu memang sebgas itu Pesan moralnya dapet horrornya dapet dan tentu kearifan lokalnya juga dapet	Setelah <i>Cleansing Data</i>
nonton di hari dan jam pertama <i>film</i> yang pantas ditunggu memang sebgas itu pesan moralnya dapet horrornya dapet dan tentu kearifan lokalnya juga dapet	Setelah <i>Case Folding</i>
nonton di hari dan jam pertama <i>film</i> yang pantas ditunggu memang bagus itu pesan moralnya dapat horornya dapat dan tentu kearifan lokalnya juga dapat	Setelah <i>Spelling Normalization</i>
<i>film</i> pantas tunggu bagus pesan moral dapat horor dapat kearifan lokal dapat	Setelah <i>Filtering</i>
“ <i>film</i> ”, “pantas”, “tunggu”, “bagus”, “pesan”, “moral”, “dapat”, “horor”, “dapat” “kearifan”, “lokal”, “dapat”	Setelah <i>Tokenizing</i>

3.4. Visualisasi Kata dengan Word Cloud

Visualisasi kata dengan *word cloud* dilakukan dengan tujuan untuk melihat kata apa saja yang sering digunakan ketika memberikan ulasan. Visualisasi kata yang digunakan ketika memberikan ulasan positif ditampilkan pada Gambar 2.



Gambar 2. Word Cloud Ulasan Positif

Gambar 2 menunjukkan gambaran kata-kata yang digunakan untuk memberikan ulasan positif. Visualisasi dengan *word cloud* menunjukkan bahwa semakin besar ukuran kata artinya semakin sering kata itu digunakan ketika memberikan ulasan. Kata yang sering digunakan untuk ulasan positif yaitu kata “bagus”, “thread”, “film”, “cerita”, dan kata “sesuai”. Sedangkan visualisasi kata yang digunakan ketika memberikan ulasan negatif ditampilkan pada Gambar 3.



Gambar 3. Word Cloud Ulasan Negatif

Gambar 3 menunjukkan gambaran kata-kata yang digunakan untuk memberikan ulasan negatif. Sama seperti visualisasi *word cloud* untuk ulasan positif, visualisasi dengan *word cloud* untuk ulasan negatif juga menunjukkan bahwa semakin besar ukuran kata artinya semakin sering kata itu digunakan ketika memberikan ulasan. Kata yang sering digunakan untuk ulasan negatif yaitu kata “film”, “tidak”, “kurang”, dan kata “biasa”.

3.5. Hasil Asosiasi Teks

Setelah memperoleh kata-kata yang digunakan menggunakan *word cloud*, langkah selanjutnya adalah mencari asosiasi antar kata yang sering digunakan yang muncul bersamaan dengan kata lain ketika memberikan ulasan. Hasil asosiasi yang digunakan merupakan nilai koefisien yang bernilai positif dan ketika dilakukan uji hipotesis H_0 ditolak.

Misalkan akan dicari Asosiasi Teks kata “bagus” dan kata “akting” pada ulasan positif. Langkah yang dilakukan terlebih dahulu adalah memberi skor pada kata tersebut. Jika kata “bagus” muncul pada suatu ulasan, maka diberi skor sebesar = 1. Namun apabila kata “bagus” tidak muncul pada suatu ulasan, maka diberi skor sebesar = 0. Demikian juga kata “akting”, jika muncul diberi skor sebesar = 1 dan apabila tidak muncul diberi skor sebesar = 0. Hasil perhitungan asosiasi antar kata yang terbentuk untuk ulasan positif dengan menggunakan Persamaan (1) ditampilkan pada Tabel 4.

Tabel 4. Hasil Asosiasi Kata Pada Ulasan Positif

	Bagus	Thread	Keren
Tokoh	0,30	Sesuai	0,40
Akting	0,30	Cerita	0,40
Sinematografi	0,30	Visualisasi	0,20

Berdasarkan hasil yang diperoleh pada Tabel 4, kata-kata yang diambil merupakan kata yang banyak digunakan ketika memberikan ulasan. Kemudian dari kata-kata tersebut berasosiasi dengan kata-kata lain. Hubungan antar kata yang terbentuk berdasarkan Tabel 4 menunjukkan hubungan yang sedang, kecuali untuk asosiasi antara kata “thread” dan kata “visualisasi” termasuk ke dalam hubungan yang lemah. Kata “bagus” berasosiasi dengan kata “tokoh” dengan nilai asosiasi sebesar 0,30, kata “akting” dengan nilai asosiasi sebesar 0,30, dan kata “sinematografi” dengan nilai asosiasi sebesar 0,30. Berdasarkan asosiasi kata yang terbentuk, informasi yang diperoleh adalah bahwa penonton memberikan ulasan bagus mengenai tokoh, akting, dan sinematografi dalam *Film KKN* di Desa Penari.

Selanjutnya untuk menentukan asosiasi kata pada ulasan negatif sama seperti pada ulasan positif. Suatu kata yang muncul pada suatu ulasan akan diberi skor sebesar=1 serta kata yang tidak muncul akan diberi skor sebesar = 0. Hasil perhitungan asosiasi untuk ulasan negatif dengan menggunakan Persamaan (1) ditampilkan pada Tabel 5.

Tabel 5. Hasil Asosiasi Kata Pada Ulasan Negatif

	Film	Kurang
Biasa	0,47	Hantu
		0,23

Berdasarkan Tabel 5, kata “*film*” berasosiasi dengan kata “biasa” dengan nilai asosiasi sebesar 0,47. Hubungan asosiasi antara kata “*film*” dengan kata “biasa” termasuk ke dalam hubungan yang sedang. Informasi yang diperoleh dari asosiasi kata yang terbentuk adalah bahwa *film* tersebut adalah *film* yang biasa. Kata “kurang” berasosiasi dengan kata “hantu” dengan nilai asosiasi sebesar 0,23. Hubungan asosiasi antara kata “kurang” dan kata “hantu” termasuk ke dalam hubungan yang lemah. Informasi yang diperoleh dari asosiasi kata yang terbentuk adalah bahwa hantu yang ditampilkan pada *film* tersebut kurang seram.

4. KESIMPULAN

Hasil *Web Scraping* pengambilan komentar *twitter* pada tanggal 30 April 2022 didapatkan komentar sebanyak 866 komentar. Komentar yang digunakan merupakan komentar yang berisi ulasan, sehingga setelah diseleksi terdapat 116 ulasan positif dan 83 ulasan negatif. Komentar yang diperoleh tidak dapat dianalisis secara langsung, sehingga perlu melalui tahapan *text preprocessing*. Adapun proses *text preprocessing* yang dilakukan yaitu *cleansing data*, *case folding*, *spelling normalization*, *filtering*, dan *tokenizing*.

Berdasarkan hasil Asosiasi Teks yang dilakukan, diperoleh informasi untuk ulasan positif penonton memberikan penilaian yang bagus terhadap akting, tokoh, dan sinematografi yang keren. *Film* yang ditayangkan juga sesuai dengan cerita *thread* yang ada di *twitter*. Sedangkan untuk ulasan negatif penonton menyatakan bahwa *Film* KKN di Desa Penari merupakan *film* yang biasa, hantu yang ditayangkan juga kurang seram.

DAFTAR PUSTAKA

- [1] TiNewss.com, “Jumlah Pengguna *Twitter* di Indonesia pada 2022.” Accessed: Sep. 12, 2022. [Online]. Available: <https://www.tinewss.com/indonesia-news/pr-1853618409/jumlah-pengguna-twitter-di-indonesia-pada-2022>
- [2] I. B. N. W. Manuaba, G. R. Dantes, and G. Indrawan, “Analisis Sentimen Data Provider Layanan Internet Pada *Twitter* Menggunakan Support Vector Machine (SVM) Dengan Penambahan Algoritma Levenshtein Distance,” *Jurnal Sistem Komputer dan Kecerdasan Buatan*, vol. 5, no. 2, pp. 9–17, 2022.
- [3] S. Fransiska and Yolanda, “Analisis Sentimen *Twitter* Untuk Review *Film* Menggunakan Algoritma Naive Bayes Classifier (NBC) Pada Sentimen R Programming,” *Jurnal Siliwangi Seri Sains dan Teknologi*, vol. 5, no. 2, pp. 68–71, 2019.
- [4] D. D. Ayani, H. S. Pratiwi, and H. Muhandi, “Implementasi *Web Scraping* untuk Pengambilan Data pada Situs Marketplace,” *Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 4, pp. 257–262, 2019.
- [5] R. Feldman and J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*. New York: Cambridge University Press, 2007.
- [6] E. B. Santoso and A. Nugroho, “Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook,” *Jurnal Eksplora Informatika*, vol. 9, no. 1, pp. 60–69, Sep. 2019, doi: 10.30864/eksplora.v9i1.254.
- [7] S. Fide, Suparti, and Sudarno, “Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi,” *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, 2021, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [8] F. F. Rahanto and I. Kharisudin, “Analisis Sentimen Data Ulasan Menggunakan Metode Naive Bayes Studi Kasus The Wujil Resort & Conventions pada Situs Tripadvisor,” *UNNES Journal of Mathematics*, vol. 10, no. 1, pp. 56–62, 2021.
- [9] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2006.
- [10] D. Alita and A. Rahman, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *Jurnal Komputasi*, vol. 8, no. 2, pp. 50–58, 2020.

- [11] D. P. Santoso and W. Wibowo, “Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store,” *Jurnal Sains dan Seni ITS*, vol. 11, no. 2, pp. 2337–3520, 2022.
- [12] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, “Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 239, Jul. 2020, doi: 10.26418/justin.v8i3.39118.
- [13] N. S. Wardani, A. Prahutama, and P. Kartikasari, “Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial,” *Jurnal Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [14] C. A. DePaolo and K. Wilkinson, “Get Your Head Into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data,” *Journal TechTrends*, vol. 58, no. 3, pp. 38–44, 2014.
- [15] F. S. Pamungkas and I. Kharisudin, “Analisis Sentimen dengan SVM, Naïve Bayes, dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial *Twitter*,” *PRISMA (Prosiding Seminar Nasional Matematika)*, vol. 4, pp. 628–634, 2021.
- [16] A. Novianti and E. Elisa, “Penentuan Aturan Asosiasi Pola Pembelian pada Minimarket dengan Algoritma Apriori,” *Technology and Science (BITS)*, vol. 2, no. 1, pp. 64–70, 2020.
- [17] S. Nugroho, S. Akbar, and R. Vusvitasari, “Kajian Hubungan Koefisien Korelasi Pearson (r), Spearman-rho (ρ), Kendall-Tau (τ), Gamma (G), dan Somers (dxy),” *Jurnal Gradien*, vol. 4, no. 2, pp. 372–381, 2008.
- [18] M. Nurwahid, “Korelasi Antara Motivasi Belajar Siswa dalam Pembelajaran Online dengan Hasil Belajar Matematika di Masa Pandemi,” *Jurnal Cendekia: Jurnal Pendidikan Matematika*, vol. 05, no. 02, pp. 1127–1137, 2021.
- [19] D. Dewanti, Suparti, and A. Prahutama, “Pemodelan Indeks Harga Saham Gabungan (IHSG) dan Jakarta Islamic Index (JII) Menggunakan Regresi Birespon Spline Truncated Berbasis GUI R,” *Statistika*, vol. 8, no. 2, pp. 134–143, 2020.