Jurnal Pendidikan dan Konseling
Volume 4 Nomor 6 Tahun 2022
E-ISSN: 2685-936X dan P-ISSN: 2685-9351
Universitas Pahlawan Tuanku Tambusai

Analisis Sentimen Tempat Wisata Di Jakarta Pasca Covid -19 Dengan Algoritma *Naïve Bayes*

Nabila Aurelia Rahma[1], Garno[2], Nina Sulistiyowati[3]
[1,2,3] Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
Email : [1]nabila.aurelia18208@student.unsika.ac.id, [2]garno@staff.unsika.ac.id,
[3]nina.sulistio@unsika.ac.id

Abstrak

Pariwisata merupakan sektor yang menjadi imbas dari kebijakan PPKM, karena mengalami penurunan pengunjung. Hal tersebut menjadikan banyak pariwisata yang mengalami kerugian yang besar. Penurunan pengunjung menyebabkan tempat wisata harus memikirkan cara untuk mengembalikan pengunjung seperti saat sebelum pandemi covid - 19 agar tidak mengakibatkan kerugian yang sangat besar. Oleh karena itu dilakukannya penelitian menggunakan media sosial *Twitter* untuk mencari sebuah opini atau tanggapan dari pengunjung tempat wisata di jakarta pasca covid — 19 yaitu Dufan dan TMII yang merupakan tempat paling sering dikunjungi pada 2020. Analisis sentimen digunakan untuk mengolah opini tersebut dengan menggunakan algoritma *Naïve bayes Classifier* dan Metodologi *KDD* (*Knowledge Discovery in Database*) dengan tahapan yaitu *data selection, Pre Processing, transformation data, data mining* dan *evaluation*. Data yang digunakan berjumlah 9729 yang diambil dari *Twitter* dengan *keyword* dufan dan tmii. Penelitian ini menggunakan transformasi *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan melakukan pengujian menggunkan pembagian empat model yaitu 90:10, 80:20, 70:30, dan 60:40. Model 90:10 mempunyai nilai skor akurasi tertinggi yaitu 65%, *precision* 53%, *Recall* 51% dan *F-Measure* 50%. Sedangkan pengujian performa model dengan menggunakan nilai AUC menghasilkan nilai 0.71.
Keywords: Analisis Sentimen, *Naive bayes*, Turis, *Twitter*,TF-IDF

Abstract

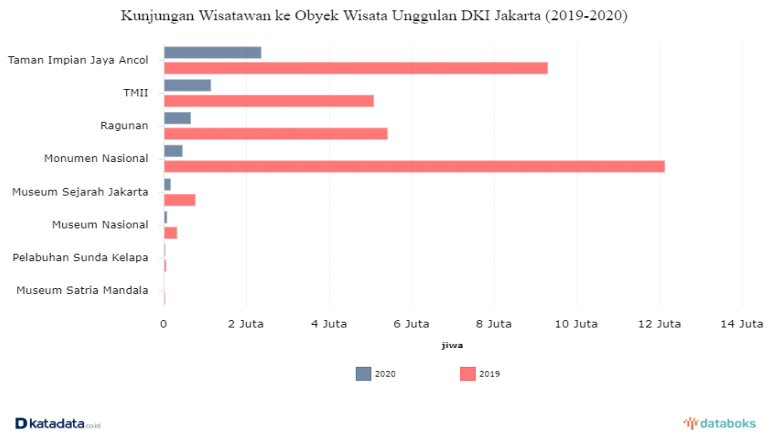
Tourism is a sector that is the impact of the PPKM policy, because it has decreased visitors. This causes a lot of tourism to suffer huge losses. The decline in visitors causes tourist attractions to have to think of ways to return visitors as they were before the COVID-19 pandemic so as not to cause huge losses. Therefore, a study was conducted using Twitter social media to seek an opinion or response from visitors to tourist attractions in post-covid Jakarta, namely "Dufan" and "TMII" because they are the most frequently visited places. Sentiment analysis can be used to process the opinion using the Naïve bayes Classifier algorithm and the KDD (Knowledge Discovery in Database) methodology with several stages such as data selection, Pre Processing, data transformation, data mining, evaluation. The data used are 9729 taken from Twitter with the keywords "dufan" and "tmii. This research uses Term Frequency-Inverse Document Frequency (TF-IDF) transformation by testing using 4 models 90:10, 80:20, 70:30, and 60:40. The 90:10 model has the highest accuracy score of 65%, precision 53%, Recall 51% dan F-Measure 50%. While testing the performance of the model using the AUC value produces a value of 0.71.
Keywords: *Sentiment Analysis, Naïve bayes, Touris, Twitter, TF-IDF*

PENDAHULUAN

Pemberlakuan Pembatasan Kegiatan Masyarakat atau PPKM merupakan kebijakan yang dibentuk oleh pemerintah untuk menekan angka penyebaran virus Covid-19. Kebijakan tersebut mengharuskan segala kegiatan yang dilakukan sehari - hari untuk keluar rumah diganti untuk melakukannya di dalam rumah. Seperti bersekolah, bekerja, bahkan untuk membeli makanan dilakukan dengan melakukan pengiriman tanpa kontak fisik. Kebijakan tersebut membuat masyarakat dikurung di rumah nya, sehingga tidak bisa bepergian kemana - mana.

Pariwisata merupakan sektor yang menjadi imbas dari kebijakan tersebut, dikarenakan diperlukannya masyarakat yang berkunjung. Hal tersebut menjadikan banyak pariwisata yang mengalami kerugian yang besar (Kemenkraf, 2020). Setelah dua tahun pandemi berlangsung dan virus covid - 19 berkurang, kebijakan untuk tetap dirumah sudah tidak berlaku. Sehingga hal tersebut menjadikan titik balik tempat - tempat pariwisata untuk bangkit kembali.

Salah satu pusat pariwisata terbesar yaitu di DKI Jakarta, banyak pariwisata yang ada di Jakarta setelah masa pandemi covid-19 sudah mulai bangkit dan kembali beroperasi. Permasalahan yang terjadi yaitu berkurangnya pengunjung dikarenakan masih adanya ketakutan terhadap pandemi covid - 19, bisa dilihat pada gambar 1.1 yang menunjukan jumlah kunjungan wisata di tempat wisata unggulan di Jakarta pada tahun 2019 — 2020 pasca covid -19 :



Gambar 1. 1 Data pengunjung Objek Wisata DKI Jakarta 2019 — 2020
(Sumber : katadata.co.id)

Dari beberapa tempat wisata yang ada di Jakarta, di gambar tersebut . Bisa dilihat pada tempat wisata Taman Impian Jaya Ancol dan Taman Mini Indonesia Indah merupakan 2 tempat pariwisata yang memiliki pengunjung teratas pada 2020. Di Taman Jaya Impian Ancol mempunyai beberapa tempat wisata salah satu yang menjadi favorit adalah Dunia Fantasi (Dufan) (kumpulaninfo.com, 2021). Pada tahun 2019 sebelum adanya wabah covid -19 pengujung dari wisata unggulan tersebut mencapai angka 12 juta pengunjung, akan tetapi setelah mengalami covid — 19 dan pemberlakuan PPKM angkanya langsung menurun drastis sekitar 3 juta pengujung.

Penurunan yang sangat signifikan yang terjadi menyebabkan tempat wisata harus memikirkan cara untuk mengembalikan pengunjung seperti saat sebelum pandemi covid - 19 agar tidak mengakibatkan kerugian yang sangat besar. Dibutuhkannya pendapat dari pengunjung - pengunjung yang ada untuk meningkatkan daya tarik pengunjung agar banyak yang mengunjungi tempat wisata, serta meningkatkan kualitas tempat wisata agar lebih dari sesaat sebelum terjadinya pandemi. Pembiasaan untuk menjalan kehidupan baru setelah pandemi juga menjadi tantangan untuk tempat -

tempat wisata, walau pandemi sudah berkurang tetapi pencegahan yang dilakukan tempat wisata akan berakibat baik untuk pengunjung. Pendapat dari pengunjung menjadikan bahan evaluasi untuk para tempat wisata. Oleh karena itu dilakukannya penelitian terhadap sentimen para pengunjung tempat wisata yang ada di Jakarta. Pendapat tersebut bisa berupa tanggapan positif, tanggapan negatif maupun tanggapan netral. Hal tersebut bisa membantu pengelola tempat wisata untuk mengetahui tingkat kualitas dan layanan yang diberikan terhadap pengunjung, untuk membantu meningkatkan jumlah pengunjung yang ada.

Media sosial merupakan salah satu media untuk para pengunjung untuk memberikan tanggapan dan komentar saat mereka mengunjungi tempat wisata. Salah satu media yang sering digunakan untuk memberikan komentar pengunjung yaitu media sosial *Twitter*. *Twitter* adalah salah satu dari banyaknya *platform* media sosial yang telah banyak penggunanya di Indonesia. Menurut data pada Statista, Indonesia menempati urutan kelima sebagai pengguna *Twitter* terbanyak di dunia yakni dengan jumlah penggunanya mencapai angka 18.45 juta (Statista, 2022). *Twitter* merupakan *platform* bebas bagi para pengguna untuk menyampaikan pendapat mereka salah satunya untuk para pengunjung tempat wisata. Hal tersebut bisa dijadikan bahan referensi untuk para tempat wisata mengevaluasi tempat wisatanya.

Oleh karena itu dilakukannya penelitian menggunakan media sosial *Twitter* untuk mencari sebuah opini atau tanggapan dari pengunjung tempat wisata di jakarta pasca covid - 19. Opini atau sentimen pengunjung yang dalam istilah *Twitter* biasa disebut sebagai ciutan ini dapat berupa opini negatif, positif dan netral. Akan tetapi jumlah data tersebut cukup banyak sehingga dibutuhkan suatu metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentimen. Adapaun untuk mengolah data dari opini para pengunjung yang didapat pada media sosial *Twitter* yaitu dengan menggunakan proses *Knowledge Discovery in Database* (*KDD*). Proses ini untuk dimaksudkan untuk menjalani ekstrasi informasi dari data opini para pengujung pariwisata agar mendapatkan pola informasi berbasis pengetahuan yang dapat digunakan dalam membantu peningkatan kualitas pariwisata (Mirza, 2018). *KDD* sebagai salah satu metode teknis yang memuat berbagai tahapan khusus yang bersifat teknis yang dapat membantu penelitian ini. *KDD* adalah suatu proses model yang bersifat lebih lengkap dan akurat (Shafique & Qaiser, 2014).

Proses pengekstrasian informasi agar mendapatkan pola informasi berbasis pengetahuan dibutuhkan suatu algoritma didalam proses data mining yaitu dengan menggunakan algoritma *naïve bayes*. Algoritma *Naïve bayes* adalah algoritma klasifikasi sederhana yang mana menghitung sekumpulan probabilitas dengan cara menjumlahkan dan mengkombinasikan nilai dari dataset yang diberikan. Algoritma *Naïve bayes* akan digunakan pada penelitian ini dalam proses analisis sentimen pengujung wisata di media sosial *Twitter* terhadap tempat pariwisata di Jakarta pasca covid -19. Dari berbagai referensi penelitian, klasifikasi Algoritma *Naïve bayes* lebih banyak disukai dikarenakan kesederhanaanya dan juga kecepatannya (Basit, 2020). Walaupun klasifikasi Algoritma *Naïve bayes* bisa dikatakan klasifikasi yang sederhana, tetapi hasil yang didapat dari pada klasifikasi Algoritma *Naïve bayes* ini sering sekali mencapai performa yang serupa dengan algoritma klasifikasi lainnya seperti contoh *Neural Network classifier & Decision tree.* Klasifikasi Algoritma *Naïve bayes* ini bisa memberikan akurasi yang tinggi juga cepat di dalam memproses data yang ada di dalam jumlah yang sangat banyak (Nugroho & Cholissodin, 2021).

Penelitian tentang analisis sentimen sebelumnya dilakukan oleh Kautsar Ramadhan S. dan Kemas Muslim L (Sugiharto & Lhaksmana, 2018) tentang analisis sentimen terhadap toko online menggunakan *naïve bayes* pada media sosial *Twitter* menggunakan metode *naïve bayes classifier* dengan data training diambil 900 ciutan dan data testing diambil 300 *tweet* secara manual. Penelitian menjelaskan akurasi terbaik yang dihasilkan oleh algoritma adalah akurasi dari data

Tokopedia yaitu sebesar 83.97% dibandingkan data Lazada yang bernilai 75.26%. Selain itu recall pada data Tokopedia lebih besar dengan nilai 95.48% dibandingkan data Lazada yang bernilai 87.50%, precision data Tokopedia lebih besar dengan nilai 86.12% dibandingkan data Lazada yang bernilai 80.65% dan f-measure data Tokopedia lebih besar dengan nilai 90.56% dibandingkan data Lazada yang bernilai 83.93%. kurangnya atribut yang digunakan pada data training serta adanya perbedaan hasil dari pelabelan secara manual dengan hasil prediksi dari klasifikasi model merupakan kekurangan dari penelitian ini. Penelitian yang dilakukan oleh (Alaei, Becken, & Stantic, 2019) membandingkan beberapa metode analisis sentimen dalam topik pariwisata, kesimpulannya adalah bahwa menggunakan metode lexicon memiliki peningkatan hasil sentimen paling tinggi dibandingkan dengan metode lainnya. Pada penelitian ini penulis akan menggunakan sumber daya kamus leksikon yaitu InSet (Indonesian *Sentiment Lexicon*) yang dikembangkan oleh (Koto & Rahmaningtyas, 2017).

Berdasarkan pembahasan sebelumnya maka dilakukan sebuah penelitian tentang bagaimana menganalisis sentimen terhadap tempat pariwisata di Jakarta pasca covid - 19, dengan cara mengklasifikasikan komentar dan ulasan pengunjung, menganalisis dan mengevaluasi tentang sentimen analisis pengunjung tempat wisata di Jakarta pasca covid-19 dengan *naïve bayes*.

METODE

Metodologi penelitian yang digunakan pada penelitian ini adalah metode *Knowledge Discovery in Database* (*KDD*). Tahapan pada metodologi *KDD* yaitu *data selection, Pre Processing, transformation, data mining, evaluation*.

HASIL DAN PEMBAHASAN

Hasil Penelitian yang telah dilakukan adalah melakukan analisis sentimen dari data *tweet* mengenai pariwisata di Jakarta pasca pandemi Covid-19. Kemudian data *tweet* tersebut diklasifikasikan ke dalam 3 kelas yaitu positif, negatif, dan netral dengan mengimplementasikan algoritma klasifikasi *Naive bayes Classifier*. Evaluasi sistem dengan menggunakan *Confussion Matrix* untuk mengetahui nilai *accuracy, precision, recall, dan f-measure* dari model tersebut.

- ***Data Selection***

	username	tweetcreatedts	text
0	febskadlue	2022-10-05 08:18:27	@shintasucip @convomfs Tipe yg belakang mobiln...
1	septasha	2022-10-05 08:18:24	@wiragalam abis main main didepan kipas angin ...
2	Alfianaprn	2022-10-05 08:05:43	@innerbiatch @numblittlebug20 Bangke mukanya s...
3	abhiseva7602	2022-10-05 08:04:29	Untuk merayakan HUT 8 Oktober nanti, XL Axiata...
4	voizoombo	2022-10-05 08:01:46	chris, kita nggak jadi ketemu di dufan, nanti ...
5	PaxSanguina	2022-10-05 07:59:34	@gilangabasi @innerbiatch @numblittlebug20 kek...
6	hannasmnlg_	2022-10-05 07:57:11	pengen ke dufan tapi bukan buat main, cuma pen...
7	Kamis22281667	2022-10-05 07:51:58	Untuk merayakan HUT 8 Oktober nanti, XL Axiata...
8	onglyongbok	2022-10-05 07:51:09	@Irma_hasanah @wooncrysan dia mau nonton the r...
9	horandust	2022-10-05 07:47:20	@tanyakanrl dufan doang pas smp. harusnya pas ...

Hasil dari *data* selection yang dilakukan dengan cara melakukan teknik *Crawling data* berupa *tweet* dari media sosial *Twitter* menggunakan kata kunci dufan dan tmii dan Bahasa yang digunakan yaitu bahasa Indonesia.. Proses *Crawling data* ini menggunakan akses *Twitter* API. Berikut sampel dataset awal hasil *Crawling* yang dapat dilihat pada Gambar 4.1. Jumlah data awal hasil *Crawling data*

ini berjumlah 9729 yang merupakan data gabungan dari 3 kali melakukan *Crawling* dan waktu yang digunakan pada proses *Crawling data* ini yaitu pada bulan September 2022. Pada dataset awal ini terdiri dari beberapa atribut, diantaranya yaitu *username*, *tweet*createdts, dan *text*. Penjelasan dari masing-masing atribut tersebut dapat dilihat pada Tabel 4.1 berikut.

Tabel 4. 1 Atribut Pada Dataset

No	Atribut	Penjelasan
1	*Username*	Username dari akun *Twitter* yang membuat cuitan tersebut.
2	*Tweetcreatedts*	Waktu dan tanggal dari pembuatan cuitan tersebut.
3	*Text*	Merupakan cuitan dari pengguna *Twitter* berdasarkan *username* tertentu.

Data yang telah di *Crawling* masih sangat tidak terstruktur dan memiliki banyak *noise* seperti tanda baca, angka, simbol, kata yang tidak baku, yang tidak dibutuhkan dalam proses klasifikasi, maka dataset ini akan di proses pada selanjutnya di tahap *Pre Processing*. Kemudian juga data di seleksi, menyeleksi twit yang tidak keluar konteks penelitian, data yang awalnya 9729 diseleksi menjadi 7995 data.

- **Hasil *Pre-Proccesing***

Hasil melakukan pengumpulan dataset, ditemukannya bahwa jenis dataset yang akan digunakan masih kurang ideal untuk dilakukan proses data mining, maka dari hal tersebut perlu dilakukan tahapan *pre proccesing* untuk menghilangkan kata — kata atau karakter — karakter yang tidak dibutuhkan.

- ***Cleaning***

Proses *cleaning* yang merupakan penghapusan URL, angka, tanda baca, atau simbol-simbol yang tidak diperlukan dalam klasifikasi. Berikut merupakan contoh hasil penerapan tahap *cleaning*, bisa dilihat pada Tabel 4.2.

Tabel 4. 2 Sample Hasil Data Cleaning

No	Sebelum	Sesudah
1	Dufan, elu kaga mau kah promo jadi 120rb lagi? ðŸ¥¹ https://t.co/Otywiz17gs	Dufan elu kaga mau kah promo jadi rb lagi
2	kalo kalian pas lagi ke dufan trs belum pernah naik wahana kereta misteri, cobain deh. at least harus coba 1x naik wahana iniðŸ¥°	Kalo kalian pas lagi ke dufan trs belum pernah naik wahana kereta misteri cobain deh at least harus coba x naik wahana ini

Pada contoh hasil tersebut, sebelum dilakukan proses *cleaning* masih terdapat URL.karakter,dan angka yang tidak dibutukan untuk proses klasifikasi, setelah dilakukan proses tahap *cleaning tweet,* data sudah bersih dari URL, angka, tanda baca atau simbol.

Case Folding

Setelah dilakukannya proses *Cleaning,* data tersebut kemudian dilakukan proses *Case Folding* di mana semua huruf kapital dalam dataset diubah menjadi huruf kecil. Hal ini dilakukan agar semua data yang akan diproses memiliki penyeragaman karakter sehingga ketika proses *Pre Processing* selanjutnya yaitu *Tokenizing* lebih muudah dilakukan Berikut contoh hasil tahap *case folding*, bisa dilihat pada Tabel 4.3.

Tabel 4. 3 Sample Hasil Case Folding

No	Sebelum	Sesudah
1.	x lebih menyenangkan daripada main di Dufan	x lebih menyenangkan daripada main di dufan
2.	Mana yang mau ke dufan Ayo main tornado halilintar kora hysteria baling pokoknya semua yang menantang maut	mana yang mau ke dufan ayo main tornado halilintar kora hysteria baling pokoknya semua yang menantang maut

Hasil dari dilakukan tahap *case folding* bisa dilihat semua huruf yang ada pada dalam dataset menjadi huruf kecil, Termasuk huruf yang seharusnya kapital. Dapat dilihat pada tabel 4.3 diatas kata Ayo pada contoh dirubah menjadi huruf kecil yaitu ayo.

Tokenizing

Hasil dari *Tokenizing* dilakukannya pengubahan data *tweet* yang sebelumnya masih berbentuk kalimat akan dipecah menjadi bentuk kata perkata. Proses ini akan mempermudah saat masuk ke tahap transformasi dengan tidak memproses kalimatnya tapi memproses kata demi kata dari kalimat tersebut. Hasil dari proses *tokenzing* dapat di lihat pada Tabel 4.4.

Tabel 4. 4 Sample Hasil *Tokenizing*

No	Sebelum	Sesudah
1.	x lebih menyenangkan daripada main di dufan	x, lebih, menyenangkan, daripada, main, di, dufan
2.	mana yang mau ke dufan ayo main tornado halilintar kora hysteria baling pokoknya semua yang menantang maut	mana, yang, mau, ke, dufan, ayo, main, tornado, halilintar, kora, hysteria, baling, pokoknya, semua, yang, menentang, maut

Pada Tabel 4.4 menunjukkan hasil pada tahap *Tokenizing* di mana data *tweet* yang awalnya berupa kalimat berhasil dipecah menjadi kata per kata. Adapun karakter pemisah dari tiap kata tersebut ditandai dengan tanda kutip satu yang berada di tiap awal dan akhir huruf dalam tiap kata.

Filtering

Data yang sudah melalui tahap *Tokenizing* kemudian harus di*filter* yakni menyaring kata-kata yang relevan guna proses klasifikasi selanjutnya. . Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan *Stopword.* Contoh kata yang tidak diperlukan yaitu kata yang, dan, di, dari, nya. Pada proses ini akan menggunakan dokumen berisikan kata — kata *stopword*, yang digunakna untuk membantu menyaring kata — kata yang tidak digunakan pada proses klasifikasi. proses ini menggunakan bantuan library nltk yang ada pada bahasa pemrograman *Python*. Hasil dari proses *stopword removal*, dapat dilihat pada Tabel 4.5.

Tabel 4. 5 Sampel Hasil *Filtering*

No	Sebelum	Sesudah
1	x, lebih, menyenangkan, daripada, main, di, dufan	lebih,menyenangkan, main dufan
2	mana, yang, mau, ke, dufan, ayo, main, tornado, halilintar, kora, hysteria, baling, pokoknya, semua, yang, menentang, maut	dufan, ayo, main, tornado, halilintar, kora, baling, pokoknya,semua, menantang maut

Tabel 4.5 menunjukkan hasil pada tahap *Filtering* atau *stopword removal* di mana pada tahap tersebut berhasil dilakukan penghapusan kata-kata konjungsi, yaitu pada contoh diatas kata mana, daripada dan yang.

- ***Stemming***

Hasil terakhir dari *Pre Processing* yaitu *Stemming.* Dilakukannya penghapusan imbuhan dalam sebuah kata yang terdapat pada awal, akhir, ataupun kombinasi dari keduanya Tujuan dari tahap ini yakni untuk mengurangi frekuensi dari sebuah kata turunan. Pada tahap *Stemming* ini menggunakan *library* Sastrawi pada Python. Hasil dari tahap *Stemming* dapat dilihat pada Tabel 4.6 berikut.

Tabel 4. 6 Sample Hasil *Stemming*

No	Sebelum	Sesudah
1.	lebih, menyenangkan, main dufan	lebih, senang, main, dufan
2.	dufan, ayo, main, tornado, halilintar, kora, baling, pokoknya,semua, menantang maut	dufan, ayo, main, tornado, halilintar, kora, baling, pokok,semua, menantang maut

Pada Tabel 4.6 menunjukkan hasil pada tahap *pre-processing* yang terakhir yaitu *Stemming* di mana dilakukan penghapusan imbuhan yang terdapat pada suatu kata baik di awal, akhir, ataupun kombinasidari keduanya.

Setelah dilakukannya tahap *Pre Processing* dataset yang awalnya berjumlah 7995 menjadi 3424, dikarenakan data yang dibuang merupakan data yang tidak diperlukan pada proses klasifikasi. Dapat dilihat pada gambar 4.2 dibawah.

	Tweet
0	tipe mobil sticker mekarsari tmii dufan dll
1	abis main main depan kipas angin dufan
2	bangke muka spek maskot dufan trnyata sorry kb...
3	raya hut oktober xl axiata ajak bestie seru se...
4	chris nggak ketemu dufan ketemu venue aja yaaa...
...	...
3419	beda apa sinting udah grafik source tmii kpop ...
3420	alhamdulillah mudah tasik singaparna gerak tmii
3421	alhamdulillah info baru guruagung tmii
3422	pinter main tmii
3423	gak tmii pdhl yuk tengok

3424 rows × 2 columns

- ***Transformation Data***

Hasil dari melakukan tahap *preproccesing*s selanjutnya akan dilakukannya tahap *transformation* data. Sebelum melakukan *transformation* data dilakukan proses pembobotan sentimen untuk melabelkan data, yang mana pada proses ini dilakukan penghitungan skor sentimen dengan cara mencari jumlah kata bersentimen positif dan kata bersentimen negatif. Skor sentimen tersebut didapatkan dari jumlah polaritas sentimen positif dan negatif yang terdapat pada *dataset.* Kamus *Lexicon* yang digunakan pada penelitian ini adalah kamus *Lexicon* dan *Negative Word* dari penelitian sebelumnya yang dilakukan oleh Koto dan Rahmaningtyas (Koto & Rahmaningtyas, 2017) guna membobot ulang sentimen dari masing- masing *tweet* dan menilai suatu sentimen lebih baik menurut komputer. Hasil dari penambahan kamus *lexicon* dapat dilihat pada Gambar 4.3.

Pada Gambar 4.3 di atas menunjukkan kamus *lexicon* yang ditambahkan Dengan memberi nilai 1 untuk kata yang memiliki kelas positif dan nilai -1 untuk kata dengan kelas negatif. Sedangkan hasil dari penambahan kamus *negative words* terdapat pada Gambar 4.4 berikut

Setelah menambahkan kamus *lexicon* dan *negative words,* maka selanjutnya dilakukan perhitungan skor sentimen dengan menghitung polaritas tiap kalimat yang ada pada data *tweet* tersebut. Hasil dari dilakukannya pembobotan sentimen dengan kamus *lexicon* dan *negative words* tersebut dapat dilihat pada Gambar 4.5 berikut.

	Tweet	label	weight
0	tipe mobil sticker mekarsari tmii dufan dll	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
1	abis main main depan kipas angin dufan	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
2	bangke muka spek maskot dufan trnyata sorry kb...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
3	raya hut oktober xl axiata ajak bestie seru se...	positive	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1...
4	chris nggak ketemu dufan ketemu venue aja yaaa...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
5	maskot dufan	neutral	0 + 0 = 0
6	ken dufan main ken hunting foto sepi hidup woy	positive	0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 = 1
7	hasanah nonton the rose aja kayak kak ngambil ...	positive	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
8	dufan doang pas smp pas sma jogja tp keburu co...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
9	foto dokyeom dufan konsumsi pribadi doang gima...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
10	dufan kak	neutral	0 + 0 = 0
11	ihhh donng main dufan mo wahana alap alap yagak	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
12	bingo coret garis kategori wahana dufan merk c...	positive	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
13	demen banget ngajak dufan siiih vertigo	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
14	udh bgt ngantri udh kaya ngantri wahana dufan ...	negative	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + (1 * -1) + ...
15	duit drmn dufan	neutral	0 + 0 + 0 = 0
16	jisung kaya anak sekolah bolos dufan	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
17	dufan untung pas aneh rem banget anak lepas la...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
18	semi iya pas smp bawa dufan duduk aja nungguin...	negative	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
19	pas bgt yaa liat postingan podcastancur emang ...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0

Gambar 4. 5 Sample Hasil Pembobotan Sentimen

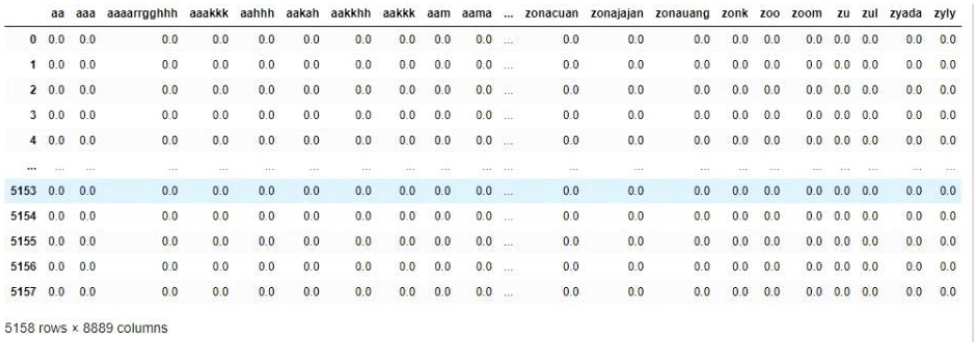
Dapat dilihat pada gambar 4.5 dilakukan perhitungan dengan memberi skor pada setiap kata dalam kalimat dan dijumlahkan semuanya, apabila jumlah dari skor sentimen tersebut nilainya >0 maka sentimen tersebut masuk ke dalam kelas positif, apabila skor sentimen <0 maka kalimat tersebut masuk ke dalam kelas negatif, dan apabila nilai sentimen adalah selain nilai positif dan negatif tersebut maka kalimat sentimen tersebut masuk ke dalam kelas netral. Setelah dilakukan pembobotan sentimen menggunakan kamus lexicon dan negative word, dilakukannya juga validasi dengan pakar Bahasa Indonesia guna memvalidasi hasil pembobotan sentimen yang dilakukan menggunakan kamus. Sehingga ada beberapa sentimen yang berubah selah validasi, berikut hasil rincian jumlah data yang memiliki label positif, negatif, dan netral dari hasil pembobotan sentimen dengan validasi dapat dilihat pada Tabel 4.7.

Tabel 4. 7 Jumlah Pada Setiap Label

Label/Kelas	Jumlah	Validasi
Positif	350	1679
Negatif	301	534
Netral	2773	1211
Total	**3424**	**3424**

Hasil dari pembobotan sentimen dengan kamus *lexicon* dan *negative words* berhasil mengelompokkan sebanyak 5158 ke dalam 3 kelas yaitu 715 data ke dalam kelas positif, 625 data ke dalam kelas negatif, dan 3818 data ke dalam kelas netral. Setelah di validasi oleh pakar Bahasa Indonesia jumlah kelas positif menjadi 1679 data, jumlah kelas negatif menjadi 534 data, dan jumlah kelas netral menjadi 1211 data.

Selanjutnya setelah dilakukan pembobotan sentimen dengan kamus lexicon dan negative words maka akan dilakukan pembobotan kata dengan mengimplementasikan TF-IDF (Term Frequency — Inverse Document Frequency). Fungsi yang digunakan yaitu TfidfVectorizer yang terdapat pada library sklearn. Berikut merupakan hasil dari proses pembobotan TF-IDF yang diterapkan pada dataset yang dapat dilihat pada Gambar 4.6.



	aa	aaa	aaaarrgghhh	aaakkk	aahhh	aakah	aakkhh	aakkk	aam	aama	...	zonacuan	zonajajan	zonauang	zonk	zoo	zoom	zu	zul	zyada	zyly
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5153	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5154	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5155	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5156	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5157	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5158 rows × 8889 columns

Gambar 4. 6 Hasil Pembobotan TF-IDF

Pada Gambar 4.6 memperlihatkan bahwa suatu kata/*term* yang terdapat pada korpus dan diurutkan berdasarkan abjad kemudian dihitung kemungkinan suatu kata tersebut muncul dalam suatu dokumen.

- ***Data Mining***

Setelah melakukan tahap *transformation data* maka setelah itu dilakukan proses *data mining*. Tahap ini merupakan tahap klasifikasi data yang dilakukan dengan mengimplementasikan salah satu model dari algoritma *Naive bayes* yaitu *Multinomial Naive bayes* dengan menggunakan *library* sklearn.naive_bayes yang terdapat pada *Python*. Implementasi algoritma *Multinomial Naive bayes* dapat dilihat pada Gambar 4.7 berikut.

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB().fit(X_train_df, y_train)
prediction_mi = model.predict(X_test_df)
prediction_proba_mi = model.predict(X_test_df)
```

Gambar 4. 7 Library Multinomial Bayes

Dalam penelitian ini digunakan proses pengolahan data menggunakan algoritma *naïve bayes* dengan 4 skenario yang terdiri dari 90:10, 80:20, 70:30 dan 60:40. Hasil sentimen tentang pariwisata di Jakarta pasca covid - 19 ini menggunakan *naïve bayes classifier* sebagai algoritma untuk pengklasifikasian dan *Confusion Matrix* berfungsi untuk menghasilkan nilai akurasi tentang prediksi mesin terhadap data yang sudah diproses. Berikut adalah pembagian data training dan data testing.

Tabel 4. Tabel skenario pembagian *Split data*.

Presentase Data		**Jumlah Data**	
Training	**Testing**	**Training**	**Testing**
90%	10%	3081	343
80%	20%	2739	685
70%	30%	2338	1028
60%	40%	2054	1370
Total		**3424**	

- ***Evaluation***

Setelah proses pembuatan model selesai, model yang telah dibuat dari masing-masing skenario akan diuji. Nilai yang diperoleh dari hasil uji model yaitu nilai accuracy, precision, recall, dan f-measure (f1-score). Berikut perhitungan evaluasi dari setiap model skenario yang diuji.

- **Skenario 1** (90% : 10%)

	actual:negatif	actual:netral	actual:positif
predicted:negatif	4	3	8
predicted:netral	16	72	22
predicted:positif	32	40	146

```
Multinomial NB Accuracy  :  0.6472303206997084
Multinomial NB Precision :  0.5303122972847744
Multinomial NB Recall    :  0.5108518293300902
Multinomial NB F-Measure :  0.5001732454478874
```

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 90% data training dan 10% data testing

.Berdasarkan gambar 4.8, hasil dari klasifikasi menggunakan *Naive bayes* pada perbandingan 90:10 didapatkan hasil dengan nilai akurasi 65%, *precision* 53%, *recall* 51% dan *f-measure* 50%.

- **Skenario 2** (80% : 20%)

	actual:positif	actual:netral	actual:negatif
predicted:positif	8	11	12
predicted:netral	23	129	35
predicted:negatif	72	104	291

```
Multinomial NB Accuracy  :  0.6248175182481752
Multinomial NB Precision :  0.5236768088837701
Multinomial NB Recall    :  0.4891017243549718
Multinomial NB F-Measure :  0.4803307467217815
```

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 80% data training dan 20% data testing.

Berdasarkan gambar 4.9, hasil dari klasifikasi menggunakan *Naive bayes* pada perbandingan 80:20 didapatkan hasil dengan nilai akurasi 62%, *precision* 52%, *recall* 49% dan *f-measure* 48%.

- **Skenario 3** (70% : 30%)

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 70% data training dan 30% data testing.

Berdasarkan gambar 4.10, hasil dari klasifikasi menggunakan *Naive bayes* pada perbandingan 70:30 didapatkan hasil dengan nilai akurasi 63%, *precision* 53%, *recall* 49% dan *f-measure* 48%.

- **Skenario 4** (60% : 40%)

	actual:positif	actual:netral	actual:negatif
predicted:positif	13	16	18
predicted:netral	54	249	66
predicted:negatif	133	211	610

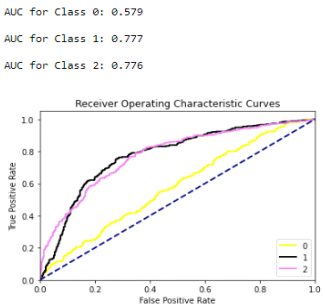
```
Multinomial NB Accuracy  :  0.6364963503649635
Multinomial NB Precision :  0.5302684968506316
Multinomial NB Recall    :  0.48902392657351124
Multinomial NB F-Measure :  0.478301177485565
```

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 60% data training dan 40% data testing

Berdasarkan gambar 4.11, hasil dari klasifikasi menggunakan *Naive bayes* pada perbandingan 60:40 didapatkan hasil dengan nilai akurasi 64%, *precision* 53%, *recall* 49% dan *f-measure* 48%.
Selain dengan menggunakan *confussion matrix* untuk mengukur nilai performansi model, nilai AUC (*Area Under Curve*) dan kurva ROC (*Receiver Operating Characteristics*) juga dapat digunakan dalam mengukur performa suatu model. Nilai AUC pada semua model *Multinomial* NB dapat dilihat pada tabel dibawah ini.

Tabel 4. Hasil Nilai AUC Setiap Skenario

Skenario	AUC
90:10	0.695
80:20	0.706
70:30	0.711
60:40	0.705



Dari tabel diatas hampir semua model skenario menghasilkan nilai AUC dengan kualitas *fair classification.* Nilai AUC terendah terdapat pada model skenario 90:10 dengan 0.695, dan nilai AUC tertinggi terdapat pada model skenario 70:30 dengan 0.711. Berikut grafik ROC dari model dengan nilai AUC tertinggi yang terdapat pada model 70:30.
Pada gambar 4.12 menunjukkan nilai AUC untuk kelas negatif dengan garis warna kuning yaitu 0.579, sedangkan kelas netral dengan garis warna hitam sebesar 0.777, dan kelas positif dengan garis warna merah muda sebesar 0.776. Berdasarkan nilai tiap kelas tersebut, maka ROC AUC *score* pada model, yaitu:
ROC AUC *Score*

Pembahasan

Penelitian analisis sentimen terhadap tempat pariwisata di Jakarta pasca covid-19 menggunakan algoritma *Naïve bayes*. *Metodologi yang digunakan adalah Knowledge Discovey in Database (KDD) yang terdiri dari Data Selection, Pre Processing, Transformation, Data Mining, dan Evaluation.* Teknik pengumpulan data yang digunakan yaitu *Crawling* data dari *Twitter* API dengan *python*. *Tweet* yang berhasil dikumpulkan sebanyak 9729 *tweet* dengan rentang waktu 1 bulan yaitu pada bulan September. Kemudia data di seleksi menjadi 7995 *tweet*.

Selanjutnya setelah data *tweet* dikumpulkan, lanjut ke tahap *Pre Processing* untuk menghapus atau membersihkan data *tweet* dari *noise* dan karakter — karakter yang tidak dibutuhkan saat proses klasifikasi, sehingga setelah melalui tahap pembersihan, data yang digunakan menjadi 3424 data *tweet*. Tahap selanjutnya merupakan tahap transformasi di mana pada tahap ini dilakukan pembobotan suatu *term*/kata yang terdapat pada dokumen dengan menggunakan metode TF-IDF. Namun sebelum mengimplementasikan TF- IDF, dilakukan pembobotan sentimen dengan

menggunakan kamus *lexicon* dan kamus *negative words* terlebih dahulu. Hasil dari pembobotan sentimen tersebut dengan bantuan pakar menghasilkan data dengan data *tweet* positif sebanyak 1679, *tweet* negatif 534, dan *tweet* netral 1211 data dari total *dataset* yaitu 3424 data. Berikut persentase dari ketiga sentimen.

Dari grafik diatas dapat dikatakan sentimen positif lebih banyak dibanding sentimen negatif yang mengartikan bahwa, sentimen pengunjung di media sosial twitter lebih banyak sentimen positif dibanding sentimen negatif dengan perbedaan 2 persen dari total data sentimen. Sentimen netral walaupun mendominasi dari total data tidak berpengaruh terhadap peningkatan kualitas tempat pariwisata di Jakarta.

Setelah pembobotan sentimen dilakukan, maka akan dilakukan proses *Split data* di mana *dataset* dibagi menjadi *data training* dan *data testing* dengan 4 skenario rasio perbandingan. Skenario pertama 90:10, skenario kedua 80:20, skenario ketiga 70:30, dan skenario keempat 60:40.

Proses selanjutnya setelah melakukan *Split data* yaitu melakukan pembobotan kata dengan mengimplementasikan TF-IDF. Setelah model klasifikasi dibuat, maka perlu dilakukannya pengujian performa yaitu dengan menggunakan *Confussion Matrix* dan ROC AUC untuk mengetahui nilai *accuracy, precision, recall, f-measure/f-1 score* dan juga nilai AUC dari model klasifikasi tersebut. Hasil perbandingan pengujian dari *4* model *Multinomial NB* dapat dilihat pada Gambar 4.14 berikut

Gambar 4.14 menunjukan perbandingan antar model skenario yang diuji. Untuk pengujian nilai *accuracy*, yang mempunyai nilai accuracy terendah terdapat pada model 80:20 dengan 62%, sedangkan untuk nilai *accuracy* tertinggi terdapat pada model 90:10 dengan 65%. Pada nilai *precision* terendah terdapat pada model 80:20 dengan 52% dan tertinggi pada Skenario 1, 3 dan 4 dengan 53%. Nilai *recall* tertinggi terdapat pada model 90:10 dengan nilai 51% dan nilai terendah terdapat pada ketiga skenario lainnya dengan 49%. *F-Measure* dengan nilai tertinggi terdapat pada model 90:10 dengan 50%, dan nilai terendah pada skenario yang lainnya yang memiliki masing — masing nilai 48%. Untuk nila AUC terendah ada pada model 90:10 dan tertinggi pada model 70:30. Dari perbandingan skor pengujian skor dari 4 model skenario, model pertama 90:10 mempunyai nilai *accuracy* skor paling tertinggi daripada model yang lain nya.

Wordcloud adalah kata yang sering muncul pada dataset. Semakin besar ukuran kata pada *Wordcloud* maka semakin tinggi frekuensi kata tersebut, artinya kata tersebut sering digunakan oleh masyarakat untuk mengirim cuitan pada *Twitter*. Berikut hasil *wordcloud* untuk tiga sentimen yang ada.

Didalam sentimen positif, kata yang paling sering di *twiit* selain keyword yang di gunakan yaitu main, pas, wahana dan seru.

Didalam sentimen netral, terdapat kata kayak, besok, dan aja yang sering di gunakan untuk mengetwit terkait keyowrd dufan dan tmii.

Di kelas sentimen negatif kata yang sering di twit untuk keyword yaitu dufan dan tmii adalah jelek, banget, kayak, dan gak,

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, diantaranya:

- Penelitian ini melakukan analisis sentimen masyarakat di twitter terhadap pariwisata di Jakarta pasca covid-19 dengan menggunakan algoritma Naive Bayes. Metodologi yang digunakan dalam penelitian ini yakni KDD (Knowledge Discovery in Database) yang terdiri dari 5 tahapan yakni tahap pertama yaitu data selection yang terdiri dari proses crawling data, lalu tahap pre-processing yang terdiri dari 5 proses didalamnya yakni data cleaning, case folding, tokenizing, filtering, dan stemming, selanjutnya tahap transformation yaitu pembobotan kata/term dengan menggunakan TF-IDF, selanjutnya tahap data mining yaitu proses klasifikasi data dengan mengimplementasikan salah satu algoritma naive bayes yaitu multinomial naive bayes, dan tahap terakhir yaitu evaluation di mana pada proses ini dilakukan pengujian dari model dengan melihat nilai parameter pada confussion matrix yakni nilai akurasi, precision, recall, dan juga f-measure/f1-score nya. Hasil pengolahan opini atau sentimen kedalam tiga kelas yaitu Positif, Netral dan Negatif melalui media sosial *Twitter* menggunakan penerapan algoritma *Naïve bayes* diperoleh hasil yaitu 1679 label positif, 534 label negatif, dan 1211 label netral. Dari perolehan sentimen tersebut dapat dikatakan sentimen positif lebih banyak dibanding sentimen negatif yang mengartikan bahwa, sentimen pengunjung di media sosial twitter lebih banyak sentimen positif dibanding sentimen negatif. Sentimen netral walaupun mendominasi dari total data tidak berpengaruh terhadap peningkatan kualitas tempat pariwisata di Jakarta.
- Tingkat akurasi analisis sentimen pada pariwisata di Jakarta pasca Pandemi Covid-19 dengan menggunakan Algoritma *Naïve bayes* diuji menggunakan metode *Split data* dengan model 4 model yaitu model 90:10, 80:20, 70:30, dan 60:40. Dengan menggunakan *confusion matrix* untuk melakukan pengujian skor, nilai *accuracy* tertingi terdapat pada model 90:10. Selain menggunakan *confusion matrix* model diuji dengan grafik ROC yang menghasilkan nilai AUC tertinggi pada model 70:30 dengan nilai 0.711.

DAFTAR PUSTAKA

Adinugroho, S., & Sari, Y. A. (2018). *Implementasi data mining menggunakan WEKA*. Universitas Brawijaya Press.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, *58*(2), 175—191.

Basit, A. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Hasil Panen Padi. *Jurnal Teknik Informatika Kaputama (JTIK) 2020*, *4*(2), 208—213.

Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. *International Conference on Discovery Science*, 1—15. Springer.

Gata, W., & Purnomo, P. P. (2017). Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS. *Format*, *6*(1), 1—13.

Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access*, *8*, 67698—67717.

Kemenkraf. (2020). Tren Pariwisata Indonesia di Tengah Pandemi. Retrieved from kemenparekraf.go.id website: https://kemenparekraf.go.id/ragam-pariwisata/Tren-Pariwisata-Indonesia-di-Tengah-Pandemi

Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *2017 International Conference on Asian Language Processing (IALP)*, 391—394. IEEE.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 538—541.

Kulkarni, A., & Shivananda, A. (2019). *Natural language processing recipes*. Springer.

kumpulaninfo.com. (2021). Dufan Ancol Jakarta. Retrieved from https://kumpulan.info/ website: https://kumpulan.info/wisata/dunia-fantasi

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.

Nugroho, R. A., & Cholissodin, I. (2021). *Implementasi Naïve Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark*. *5*(1), 301—310.

Stewart, F. (2016). *Technology and underdevelopment*. Springer.

Sugiharto, K. R., & Lhaksmana, K. M. (2018). Analisis Sentimen Terhadap Toko Online Menggunakan Naive Bayes Pada Media Sosial Twitter. *EProceedings of Engineering*, *5*(3).

Suntoro, J. (2019). *DATA MINING: Algoritma dan Implementasi dengan Pemrograman php*. Elex Media Komputindo.

Wibowo, A. (2017). Klasifikasi. Retrieved from mti.binus.ac.id website: https://mti.binus.ac.id/2017/11/24/klasifikasi/

Yulita, I. N., Abdullah, A. S., Helen, A., Hadi, S., Sholahuddin, A., & Rejito, J. (2021). Comparison multi-layer perceptron and linear regression for time series prediction of novel coronavirus covid-19 data in West Java. *Journal of Physics: Conference Series*, *1722*(1), 12021. IOP Publishing.