



Jurnal Teknologi Informasi dan Komunikasi

Vol: 13 No 02 2022

E-ISSN: 2477-3255

Diterima Redaksi: 02-10-2022 | Revisi: 02-11-2022 | Diterbitkan: 24-11-2022

Sentiment Analysis of Public Opinion Regarding Papuan Local Languages Condition Using Data Science Approach

Nur Fitriarningsih Hasan^{1*}, Aisyah², Rahman³, Herlin Wonda⁴

¹Program Studi Ilmu Komputer, Universitas Muhammadiyah Papua

²Program Studi Ilmu Komunikasi, Universitas Muhammadiyah Papua

^{3,4}Mahasiswa Program Studi Ilmu Komputer, Universitas Muhammadiyah Papua

^{1,2,3,4}Jl. Abepantai No.25 Tanah Hitam, Abepura, Jayapura, Papua, telp. 0967-582430

e-mail: ¹hi.fitri@umpapua.ac.id, ²aisyah.18@umpapua.ac.id, ³rahmanraghman@gmail.com,

⁴herlinwonda@gmail.com

Abstract

Regional languages can support economic empowerment and improvement through the tourism sector. Opinions from people's expressions in social media and online news collections in reporting the condition of regional languages often become headlines in cyberspace that number in the thousands, which can be used as new knowledge as a basis for making decisions through the mining method. This study aims to explore public opinion sentiment related to the condition of the Papuan language, sourced from text data in cyberspace using a data science approach, namely the classification method with text mining techniques using the naïve bayes algorithm. Public opinion sentiments are processed and the results are presented using word cloud visualization through 4 stages of data science, namely data collection, data preprocessing, modeling exploration and visualization analysis. The result of 778 opinions, 92% tend to have a positive sentiment. The analysis of public opinion sentiment is carried out by the naïve bayes algorithm which has an algorithm model accuracy of 78% and a precision of 88%. The machine learning model that was built and the word cloud visualization analysis succeeded in providing new insights regarding the condition of the Papuan language.

Keywords: public opinion sentiment; papuan language; data science; naïve bayes; word cloud

Analisis Sentimen Opini Publik Terkait Kondisi Bahasa Daerah Papua Menggunakan Pendekatan Data Science

Abstrak

Bahasa daerah dapat mendukung pemberdayaan dan peningkatan ekonomi melalui sektor pariwisata. Opini dari ungkapan masyarakat dalam media sosial dan kumpulan berita online dalam memberitakan kondisi bahasa daerah kerap kali menjadi tajuk dalam dunia maya yang berjumlah ribuan dapat dijadikan sebuah pengetahuan baru sebagai dasar mengambil sebuah keputusan melalui metode mining. Penelitian ini bertujuan mengeksplorasi sentimen opini publik terkait kondisi bahasa daerah Papua, bersumber dari data teks di dunia maya

menggunakan pendekatan data science yaitu metode klasifikasi dengan teknik teks mining menggunakan algoritma naïve bayes. Sentimen opini publik diolah dan hasilnya disajikan menggunakan visualisasi word cloud melalui 4 tahapan data science yaitu pengumpulan data, preprocessing data, eksplorasi modeling dan analisis visualisasi. Hasilnya dari 778 opini, 92% cenderung bersentimen positif. Analisis sentimen opini publik dilakukan oleh algoritma naïve bayes yang memiliki akurasi model algoritma sebesar 78% dan precision sebesar 88%. Model machine learning yang dibangun dan analisis visualisasi word cloud berhasil memberikan insights baru terkait kondisi bahasa Papua.

Kata kunci: sentimen opini publik; bahasa papua; data science; naïve bayes; word cloud

1. Pendahuluan

Bahasa sebagai identitas bagi sebuah bangsa [1] dapat pula mencerminkan kekayaan budaya suatu bangsa [2]. Indonesia memiliki lebih dari satu jenis bahasa yang kerap digunakan dalam komunikasi sehari-hari yaitu Bahasa Indonesia dan bahasa daerah [3]. Bahasa Indonesia adalah bahasa resmi kenegaraan, bahasa pengantar pendidikan dan sarana untuk masyarakat dalam berkomunikasi tingkat Nasional [4], bahasa daerah digunakan sebagai komunikasi antar etnis dan budaya tertentu yang telah berkembang dan menjadi warisan turun temurun masyarakat disuatu tempat[5]. Ethnologue (2016) menyebutkan Indonesia adalah negara kedua dengan bahasa terbanyak di dunia yaitu sejumlah 707 bahasa [6]-[7], kemudian tahun 2018 menjadi berjumlah 688 bahasa [6]. Pada penelitian [8]-[9] menghasilkan *insight* dimana Bahasa daerah dianggap terancam punah apabila kondisi penuturnya berjumlah sedikit serta berumur lebih dari 20 tahun, kemudian para generasi tua sering kali tidak berbicara menggunakan bahasa tersebut kepada keluarga dan lingkungan mereka sendiri [3]-[10].

Provinsi Papua memiliki bahasa daerah terbanyak di Indonesia yaitu sejumlah 326, namun juga termasuk yang mengalami pergeseran juga kepunahan, salah satunya adalah bahasa Nafri [3]. Padahal bahasa daerah dapat mendukung pemberdayaan dan peningkatan ekonomi melalui sektor pariwisata [11], hal ini termasuk dalam renstra Dinas Pariwisata Provinsi Papua 2019-2023 dimana misi pertama adalah memantapkan kualitas dan daya saing SDM dengan sasaran terwujudnya internalisasi nilai adat dan budaya asli tanah Papua dalam berbagai bidang kehidupan masyarakat [12]. Maka pada era teknologi kini tentu bukan sesuatu yang mustahil dalam mewujudkan misi ini. Menurut penelitian [13] upaya pelestarian bahasa daerah dapat dimulai dari kegiatan dan lingkungan akademik[14]. Hal inilah yang menjadi urgensi dalam penelitian ini. Penelitian-penelitian terdahulu melakukan analisis terhadap kondisi bahasa daerah menggunakan metode konvensional yang terbatas pada *resources* dan ruang lingkup. Penelitian kami berfokus pada menganalisa ratusan hingga ribuan data opini publik yang tersebar di internet dari masa lampau hingga kini mengenai bahasa daerah Papua menggunakan pendekatan *data science* dengan *resources* dan ruang lingkup tidak terbatas.

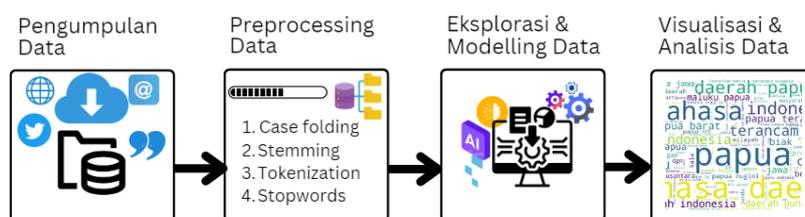
Teknologi yang semakin berkembang dapat menjadi tantangan sekaligus memberikan manfaat khususnya terhadap kondisi bahasa daerah [8]-[15]. Perbincangan dan opini masyarakat dalam bermedia sosial serta isu dan fakta yang dilontarkan oleh pemerhati sastra dan budaya dalam memberitakan kondisi bahasa daerah kerap kali menjadi headline dan perdebatan dalam dunia maya [16]-[17]. Opini dari ungkapan masyarakat pada media sosial serta kumpulan berita pada platform media online dapat menjadi sebuah *insight* untuk memperkaya pengetahuan [18]. Melalui analisis sentimen opini publik dengan memanfaatkan machine learning (ML) yang merupakan salah satu cabang ilmu artificial intelligence (AI), informasi diolah menjadi sebuah

pengetahuan baru terhadap isu yang berkembang [19]. Menggunakan twitter API pada library tweepy dan kata kunci pada portal berita online di google dengan bahasa pemrograman Python 3.9, kemudian diproses dengan teknik text mining menggunakan algoritma ML yaitu naïve bayes. Penelitian sentimen analisis menggunakan algoritma ini rata-rata menghasilkan akurasi 74% [20], 77.2% [21], 86.8% [22], 88.89% [23], dan 86% [24]. Naïve bayes menjadi algoritma yang paling sering digunakan karena kinerjanya yang baik [25]-[26]. Hal inilah yang menjadi dasar dalam memilih algoritma pemodelan ML untuk eksplorasi data dan visualisasi data science.

Penelitian ini bertujuan mengeksplorasi mendalam terkait kondisi bahasa daerah Papua bersumber dari data *text* di dunia maya tahun 2017-2022, menghasilkan model ML untuk analisis sentimen opini publik menggunakan pendekatan *data science*. Sentimen opini publik diolah dan hasilnya disajikan menggunakan visualisasi data berupa grafik dan *word cloud*.

2. Metode Penelitian

Penelitian ini menggunakan tahapan metodologi *data science* seperti terlihat pada Gambar 1. *Data science* adalah kemampuan melakukan pengolahan data untuk menghasilkan *insights* melalui visualisasi data yang tepat dari data kuantitatif yang terstruktur maupun tidak terstruktur [27]-[28], menggabungkan kemampuan statistik, *programming* dan pengetahuan bisnis serta kemampuan komunikasi (*storytelling*) [29].



Gambar 1. Metodologi data science

Pada Gambar 1 merupakan diagram alir penelitian analisis sentimen opini publik terkait kondisi bahasa daerah Papua menggunakan pendekatan *data science*, yaitu metode klasifikasi dengan teknik teks mining menggunakan algoritma naïve bayes. Terdapat empat tahapan yaitu pengumpulan data, *preprocessing data* yaitu mempersiapkan data mentah menjadi dataset sebelum dilakukan tahap selanjutnya yaitu eksplorasi dan *modelling data*. Tahap eksplorasi dan *modelling data* dengan cara membagi data menjadi data uji data latih untuk selanjutnya diproses oleh ML menggunakan algoritma naïve bayes. Hasil dari model ML berupa data yang telah bersentimen positif dan negatif. Data *modelling* selanjutnya menghasilkan *term frequency word*, *TF-word* ini dibuatkan visualisasi *word cloud* untuk dilakukan analisis dan penarikan kesimpulan atau *insights* terkait kondisi bahasa daerah Papua. Tools yang digunakan pada proses pengambilan data, *preprocessing*, eksplorasi, *modelling* ML, dan analisis visualisasi *word cloud* menggunakan bahasa pemrograman Python 3.9 (Jupyter). Sementara tampilan visualisasi seperti *bar chart*, *line chart*, *donut/pie chart* menggunakan Microsoft Power BI.

2.1. Pengumpulan Data

Data scrapping adalah proses pengumpulan data dan informasi dalam bentuk yang lebih spesifik sesuai dengan topik atau *keyword* yang ditentukan [30]. *Flowchart* metodologi pengumpulan data terdapat pada Gambar 2.

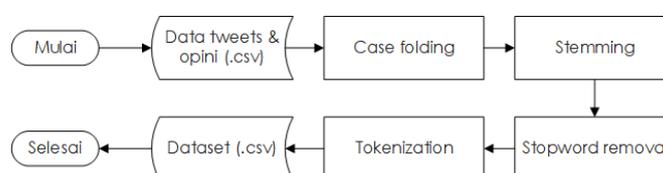


Gambar 2. Flowchart pengumpulan data

Keyword pada penelitian ini menggunakan #BahasaDaerah, #BahasaDaerahPapua, #BahasaPapua dan #KondisiBahasaPapua dalam 5 tahun terakhir yaitu tahun 2017-2022. Pada tahap ini data yang dikumpulkan adalah data teks tidak terstruktur (.csv) bersumber dari sosial media twitter dan opini-opini publik pada platform portal berita online maupun website lainnya di google.

2.2. Preprocessing Data

Pada tahap kedua data yang didapatkan merupakan data kompleks yang tidak lengkap, berantakan dan tidak akurat, sehingga perlu dilakuakn *preprocessing* terlebih dahulu dengan metode *data cleaning* [31]. Gambar 3 adalah *flowchart preprocessing data*.



Gambar 3. Flowchart preprocessing data

Proses pembersihan data yang dilakukan adalah *case folding*, *stemming*, *stopword removal*, dan *tokenizing* [32]. *Natural Language Toolkit* (NLTK) adalah *library* Python untuk bekerja dengan pemodelan teks [33]. Setelah melalui tahap *preprocessing* data, selanjutnya akan digunakan sebagai dataset untuk dilakukan eksplorasi dan *modelling*.

a) *Case folding*

Menghapus karakter-karakter khusus pada sebuah string atau kalimat misalnya koma (,), tanda petik (‘), tanda tanya (?) dan karakter lainnya.

b) *Stemming*

Proses ini berupa penghilangan semua imbuhan (*affix*) yang terdiri dari awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan duplikasi. Proses *stemming* dilakukan untuk mendapatkan kata dasar dari kata berimbuhan.

c) *Stopwords*

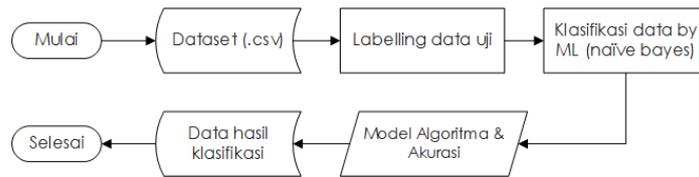
Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna untuk digunakan di *text mining*, maka kata-kata umum ini dihapus.

d) *Tokenization*

Tokenisasi adalah proses pemotongan seluruh urutan karakter menjadi satu potongan kata. Proses ini untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token menggunakan modul NLTK.

2.3. Eksplorasi dan Modelling Data

Pada tahap ini ada dua kegiatan utama yaitu eksplorasi dan *modelling* data. Dataset yang sudah siap diolah selanjutnya dibagi menjadi dua bagian yaitu data uji dan data latih dengan perbandingan sekitar 20:80. *Flowchart* kegiatan eksplorasi dan *modelling* data ada pada Gambar 4.



Gambar 4. Flowchart eksplorasi dan modelling data

Pada data uji dilakukan *labelling* secara manual, yaitu label positif (1) dan label negatif (0). Label positif seperti kata-kata baik, bertahan, lestari, keren, bertahan, upaya, dan lain lain. Sementara label negatif yaitu yang mengandung kata jelek, punah, hilang, bodoh dan sebagainya. Selanjutnya oleh ML data latih akan diklasifikasi menggunakan algoritma naïve bayes berdasarkan label yang ada pada data uji. Rumus algoritma naïve bayes yang digunakan untuk klasifikasi seperti pada persamaan (1) dan persamaan (2) [34].

$$P(v_j) = \frac{|doc_i|}{|training|} \quad (1)$$

$$P(a_i|v_j) = \frac{n_i+1}{n+|kosakata|} \quad (2)$$

$P(v_j)$ yaitu probabilitas setiap dokumen pada sekumpulan dokumen, $P(a_i|v_j)$ yaitu probabilitas kemunculan kata a_i pada suatu dokumen dengan kategori kelas v_j , $|doc_i|$ yaitu frekuensi dokumen pada setiap kategori, $|training|$ yaitu jumlah dokumen latih (*training*) yang ada, n_i frekuensi kata ke-k pada setiap kategori dan $|kosakata|$ adalah jumlah kosakata yang ada pada dokumen uji.

Setelah dilakukan *training*, hasil klasifikasi dari ML yaitu berupa model algoritma dan akurasi terhadap model. Akurasi yang digunakan pada kegiatan ini adalah *confusion matrix*. *Confusion Matrix* merupakan metode evaluasi yang dapat digunakan untuk menghitung kinerja algoritma dari proses klasifikasi, perhitungan dan pencarian hasil akurasi model menggunakan persamaan (3) [21].

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Pada persamaan (3) dimana TP (*true positive*) adalah hasil prediksi positif dan itu benar, TN (*true negatif*) adalah hasil prediksi negatif dan itu benar, FP (*false positive*) adalah hasil prediksi positif dan itu salah, FN (*false negatif*) adalah hasil prediksi negatif dan itu salah. Model algoritma terbaik yang akan menjadi dasar dalam membuat visualisasi *word cloud* [35].

2.4. Analisis dan Visualisasi Data

Word cloud adalah gambar yang menunjukkan daftar kata-kata didalam sebuah teks dalam bentuk abstraksi visual yang intuitif. *Word cloud* juga dapat membantu dalam membandingkan dan membedakan dua bagian teks yang berbeda untuk menemukan kesamaan kata di antara keduanya [19]. Langkah awal yang dilakukan adalah melakukan instalasi *library word cloud* [36], dibutuhkan *library* tambahan lainnya seperti *numpy*, *pandas*, *matplotlib* dan *image*.

Visualisasi *word cloud* dihasilkan melalui perhitungan frekuensi *terms* dan pembobotan tiap-tiap *terms*. Setiap kata dianggap sebagai token dan dihitung berapa sering suatu kata muncul [37]. Semakin sering suatu kata muncul, maka semakin besar ukuran kata yang tampil dalam *word cloud* [38]. Formula pembobotan *term frequency (TF)* pada *word cloud* ditunjukkan pada persamaan (4) [33].

$$TF(i, j) = \frac{\text{term } i \text{ frequency in document}}{\text{total words in document}} \quad (4)$$

Dimana, *i* adalah *term* dan *j* adalah dokumen.

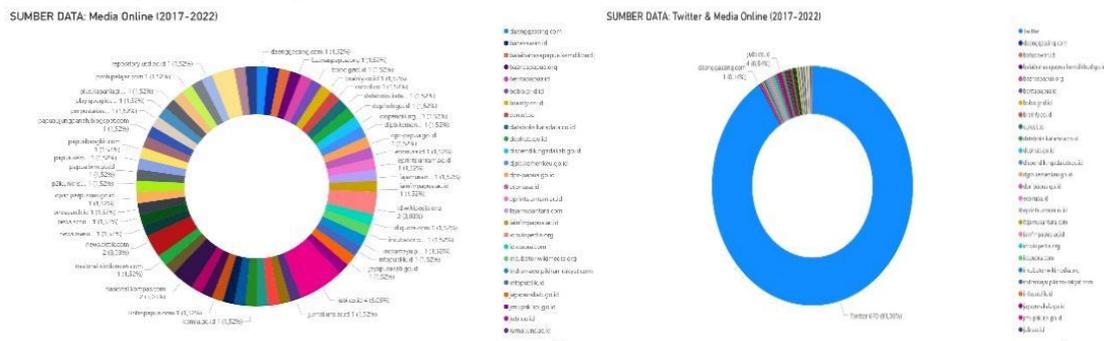
3. Hasil dan Pembahasan

Bagian ini menjelaskan hasil dari penelitian yang telah dilakukan. Hasil yang didapatkan dielaborasi sesuai dengan tahapan pada metode penelitian.

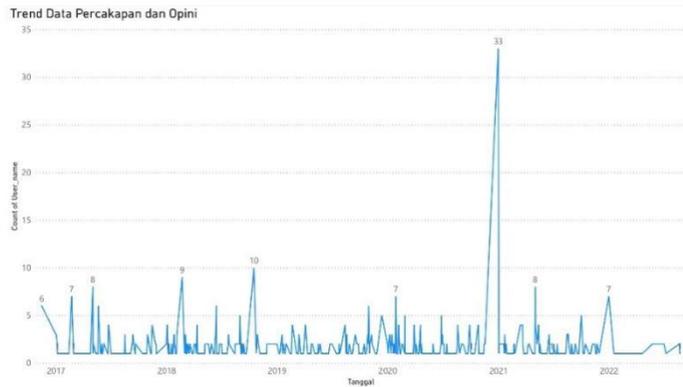
3.1. Hasil Pengumpulan Data

Data untuk analisa diambil dari *twitter* dan *platform* berita *online* serta *website* yang ada di *google*, mulai tahun 2017 hingga tahun 2022 menggunakan kata kunci #bahasapapua, #bahasadaerahpapua. Data yang didapatkan sebanyak 1,156 data. Data yang diambil untuk dianalisa terdiri dari *tweet* dan kalimat atau tulisan teks dari platform *tweeter*, *headline news* pada situs *platform* berita *online*, dan *tag key* pada *website* yang ada di *google*. Gambar 5 dan Gambar 6 merupakan hasil visualisasi sumber data dari hasil pengumpulan data.

Pada Gambar 5 dan 6 dapat dilihat bahwa opini terkait #BahasaDaerah, #BahasaDaerahPapua, #BahasaPapua dan #KondisiBahasaPapua lebih banyak ditemukan pada *platform* media sosial yaitu *twitter* sejumlah 720 data atau 91,03%. Sedangkan sisanya pada situs berita *online* kemunculan kata kunci ini memiliki rata-rata 1-3 data saja per situs (1,52%–3,03%) namun jumlah situs berita *online* yang memberitakan opini terkait bahasa daerah Papua sangat banyak yakni sejumlah 436 data dari portal berita. Jika dibandingkan data yang terkumpul dari *twitter* dan situs berita *online* yaitu 9:1 data *twitter* lebih banyak menampilkan opini terkait bahasa daerah Papua. Situs yang cukup sering mengangkat opini ini adalah *twitter.com*, *jubi.co.id*, *news.detik.com*, *nasional.kompas.com*, *jayapurakab.go.id* dan diikuti 1,52% situs lainnya. Situs-situs ini muncul karena kata kunci yang dicari terdapat pada situs tersebut, sehingga dalam penelitian ini tidak dilakukan skenario penentuan awal dari situs mana yang akan dilakukan penarikan data. *Trend* percakapan dan opini tinggi berbentuk grafik ditampilkan Gambar 7.



Gambar 5. Data pada portal berita *online* **Gambar 6.** data dari *twitter* vs berita *online*



Gambar 7. Grafik tren data percakapan dan opini publik

Pada Gambar 7 terlihat ada sembilan kali percakapan dan opini tertinggi terkait bahasa daerah Papua yaitu pada tahun 2022 sebanyak 1 kali sejumlah 7 opini, tahun 2021 sebanyak 2 kali sejumlah 41 opini, tahun 2020 sebanyak 1 kali sejumlah 7 opini, tahun 2019 sebanyak 1 kali sejumlah 19 opini, tahun 2018 sebanyak 1 kali sejumlah 9 opini dan tahun 2017 sebanyak 3 kali sejumlah 21 opini. Jika dilihat pada grafik gambar 7, terjadi percakapan atau opini tertinggi terkait bahasa daerah Papua pada tahun 2021 sebanyak 33 data. Hal ini mengartikan terdapat 33 user yang membahas terkait bahasa Papua. Pada tanggal 1 Januari 2021 percakapan tinggi pertama bersumber dari tweet @kamusbesarkbbi: “keberlanjutan revitalisasi bahasa daerah papua tergantung kebijakan jurnal soreang jurnal soreang <https://journalsoreang.com> pikiran rakyat pendidikan pr 1015011920 keberlanjutan revitalisasi bahasa daerah papua tergantung kebijakan kamus kbbi bahasa”, kemudian disambung dengan reply ataupun retweet oleh @sundingjalila: “jujur langkah bagus sih yg lakukan dinas pendidikan papua dimana murid paud kelas 1 sd menerapkan muatan lokal pembelajaran bahasa daerah jayawijaya kalo sd sma menerapkan biar bahasa daerah jaga”, dan dilanjutkan oleh 31 akun lainnya.

3.1. Hasil Preprocessing Dataset

Hasil pengumpulan data selanjutnya dilakukan *preprocessing* dan *cleaning* yaitu untuk menghilangkan data terduplikat, redundansi, imbuhan tidak penting dan lainnya. Proses *data cleaning* dapat dilihat pada Gambar 8 dan 9.

```

In [82]: # stop removal
# stop_remove('stopwords')
stopword = nltk.corpus.stopwords.words('indonesian')

def remove_stopwords(text):
    text = [word for word in text if word not in stopword]
    return text

df['STOP_REMOVAL'] = df['TOKENIZATION'].apply(lambda x: remove_stopwords(x))
df.head(5)
    
```

id	nama	username	isi	tanggal	TWEETS	NEW_USERNAME	TOKENIZATION	STOP_REMOVAL
0	Fakta Google	fatagoogle	kata papua berasal dari bahasa meyo yang artinya...	27/01/2017	fatagoogle		[kata, papua, berasal, dari, bahasa, meyo, yang, artinya, ...]	[kata, papua, berasal, meyo, yang, arti, ...]
1	Fakta Google	fatagoogle	pada night merupakan negara dengan jumlah ba...	24/01/2017	fatagoogle		[pada, night, merupakan, negara, dengan, jumlah, ba, ...]	[pada, night, merupakan, negara, dengan, jumlah, ba, ...]
2	Uniqke Shop	dagangunqke	kumpulan kata kata bahasa papua lucu dan meng...	23/01/2017	dagangunqke		[kumpulan, kata, kata, bahasa, papua, lucu, dan, meng, ...]	[kumpulan, kata, kata, bahasa, papua, lucu, dan, meng, ...]
3	Jaharmanan	ididididididid	kebanggaan bahasa papuan papua	19/01/2017	ididididididid		[kebanggaan, bahasa, papuan, papua]	[kebanggaan, bahasa, papuan, papua]
4	ini	iniini	kebanggaan bahasa papua dan hayoo	25/01/2017	iniini		[kebanggaan, bahasa, papua, dan, hayoo]	[kebanggaan, bahasa, papua, dan, hayoo]

Gambar 8. Proses stemming, case folding, stopwords dan tokenisasi

```

In [79]: # fungsi cleaning text tweet
def remove_punct(text):
    text = re.sub('[^a-zA-Z0-9]', '', str(text))
    text = re.sub('http://.*', '', text)
    text = re.sub('@.*', '', text)
    return text

# df['CLEANING'] = df['tweet'].apply(lambda x: remove_punct(x))
df['CLEANING'] = df['tweet'].apply(lambda x: remove_punct(x))

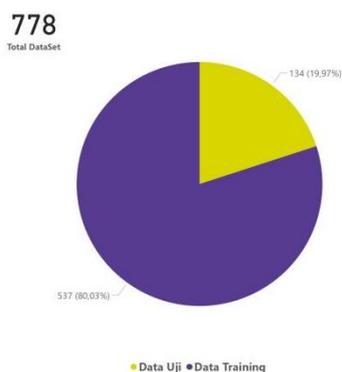
In [80]: # fungsi cleaning username
def cleaning_text(text):
    text = re.sub('[@]', '', text)
    return text

df['NEW_USERNAME'] = df['Username'].apply(lambda x: cleaning_text(x))
df.head(5)
    
```

id	nama	username	isi	tanggal	TWEETS	NEW_USERNAME
0	Fakta Google	fatagoogle	kata papua berasal dari bahasa meyo yang arti...	27/01/2017	kata papua berasal dari bahasa meyo yang arti...	fatagoogle
1	Fakta Google	fatagoogle	pada night merupakan negara dengan jumlah ba...	24/01/2017	pada night merupakan negara dengan jumlah ba...	fatagoogle
2	Uniqke Shop	dagangunqke	kumpulan kata kata bahasa papua lucu dan meng...	23/01/2017	kumpulan kata kata bahasa papua lucu dan meng...	dagangunqke
3	Jaharmanan	ididididididid	kebanggaan bahasa papuan papua	19/01/2017	kebanggaan bahasa papuan papua	ididididididid
4	ini	iniini	kebanggaan bahasa papua dan hayoo	25/01/2017	kebanggaan bahasa papua dan hayoo	iniini

Gambar 9. Proses data cleaning

Data awal sejumlah 1,156 yang telah melalui proses *data cleaning* dan *preprocessing* menjadi sejumlah 778 dataset. Jumlah dataset dibagi menjadi dua yaitu data uji dan data latih dengan perbandingan 20:80 visualisasi hasil pembagian data dapat dilihat pada Gambar 10.

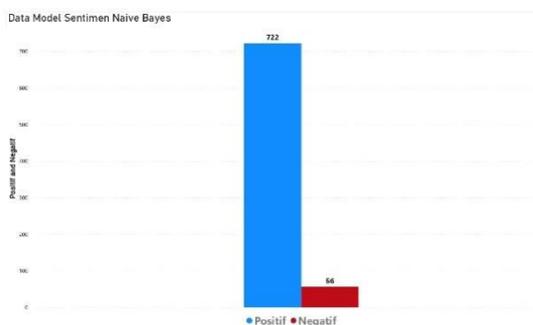


Gambar 10. Dataset

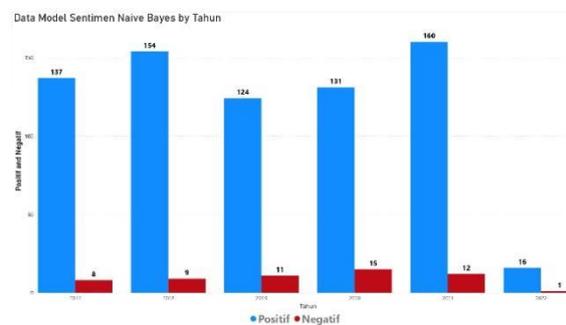
Pembagian dataset menjadi 537 data latih (80,03%) dan 134 data uji (19,97%). Pengklasifikasian (*labelling*) sentimen secara manual dilakukan pada data uji, dimana data uji ini akan menjadi data pembanding untuk proses klasifikasi naïve bayes pada data latih. Data atau kalimat yang mengandung kata negatif berlabel negatif (0) sementara yang positif berlabel positif (1).

3.2. Eksplorasi Data dan Validasi Model Machine Learning Analisis Sentimen

Dataset yang berjumlah 778 data sentimen opini publik terbagi ke dalam 2 klasifikasi, yaitu positif dan negatif. Hasilnya sebanyak 722 positif (92%) dan 56 negatif (8%). Gambar 11 merupakan hasil analisis sentimen opini publik terhadap kondisi bahasa daerah Papua secara keseluruhan. Gambar 12 merupakan hasil analisis sentimen opini publik terhadap kondisi bahasa daerah Papua per tahunnya selama 5 tahun terakhir (2017-2022).



Gambar 11. Data analisis sentimen kondisi bahasa Papua keseluruhan



Gambar 12. Data sentimen kondisi bahasa daerah Papua per tahun

Hasil analisis pada Gambar 11 dan 12 merupakan hasil sentimen yang dianalisa langsung menggunakan algoritma naïve bayes dengan melihat seberapa banyak kata-kata bernuansa negatif dalam sebuah kalimat atau opini menggunakan persamaan (1) dan (2). Kata berlabel negatif dan positif berasal dari data uji yang telah dilabeli secara manual. Data uji menjadi data pembanding untuk data latih saat proses klasifikasi oleh algoritma naïve bayes. Analisis sentimen secara keseluruhan terhadap kondisi bahasa daerah cenderung menghasilkan opini positif 722 dan negatif 54, artinya sentiment opini publik cenderung positif. Sementara untuk data sentimen per tahunnya di tahun 2017 terdapat 137 positif dan 8 negatif, tahun 2018 terdapat 154 positif dan 9 negatif, tahun 2019 terdapat 124 positif dan 11 negatif, tahun 2020 terdapat 131 positif dan 15 negatif,

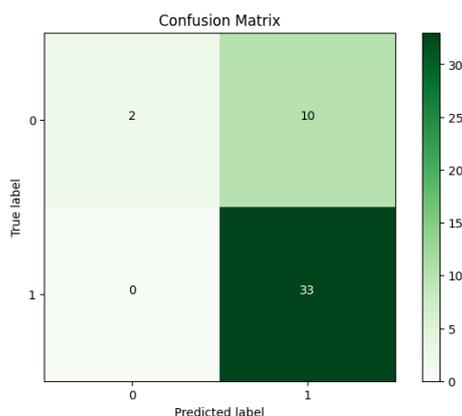
tahun 2021 terdapat 160 positif dan 12 negatif, tahun 2022 terdapat 16 positif dan 1 negatif. Jika dilihat dari grafik Gambar 11 secara keseluruhan sentimen terhadap kondisi bahasa daerah Papua lebih dominan positif, jika dilihat per tahunnya jumlah sentiment positif dan negatif mengalami kenaikan dan penurunan jumlah seperti yang terjadi dari tahun 2017-2021. Hal ini akan bisa lebih di analisis mendalam dengan analisis visualisasi *word cloud*.

Proses dan hasil klasifikasi sentimen menggunakan naïve bayes terhadap data dapat dilihat pada Gambar 13.

Tanggal	User_name	isi	Sentimen	TOKENIZATION	platform
Jumat, 01 Januari 2021	mamedrifqy	anaknya video menceritakan bahasa daerah papua terancam puna	0	['ke', 'anaknya', 'video', 'ini', 'sedikit', 'menceritakan', 'baga']	Twitter
Rabu, 01 Januari 2020	idntimesjogja	kepala program studi magister ilmu linguistik fakultas ilmu budaya	0	['menurut', 'kepala', 'program', 'studi', 'magister', 'ilmu', 'li']	Twitter
Senin, 16 Agustus 2021	hascaryo_w	geram kalo orang berasal papua orang indonesia salah orang papu	0	['ini', 'saya', 'selalu', 'geram', 'kalo', 'ada', 'orang', 'yang', 'b']	Twitter
Jumat, 27 Januari 2017	faktagoogle	papua berasal bahasa melayu rambut keriting	1	['kata', 'papua', 'berasal', 'dari', 'bahasa', 'melayu', 'yang', '']	Twitter
Selasa, 24 Januari 2017	faktagoogle	papua nugini negara bahasa 820 bahasa indonesia 500an bahasa	1	['papua', 'nugini', 'merupakan', 'negara', 'dengan', 'jumlah']	Twitter
Senin, 23 Januari 2017	dagangunique	kumpulan bahasa papua lucu bergambar http beritaheboh info ku	1	['kumpulan', 'kata', 'kata', 'bahasa', 'papua', 'lucu', 'dan', '']	Twitter

Gambar 13. Hasil klasifikasi sentimen

Pada Gambar 13 dapat dilihat tweet dari @mamedrifqy, @idntimesjogja, @hascaryo_w merupakan sentimen negatif (0) yang mengandung kata-kata diantaranya “ke”, “video”, “sedikit”, “geram”, “punah” dan lain-lain. Sementara tweet oleh akun @faktagoogle, @dagangunique merupakan sentiment positif (1) yang mengandung kata “papua”, “bahasa”, “melayu”, dan lain-lain. Validasi model ML untuk klasifikasi berguna agar memperkuat hasil kecenderungan sentimen opini dapat dilihat pada hasil pengujian model naïve bayes Gambar 14 dan Tabel 1.



Gambar 14. Hasil validasi confusion matrix

Pada Gambar 14 hasil akurasi dihitung menggunakan persamaan (3) menghasilkan akurasi sebesar 78%, hasil akurasi ini sudah cukup baik karena jika merujuk pada penelitian terdahulu akurasi dibawah 70% yaitu 66% seperti pada yang dilakukan oleh Ihsan dimana jumlah data latih memiliki hubungan dengan hasil akurasi. Kurangnya jumlah data latih mengandung sentimen menyebabkan model kurang akurat mendeteksi opini positif dan negatif [39]. Maka untuk menaikkan akurasi model ML diperlukan tambahan lebih banyak kata-kata sentimen negatif dan positif lebih spesifik dan diperlukan lebih banyak dataset opini, agar model hasil pelatihan lebih optimal lagi. Kemudian menurut Indrayuni nilai *presicion* dapat memprediksi ketepatan data dari suatu sistem [40]-[41]. Pada Tabel 1 diketahui *presicion* positif pada model klasifikasi ML naïve bayes ini lebih banyak persentasenya dibanding *presicion* negatifnya pada, yaitu 100%, jika dihitung nilai rata-rata akurasi *precision* adalah 88%.

Tabel 1. Hasil validasi model klasifikasi

Representasi Klasifikasi	ML	
	Positif	Negatif
Precision	100%	77%
Recall	17%	100%
F1-score	29%	87%

3.3. Hasil Analisis Visualisasi Word Cloud

Word cloud dikenal sebagai *tag cloud* atau *text cloud*, ini adalah cara untuk menarik bagian paling relevan dari data tekstual [33]. Visualisasi data dengan *word cloud* dilakukan berdasarkan kelas sentimen hasil klasifikasi opini oleh ML menggunakan python 3.9. Langkah awal yang dilakukan adalah melakukan instalasi *library word cloud* dengan cara `pip install word cloud`. Hasil visualisasi *word cloud* untuk sentimen opini cenderung positif dapat dilihat pada Gambar 15. Tabel 2 merupakan rincian jumlah kata yang muncul dari visualisasi *word cloud* dihitung menggunakan persamaan (4).



Gambar 15. Visualisasi word cloud positif

Tabel 2. Jumlah kemunculan kata sentimen positif

Kata	Σ Kemunculan
Bahasa	1284
Papua	1025
Daerah	553
Punah	205
Pendidikan	21
...	...
Revitalisasi	7

Dilihat dari kumpulan kata-kata yang paling sering muncul di *word cloud* sentimen positif pada gambar 15, kata-kata seperti bahasa, papua, daerah, punah, pendidikan, masyarakat, nasional, revitalisasi sangat sering muncul dapat dilihat pada Tabel 2. Hal ini dapat menunjukkan dan memberikan arti percakapan yang sering terjadi di dunia maya baik *platform* media sosial twitter maupun situs berita *online* yang sebagian besar masyarakat sadar akan kondisi bahasa daerah Papua yang terancam punah dan mendukung revitalisasi ataupun pemertahanan posisi serta penggunaan bahasa daerah Papua baik dalam skala lokal maupun Nasional. Upaya-upaya ini dapat dimulai dari sektor pendidikan. Analisis ini menjadi berhubungan dengan penelitian sebelumnya yang dilakukan oleh Windi dimana melalui pembiasaan dan arahan yang dilakukan dilingkungan akademik menggunakan teknologi atau media berbasis multimedia dapat membentuk perilaku positif, menumbuhkan pemahaman, menjadi pondasi awal setiap individu dalam membangun kesadaran menjaga kelestarian bahasa daerah agar tetap bertahan dalam perubahan jaman di era digital, serta menumbuhkan rasa memiliki bahasa daerah yang harus dilestarikan dengan cara menggunakannya dalam kehidupan sehari-hari dalam bermasyarakat [14]. Hasil visualisasi *word cloud* untuk sentimen negatif dapat dilihat pada Gambar 16 dan Tabel 3.

Pada Gambar 16 merupakan visualisasi hasil dari kumpulan kata-kata yang paling sering muncul di *word cloud* dengan sentimen negatif. Pada Tabel 3 kata-kata seperti papua, bahasa, orang, jawa, hilang, pemuda sering muncul. Hal ini menunjukkan percakapan yang sering terjadi di twitter dan situs berita *online* lainnya sebagian besar menunjukkan adanya perdebatan, kekecewaan dan kekhawatiran terhadap kondisi bahasa daerah Papua. Bukan hanya bahasa daerah Papua yang terancam hilang atau punah, namun bahasa daerah lain juga seperti bahasa daerah Jawa, Melayu dan Sulawesi. Artinya tidak hanya kondisi bahasa daerah Papua yang sedang mengalami pergeseran bahkan kepunahan namun juga bahasa daerah lain. Penelitian oleh Rabiah pun mendukung hasil analisis ini, dimana bahasa daerah Makassar semakin terlupakan di kalangan keluarga modern. Kurangnya penguasaan orang tua terhadap bahasa

dilakukan mulai dari dunia pendidikan, lingkungan akademik dan dilestarikan ditengah masyarakat. Upaya-upaya ini tentu dapat dilakukan dengan memanfaatkan teknologi, seperti teknologi internet, teknologi multimedia hingga teknologi AI.

4. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan dari pengambilan data opini di media sosial twitter dan situs berita *online* dalam 5 tahun terakhir (2017-2022) sejumlah 778 opini 92% cenderung bersentimen positif. Analisis sentimen opini publik dilakukan oleh algoritma ML naïve bayes menunjukan akurasi kinerja algoritma *machine learning* sebesar 78% dan *precision* sebesar 88%. Hasil visualisasi *word cloud* memberikan *insights* bahwa masyarakat Indonesia menyadari kondisi bahasa daerah khususnya bahasa daerah Papua mengalami kepunahan dan pergeseran di masyarakat. Menurut hasil olah opini perlu dilakukan upaya-upaya pemertahanan ataupun pelestarian bahasa daerah yang dapat dimulai dari bidang pendidikan dan akademik dengan memanfaatkan teknologi sesuai pada era digital saat ini.

Penelitian selanjutnya sebaiknya menganalisis penyebab utama dan hubungan tiap aspek variabel terkait kepunahan bahasa daerah Papua dari opini masyarakat melalui berbagai pendekatan. Setelah itu mulai melakukan kegiatan dalam rangka upaya pemertahanan bahasa daerah Papua melalui pendekatan digitalisasi ataupun teknologi lainnya. Selain itu tidak cukup sampai pada tahap penelitian, hasil analisis sebaiknya dapat dikembangkan hingga menjadi sebuah prototipe yang dapat diimplementasikan dan dirasakan manfaatnya secara langsung oleh masyarakat. Perlu adanya kerja sama dengan instansi lain terkait untuk upaya ini, seperti balai bahasa, dinas pariwisata dan pemerintah daerah.

Ucapan Terima Kasih

Kegiatan penelitian ini sepenuhnya dibiayai oleh DRTPM Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi berdasarkan Surat Keputusan nomor 0267/E5/AK.04/2022 dan kontrak nomor 166/E5/PG.02.00.PT/2022 melalui skema hibah penelitian kompetitif Nasional—PDP. Penulis mengucapkan terima kasih kepada DRTPM Ditjen Diktiristik, LLDIKTI wilayah XIV Papua—Papua Barat dan LPPM Universitas Muhammadiyah Papua atas dukungan yang diberikan. Semoga penelitian ini memberikan kontribusi ilmu yang baik dan dapat bermanfaat seluas-luasnya.

Daftar Pustaka

- [1] H. M. . Nahak, “Upaya Melestarikan Budaya Indonesia Di Era Globalisasi,” *J. Sociol. Nusant.*, vol. 5, no. 1, pp. 65–76, 2019, doi: 10.33369/jsn.5.1.65-76.
- [2] L. P. S. Antari, “Bahasa Indonesia Sebagai Identitas Nasional Bangsa Indonesia,” *J. Stilistika*, vol. 8, no. 1, pp. 23–29, 2019, [Online]. Available: <http://ejournal.unibba.ac.id/index.php/jisipol/article/view/115>.
- [3] M. Mujizah, “Preservasi Dan Revitalisasi Bahasa Dan Sastra Nafri, Papua: Sebuah Bahasa Hampir Punuh,” *Aksara*, vol. 30, no. 1, p. 75, 2018, doi: 10.29255/aksara.v30i1.217.75-88.
- [4] N. F. Hasan and M. S. A. Iribaram, “Digitizing the Papuan Regional Language Dictionary using the Rapid Application Development Method,” *Matrik J. Manag. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 710–720, 2022, doi: 10.30812/matrik.v21i3.1688.

<https://doi.org/10.31849/digitalzone.v13i2.11545>

-
- [5] M. Nurlaila, "Pengaruh Bahasa Daerah (Ciacia) terhadap Perkembangan Bahasa Indonesia Anak Usia 2 Sampai 6 di Desa Holimombo Jaya," *J. Retorika*, vol. 9, no. 2, pp. 90–163, 2016.
- [6] A. Lidwina, "Papua punya bahasa daerah terbanyak di Indonesia," 2019. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2021/05/14/papua-punya-bahasa-daerah-terbanyak-di-indonesia#>.
- [7] F. H. Tondo, "Kepunahan Bahasa-Bahasa Daerah: Faktor Penyebab Dan Implikasi Etnolinguistik," *J. Masy. Budaya*, vol. 11, no. 2, pp. 277–296, 2009, [Online]. Available: [jmb.lipi.go.id > index.php > jmb > article > download%0A](http://jmb.lipi.go.id/index.php/jmb/article/download/0A).
- [8] S. Wilian, "Pemertahanan Bahasa Dan Kestabilan Kedwibahasaan Pada Penutur Bahasa Sasak Di Lombok," in *Linguistik Indonesia*, 2010, no. 1, pp. 23–39.
- [9] I. Ibrahim, R. Ruslan, M. N. A. Asnur, Y. N. Sabata, and M. S. Kahar, "Faktor Sosial Yang Berpengaruh Terhadap Pergeseran Bahasa Lowa," *KEMBARA J. Sci. Lang. Lit. Teach.*, vol. 5, no. 2, p. 208, 2019, doi: 10.22219/kembara.vol5.no2.208-218.
- [10] S. Rabiah, "Revitalisasi Bahasa Daerah Makassar Melalui Pengembangan Bahan Ajar Bahasa Makassar Sebagai Muatan Lokal," *Din. Ilmu*, vol. 13, no. 1, pp. 51–66, 2013.
- [11] I. Pujiningrum Palimbunga, "Keterlibatan Masyarakat Dalam Pengembangan Pariwisata Di Desa Wisata Tabalansu, Papua," *J. Master Pariwisata*, vol. 05, p. 193, 2018, doi: 10.24843/jumpa.2018.v05.i01.p10.
- [12] Dinas Pariwisata Provinsi Papua, *Renstra PD Dinas Pariwisata Provinsi Papua Tahun 2019-2023*. 2019.
- [13] E. Widiyanto, "Pemertahanan Bahasa Daerah melalui Pembelajaran dan Kegiatan di Sekolah," *J. Kreda*, vol. (1) 2, pp. 1–13, 2018.
- [14] W. W. I. Utama, "Revitalisasi Bahasa Daerah Untuk Anak Usia Dini Di Tk Pertiwi Puro Pakualaman Yogyakarta," *J. Skripta*, vol. 6, no. 1, pp. 20–24, 2020, doi: 10.31316/skripta.v6i1.948.
- [15] Y. Cahyono and S. Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic," *J. Inform. Univ. Pamulang*, vol. 4, no. 3, p. 87, 2019, doi: 10.32493/informatika.v4i3.3186.
- [16] F. A. Wenando, R. Hayami, and A. J. Anggrawan, "Analisis Sentimen Pada Pemerintahan Terpilih Pada Pilpres 2019 Ditwitter Menggunakan Algoritme Naïvebayes," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 7, no. 1, pp. 101–106, 2020, doi: 10.33330/jurteks.v7i1.851.
- [17] F. A. Wenando, R. Hayami, Bakaruddin, and A. Y. Novermahakim, "Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms," in *Proceeding - 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, ICITAMEE 2020*, 2020, pp. 279–282, doi: 10.1109/ICITAMEE50454.2020.9398513.
- [18] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and D. Defina, "Analisis Sentimen Bahasa Indonesia pada Twitter Menggunakan Struktur Tree Berbasis Leksikon," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 135, 2021, doi: 10.25126/jtiik.0814133.
- [19] H. B. S. O. Mosioi and E. Mailoa, "Analisa Sentimen Publik Terkait Otonomi
-

- Khusus (OTSUS) di Papua dengan Pendekatan Sains Data,” in *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, 2021, pp. 153–156.
- [20] N. A. Lestari *et al.*, “Metode Naïve Bayes Classifier Dengan Textblob Untuk Analisis Sentimen Terhadap Pelayanan Indihome Dan First Media,” *Semin. Nas. Teknol. Inf. dan Komun. STI&K*, vol. 4, no. 1, 2020, [Online]. Available: <https://t.co/Ws2wOyU5kz>.
- [21] I. G. S. Mas Diyasa, N. M. I. Marini Mandenni, M. I. Fachrurrozi, S. I. Pradika, K. R. Nur Manab, and N. R. Sasmita, “Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1125/1/012034.
- [22] Sh. Fanissa, M. A. Fauzi, and S. Adinugroho, “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Optimasi Sisa Bahan Baku Pada Industri Mebel Menggunakan Algoritma Genetika View project Automatic Essay Scoring View project,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018, [Online]. Available: <https://www.researchgate.net/publication/322959527>.
- [23] L. Lisnawita, G. Guntoro, and M. Musfawati, “Implementation of Naïve Bayes for Classification of Learning Types,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 44–54, 2022, doi: 10.31849/digitalzone.v13i1.9825.
- [24] T. A. Sundara and S. Ekaputri Arnas, “Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme,” pp. 93–98, 2020.
- [25] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [26] D. Tuhenay and E. Mailoa, “Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM),” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 105–111, 2019, doi: 10.33387/jiko.v4i2.2958.
- [27] M. A. Aditya, R. D. Mulyana, I. P. Eka, and S. R. Widiyanto, “Penggabungan Teknologi Untuk Analisa Data Berbasis Data Science,” *Semin. Nas. Teknol. Komput. Sains*, vol. 1, no. 1, pp. 51–56, 2020.
- [28] V. Dhar, “Data Science Prediction,” *Commun. ACM*, vol. 56, no. 12, pp. 64–73, 2013, doi: 10.1145/2500499.
- [29] Y. Ardilla *et al.*, *Data Science*. Bandung: Widina Bhakti Persada Bandung, 2022.
- [30] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.
- [31] I. Romli and B. M. Putra, “Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes,” *J. Pelita Ind.*, vol. 1, no. 1, pp. 1–15, 2021.
- [32] N. F. Hasan and V. Wati, “Deteksi Cyberbullying pada Facebook Menggunakan Algoritma K-Nearest Neighbor,” *J. Smart Syst.*, vol. 1, no. 1, pp. 35–44, 2021, doi: 10.36728/jss.v1i1.1605.
- [33] T. M. Fahrudin, A. R. F. Sari, A. Lisanthoni, and A. A. D. Lestari, “Analisis

<https://doi.org/10.31849/digitalzone.v13i2.11545>

- Speech-To-Text Pada Video Mengandung Kata Kasar Dan Ujaran Kebencian Dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens Dan Visualisasi Word Cloud,” *Skatika*, vol. 5, no. 2, pp. 190–202, 2022, doi: 10.36080/skatika.v5i2.2942.
- [34] D. Berrar, “Bayes’ Theorem and Naive Bayes Classifier,” *Encycl. Bioinforma. Comput. Biol.*, vol. 1, no. 3, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.
- [35] A. R. W and A. R. Pratama, “Analisis Sentimen Kebijakan Pendidikan Di Masa Pandemi COVID-19 dengan CrowdTangle di Facebook,” *Portal J. Univ. Islam Indones.*, vol. 2, no. 2, pp. 1–8, 2020, [Online]. Available: <https://journal.uui.ac.id/AUTOMATA/article/view/19324>.
- [36] M. John, E. Marbach, S. Lohmann, F. Heimerl, and T. Ertl, “MultiCloud: Interactive Word Cloud Visualization for Multiple Texts,” in *Graphics Interface Conference*, 2018, vol. 8-11 May, pp. 25–32, doi: 10.20380/GI2018.04.
- [37] R. Yanuarti, “Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 6, no. 2, pp. 121–130, 2021, doi: 10.32528/justindo.v6i2.5503.
- [38] R. K. Dangin, F. Febriyanto, and R. P. Sari, “Perbandingan Naïve Bayes dan Support Vector Machine pada Sentimen Analisis Reputasi Brand Twitter DQLab.id,” vol. 21, pp. 99–105, 2022.
- [39] M. Ihsan, B. S. Negara, and S. Agustian, “Metode LSTM (Long short term memory) untuk Klasifikasi Sentimen Vaksin Covid-19 pada Twitter,” *Digit. Zo. J. Teknol. Inf. dan Komunikasi Jurnal Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 79–89, 2022, [Online]. Available: <https://doi.org/10.31849/digitalzone.v13i1.9950>.
- [40] E. Indrayuni, “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.1.
- [41] M. Syarifuddin, “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347.
-