

Analisis Penggunaan Kata “Gadget” dan “Gawai” dalam Konteks Era 4.0 hingga 5.0 Menggunakan Metode *Natural Language Processing* (NLP) dan Klasifikasi *Naive Bayes*

Ikmal Thariq Kadafi¹, Larasati Romadhani Yunita Putri², Affa Lelira Ibrahim³, Safira Rahmalia Putri⁴

1,2,3,4 Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran JawaTimur”

e-mail: 22083010052@studentupnjatim.ac.id¹,
22083010053@studentupnjatim.ac.id², 22083010065@studentupnjatim.ac.id³,
22083010073@studentupnjatim.ac.id⁴

Abstrak

Studi ini mengeksplorasi perubahan penggunaan bahasa seiring dengan evolusi teknologi dari era Teknologi 4.0 ke 5.0, dengan fokus pada kata "gadget" dan "gawai". Metode Natural Language Processing (NLP) digunakan untuk menganalisis frekuensi penggunaan kata-kata tersebut dan pola-pola penggunaannya dalam konteks perkembangan teknologi. Analisis ini juga melibatkan metode Naive Bayes untuk mengevaluasi klasifikasi kata "gadget" dan "gawai" oleh model. Hasil visualisasi setelah proses NLP menunjukkan frekuensi penggunaan kata "gadget" dalam jangka tahun 2017-2020 mengalami fluktuasi, berbeda frekuensi penggunaan kata "gawai" yang mengalami peningkatan dalam jangka tahun 2017-2020. Namun, jika kata "gadget" dibandingkan dengan kata "gawai" per tahun, frekuensi kata "gawai" cenderung lebih sedikit dibandingkan kata "gadget" dalam tiap tahunnya. Secara teori linguistik diakronis, fenomena ini menunjukkan bahwa preferensi bahasa masyarakat dapat berubah seiring dengan perkembangan teknologi. Hasil model Naive Bayes memperlihatkan nilai akurasi sebesar 94%, dengan nilai macro avg (rata-rata dari precision, recall, dan F1-score untuk semua kategori) sebesar 0.90. Hal ini menunjukkan bahwa model cenderung lebih baik dalam memprediksi kategori 1 (kata "gadget") dibandingkan dengan kategori 0 (kata "gawai"), karena nilai precision, recall, dan F1-score untuk kategori 1 lebih tinggi dibandingkan dengan kategori 0.

Kata kunci: *Gadget, Gawai, Natural Language Processing, Naive Bayes*

Abstract

This study explores the changes in language usage alongside the evolution of technology from the Technology 4.0 era to the 5.0 era, with a focus on the words "gadget" and "gawai." Natural Language Processing (NLP) method is used to analyze the frequency of usage of these words and their usage patterns in the context of technological development. This analysis also involves the Naive Bayes method to evaluate the classification of the words "gadget" and "gawai" by the model. The visualisation results after the NLP process show that the frequency of use of the word 'gadget' in the period 2017-2020 has fluctuated, in contrast to the frequency of use of the word 'gadget' which has increased in the period 2017-2020. However, if the word 'gadget' is compared to the word 'gadget' per year, the frequency of the word 'gadget' tends to be less than the word 'gadget' in each year. In "diachronic linguistic theory", this phenomenon shows that people's language preferences can change along with technological developments. The results from the Naive Bayes model indicate an accuracy of 94%, with a macro avg (average of precision, recall, and F1-score for all categories) of 0.90. This suggests that the model tends to perform better in predicting category 1 (word "gadget")

compared to category 0 (word "gawai"), as the precision, recall, and F1-score values for category 1 are higher than those for category 0.

Keywords: *Gadget, Gawai, Natural Language Processing, Naive Bayes*

PENDAHULUAN

Pada era teknologi yang terus berkembang pesat, perubahan dalam penggunaan bahasa seiring dengan evolusi teknologi menjadi subjek yang menarik untuk diteliti. Fenomena bahasa mencerminkan dinamika masyarakat dalam mengadaptasi kata-kata baru yang muncul seiring dengan perkembangan teknologi. Dalam konteks ini, kata "gadget" dan "gawai" menjadi perhatian karena merepresentasikan perkembangan teknologi dari era Teknologi 4.0 (2011-2019) menuju 5.0 (2019-sekarang).

Perubahan dalam penggunaan bahasa tidak hanya mencerminkan perkembangan teknologi, tetapi juga menggambarkan bagaimana masyarakat secara kolektif mengadopsi dan mengadaptasi kata-kata baru dalam kehidupan. Fenomena ini semakin menarik karena tidak hanya menyangkut aspek linguistik, tetapi juga mencerminkan transformasi budaya dan sosial yang terjadi seiring dengan perkembangan teknologi.

Teori Linguistik Diakronis, yang mempelajari perubahan bahasa sepanjang waktu, memberikan kerangka teori yang kuat untuk kajian ini. Dalam jurnal "Sinkronis dan Diakronis Linguistik" oleh Nur Fadilah dan Azizi, dijelaskan bahwa teori ini berusaha mengkaji bagaimana perkembangan bahasa yang dituturkan oleh suatu masyarakat dengan membandingkan penggunaannya pada beberapa era atau masa. Teori ini penting karena memungkinkan kita untuk memahami bagaimana kata-kata dan istilah baru muncul, berkembang, dan diadopsi oleh masyarakat. Dengan teori ini, kita dapat memahami bagaimana dan mengapa istilah seperti "gadget" dan "gawai" berkembang dalam periode Teknologi 4.0 dan 5.0.

Untuk menggali lebih dalam tentang fenomena ini, kajian ini akan menggunakan Metode *Natural Language Processing* (NLP). NLP adalah cabang dari kecerdasan buatan yang fokus pada pemahaman dan analisis bahasa manusia secara alami. Dengan bantuan NLP, kita dapat melakukan analisis yang lebih mendalam terhadap frekuensi penggunaan kata "gadget" dan "gawai" dalam konteks evolusi teknologi, serta melihat pola-pola penggunaan kata-kata ini yang dapat memberikan wawasan yang lebih kaya terkait dengan hubungan antara bahasa dan perkembangan teknologi dari Teknologi 4.0 ke 5.0.

Dalam implementasinya, algoritma *Naive Bayes* akan digunakan sebagai metode klasifikasi untuk menganalisis penggunaan kata "gadget" dan kata "gawai" selama evolusi teknologi. Dengan menggunakan algoritma ini, maka dapat diprediksi kategori kata-kata yang berdasarkan data teks, memberikan pemahaman lebih lanjut tentang bagaimana kemajuan teknologi memengaruhi penggunaan Bahasa sehari-hari.

Melalui kajian ini, diharapkan mendapat pemahaman yang lebih mendalam mengenai fenomena bahasa, serta implikasi bahasa dari perubahan era teknologi. Selain itu, hasil kajian ini juga diharapkan dapat memberikan kontribusi praktis dalam penggunaan algoritma NLP dan *Naive Bayes* dalam menganalisis data linguistik dalam konteks yang semakin kompleks dan dinamis.

Penelitian ini bertujuan untuk: menjelaskan bagaimana algoritma pengolahan bahasa alami (NLP) dapat digunakan untuk menganalisis data linguistik terkait dengan evolusi teknologi dari kata "gadget" dan "gawai", mengidentifikasi pola perubahan dalam penggunaan kata "gadget" dan "gawai" selama periode evolusi tersebut, dan meneliti bagaimana penggunaan kata "gadget" dan "gawai" berubah selama evolusi teknologi dari era Teknologi 4.0 ke 5.0.

METODE

Penelitian ini menggunakan pendekatan metode campuran. Metode kuantitatif digunakan pada NLP dan klasifikasi *Naive Bayes* sementara metode kualitatif digunakan

untuk menginterpretasi hasil Analisis Penggunaan Kata “Gadget” dan “Gawai” dengan data didapatkan melalui *scrapping* aplikasi X.

Scrapping Data

Scrapping web adalah proses pengambilan informasi secara otomatis dari halaman web, biasanya halaman web seperti HTML atau XHTML, di mana informasi tersebut digunakan untuk tujuan tertentu. Metode penelitian ini menggunakan Teknik *web scrapping* untuk mengumpulkan data dari aplikasi twitter/X yang relevan dengan topik penelitian. Sebuah Pustaka *python* yang bernama *Twikit*, dimanfaatkan untuk berinteraksi dengan *API* twitter.

Proses Scrapping Data

Dengan menggunakan *Twikit*, beberapa *post/cuitan* yang mengandung kata kunci tertentu yaitu “gawai dan “ponsel” dengan rentang waktu yang telah ditentukan akan dikumpulkan 400 *post/cuitan* dengan jumlah 20 *post/cuitan* per iterasi.

Analisis Data

Data yang sudah dikumpulkan akan disimpan dalam *DataFrame* menggunakan *library* dari *python* yaitu *Pandas* untuk memfasilitasi analisis secara lebih lanjut. Analisis data digunakan untuk mengidentifikasi pola dan tren yang relevan dengan topik penelitian.

Dengan menggunakan teknik *Scrapping Data* ini, maka dapat diakses dan dianalisis dari dari twitter/X secara efisien, dan memungkinkan untuk wawasan lanjutan mengenai topik penelitian.

Natural Language Processing (NLP)

Natural Language Processing (NLP) atau pemrosesan bahasa alami merupakan salah satu cabang dari kecerdasan buatan (AI) yang berhubungan dengan interaksi antara komputer dan manusia menggunakan bahasa alami.

Cara kerja NLP akan melibatkan suatu sistem yang beroperasi secara semi-otomatis. Pada tahap awal penggunaannya, diperlukan NLP *machine learning* untuk memahami data dengan akurat. Setelah itu, mesin akan membuat algoritma yang dapat memahami kata-kata yang ditemukan. NLP dapat dibagi menjadi beberapa tahap, yaitu:

Case Folding

Tujuan dari *case folding* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf dari 'a' sampai 'z' yang akan dipertahankan, sedangkan karakter selain huruf akan dihapus dan dianggap sebagai pemisah kata. Proses ini dilakukan tanpa menggunakan *library* eksternal di *Python*, melainkan dengan memanfaatkan modul yang tersedia dalam bahasa pemrograman tersebut.

Tokenization

Proses ini mengubah teks menjadi unit-unit yang lebih kecil yang disebut dengan token. Token merupakan unit dasar dalam pemrosesan teks, bisa berupa kata, frasa atau karakter terpisah tergantung pada jenis tokenisasi yang digunakan.

Tokenisasi dibedakan menjadi tiga jenis, yaitu:

1. Tokenisasi kata. Metode ini memecah teks menjadi kata-kata individual, yang merupakan metode yang paling umum dan sangat efektif untuk bahasa seperti bahasa Inggris.
2. Tokenisasi karakter. Teks dibagi menjadi karakter individual. Teknik ini berguna untuk bahasa yang tidak memiliki batasan kata yang jelas atau untuk tugas yang memerlukan pemeriksaan menyeluruh, seperti perbaikan ejaan.
3. Tokenisasi subkata. Metode ini memecah teks menjadi unit yang lebih kecil atau lebih besar dari satu kata penuh dengan menyeimbangkan tokenisasi kata dan karakter. Misalnya, kata "Chatbots" dapat diberi token "Chat" dan "bots". Metode ini sangat bermanfaat untuk bahasa yang membentuk makna dengan menggabungkan komponen yang lebih kecil atau ketika berhadapan dengan kata-kata yang tidak termasuk dalam kosakata dalam tugas-tugas pengolahan bahasa natural.

Stop Word Removal

Stop word removal adalah salah satu tahapan dalam pemrosesan teks yang bertujuan untuk menghapus kata-kata yang dianggap tidak memiliki nilai informasi penting

dalam konteks analisis tertentu. Kata-kata ini termasuk kata penghubung, kata seru, kata ganti, dan kata-kata umum lainnya yang sering muncul namun tidak memberikan kontribusi signifikan dalam pemahaman makna teks. Tujuan utama tahap ini adalah untuk mengurangi korpus teks, menghilangkan *noise*, dan meningkatkan kualitas analisis yang dilakukan.

Metode Prediksi *Naive Bayes*

Naive Bayes adalah sebuah pengelompokan statistik yang dapat digunakan untuk memprediksi probabilitas anggota *class*. Jika *Naive Bayes* diaplikasikan pada database dengan *big data*, itu akan mempunyai akurasi dan kecepatan yang sangat kuat. Berikut merupakan rumus dari persamaan *Naive Bayes* :

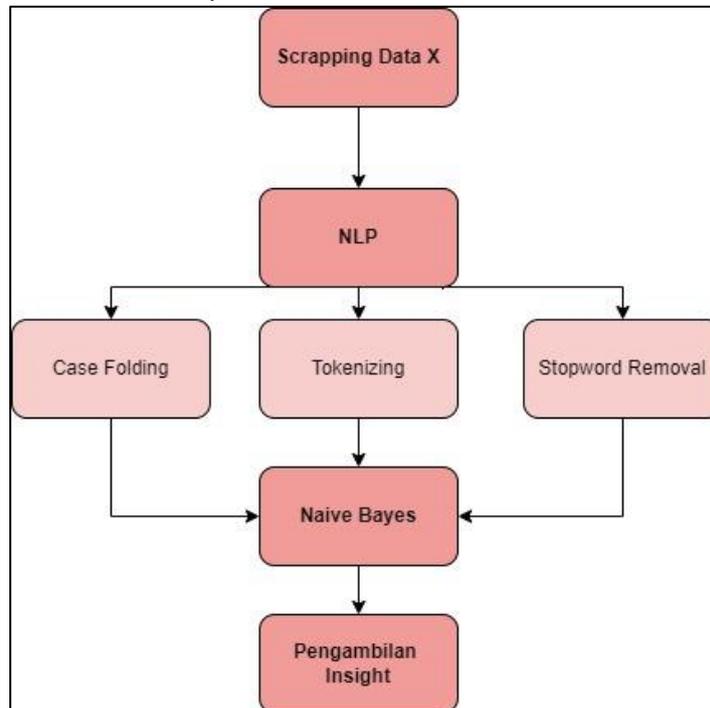
$$P(y_i|X) = \frac{P(X|y_i) \times P(y_i)}{P(X)}$$
, di mana:

- $P(y_i|X)$ adalah probabilitas bahwa sebuah data x masuk dalam kelas y .
- $P(X|y_i)$ adalah probabilitas bahwa data x muncul dalam kelas y .
- $P(y_i)$ adalah probabilitas prior dari kelas y .
- $P(X)$ adalah probabilitas dari data x itu sendiri.

Dengan kata lain, rumus ini digunakan untuk menghitung probabilitas bahwa data x termasuk dalam kelas y berdasarkan seberapa sering kelas y terjadi, dan seberapa sering data x muncul dalam kelas y .

Alur Analisis

Berikut adalah *flowchart* atau tahapan analisis ini:



HASIL DAN PEMBAHASAN

Data

Data yang berhasil di-*scrapping* ada sebanyak 3.864 data. Berikut adalah gambaran data yang akan dianalisis:

	text	datetime
0	Mau 16 gadget ini? NB. PCnya ga termasuk 🤔🤔🤔 M...	2017-04-26 12:57:55
1	Jangan lihat2 toko gadget klo baru bangun tidu...	2017-04-27 00:01:48
2	Gadget ni Ade 1 kebaikan tapi bejuta keburukan...	2017-04-28 15:56:44
3	Sudah follow semua sosial media @SelularID dan...	2017-04-27 06:25:12
4	Beberapa temen yg pamer gadget lawasnya di sin...	2017-04-28 13:52:01

Data di atas diambil dalam rentang tahun 2017-2020 di mana setiap data nya memuat kata “gadget” atau “gawai”. Kolom “datetime” memuat informasi kapan “text” tersebut dikirim.

Penggunaan NLP dalam Menganalisis Kata “Gadget” Dan “Gawai.”
Case Folding

	text	datetime	text_lower
0	Mau 16 gadget ini? NB. PCnya ga termasuk 🤔🤔🤔 M...	2017-04-26 12:57:55	mau 16 gadget ini? nb. pcnya ga termasuk 🤔🤔🤔 m...
1	Jangan lihat2 toko gadget klo baru bangun tidu...	2017-04-27 00:01:48	jangan lihat2 toko gadget klo baru bangun tidu...
2	Gadget ni Ade 1 kebaikan tapi bejuta keburukan...	2017-04-28 15:56:44	gadget ni ade 1 kebaikan tapi bejuta keburukan...
3	Sudah follow semua sosial media @SelularID dan...	2017-04-27 06:25:12	sudah follow semua sosial media @selularid dan...
4	Beberapa temen yg pamer gadget lawasnya di sin...	2017-04-28 13:52:01	beberapa temen yg pamer gadget lawasnya di sin...

Pada tahap ini, semua teks akan diubah ke dalam bentuk huruf kecil (lowercase) untuk konsistensi dan kemudahan pemrosesan. Misalnya, kata "Mau" akan diubah menjadi "mau", dan hal ini dilakukan agar tidak ada perbedaan dalam pemrosesan antara "Mau", "mau", atau "MAU". Setelah diproses, teks akan disimpan dalam tabel yang bernama *text_lower*. Hal ini memastikan bahwa data teks tersimpan dengan format yang seragam dan siap untuk dianalisis.

Tokenization

text_lower	cleaned_text
mau 16 gadget ini? nb. pcnya ga termasuk 🤔🤔🤔 m...	[mau, gadget, ini, nb, pcnya, ga, termasuk, mo...
jangan lihat2 toko gadget klo baru bangun tidu...	[jangan, lihat, toko, gadget, klo, baru, bangu...

Dapat dilihat perbedaan pada tabel *text_lower* dan *cleaned_text* setiap kalimat dipecah menjadi bagian-bagian yang lebih terperinci, seperti frasa-frasa atau kata-kata individual serta dihilangkannya karakter-karakter tertentu. Tujuan dari tahap ini adalah untuk memudahkan pengambilan kata-kata tertentu seperti “gadget” dan “gawai”, karena kata-kata tersebut telah terpisah dari konteks kalimat yang lebih besar. Dengan demikian, kita dapat secara lebih efisien menganalisis frekuensi penggunaan kata-kata tersebut dalam teks yang telah diproses, serta melakukan penelitian yang lebih mendalam terkait dengan kata-kata kunci yang ditargetkan.

Stop Words Removal

cleaned_text	text_WSW	clean_text
[mau, gadget, ini, nb, pcnya, ga, termasuk, mo...	[gadget, nb, pcnya, monggo, tkp, bgt]	gadget nb pcnya monggo tkp bgt
[jangan, lihat, toko, gadget, klo, baru, bangu...	[lihat, toko, gadget, bangun, tidur, kirain]	lihat toko gadget bangun tidur kirain

Dalam perbandingan antara tabel *cleaned_text* dan *clean_text*, terdapat perbedaan yang signifikan terutama dalam penghapusan kata-kata yang kurang memiliki arti atau tidak memberikan kontribusi penting dalam konteks tertentu. Proses penghapusan kata-kata

tersebut bertujuan untuk membersihkan teks dari *noise* atau informasi yang kurang relevan, sehingga meningkatkan kualitas analisis data. Contohnya seperti kata-kata penghubung (contoh: "dan", "atau"), kata-kata umum (contoh: "yang", "ini"), dan beberapa kata lainnya yang sering muncul namun tidak memberikan makna khusus dalam analisis teks.

Dengan menghapus kata-kata yang kurang bermakna ini, teks yang dihasilkan menjadi lebih fokus pada informasi yang relevan dan penting. Hal ini dapat meningkatkan efisiensi dalam proses analisis teks, terutama dalam konteks seperti analisis sentimen, klasifikasi dokumen, atau pencarian informasi yang spesifik.

Naive Bayes Data Labeling

extracted_words	word_count	label
gawai	{'gawai': 1}	0
gadget	{'gadget': 1}	1
gadget	{'gadget': 1}	1
gadget	{'gadget': 1}	1

Langkah selanjutnya adalah membuat kolom baru Bernama "extracted_words" untuk mengekstrak kata "gadget" dan "gawai" dari setiap teks di kolom tersebut. Kata "gadget" dan "gawai" mungkin muncul beberapa kali dalam teks. Sehingga dibentuknya kolom "word_count" yang menghitung berapa kali kata "gadget" dan "gawai" muncul di setiap teks. Selanjutnya gunakan kolom Label untuk memberi label pada kata Gadget dan Gadget yang digunakan dalam proses prediksi menggunakan metode Naive Bayes.

Hasil Klasifikasi

	precision	recall	f1-score	support
[1 1 1 ... 1 1 1]				
0	1.00	0.72	0.84	263
1	0.92	1.00	0.96	871
accuracy			0.94	1134
macro avg	0.96	0.86	0.90	1134
weighted avg	0.94	0.94	0.93	1134

- a. Untuk kategori 0 atau kata "gawai"
 - Didapat nilai *precision* sebesar 1.00, yang berarti dari semua prediksi yang dilakukan model untuk kategori 0, semuanya benar (tidak ada *false positive*).
 - Didapat nilai *recall* sebesar 0.72, yang berarti model dapat mengidentifikasi sekitar 72% dari semua contoh yang sebenarnya termasuk dalam kategori 0.
 - Didapat nilai *F1-score* sebesar 0.84, yang berarti memiliki performa yang baik dari keseluruhan model. Nilai ini mengevaluasi seberapa baik model dapat mengklasifikasi kedua kelas secara seimbang, mempertimbangkan nilai *precision* dan *recall* dari masing-masing kelas.
 - Didapat nilai *support* sebesar 263, yang merupakan jumlah total contoh yang termasuk dalam kategori 0.
- b. Untuk kategori 1 atau kata "gadget"
 - Didapat nilai *precision* sebesar 0.92, yang berarti model mampu memprediksi Sebagian besar contoh kategori 1 secara tepat (tanpa *false positive*).

- Didapat nilai *recall* sebesar 1.00, yang berarti model dapat mengidentifikasi semua contoh yang sebenarnya termasuk dalam kategori 1.
- Didapat nilai F1-score sebesar 0.96, yang berarti memiliki performa yang baik dari keseluruhan model. Nilai ini mengevaluasi seberapa baik model dapat mengklasifikasi kedua kelas secara seimbang, mempertimbangkan nilai *precision* dan *recall* dari masing-masing kelas.
- Didapat nilai *support* sebesar 871, yang merupakan jumlah total contoh yang termasuk dalam kategori 1.

Secara keseluruhan, model *Naive Bayes* memiliki akurasi sebesar 94%, dengan nilai *macro avg* (rata-rata dari *precision*, *recall*, dan F1-score untuk semua kategori) sebesar 0.90. Hal ini menunjukkan bahwa model cenderung lebih baik dalam memprediksi kategori 1 (kata "gadget") dibandingkan dengan kategori 0 (kata "gawai"), karena nilai *precision*, *recall*, dan F1-score untuk kategori 1 lebih tinggi dibandingkan dengan kategori 0.

Perbandingan Data *Test* dan Data *Train*

	index	aktual	prediksi
0	1327	0	1
1	2692	1	1
2	2424	1	1
3	310	1	1
4	1158	1	1
...
1129	3034	1	1
1130	3610	1	1
1131	2732	1	1
1132	1784	1	1
1133	3804	1	1

1134 rows x 3 columns

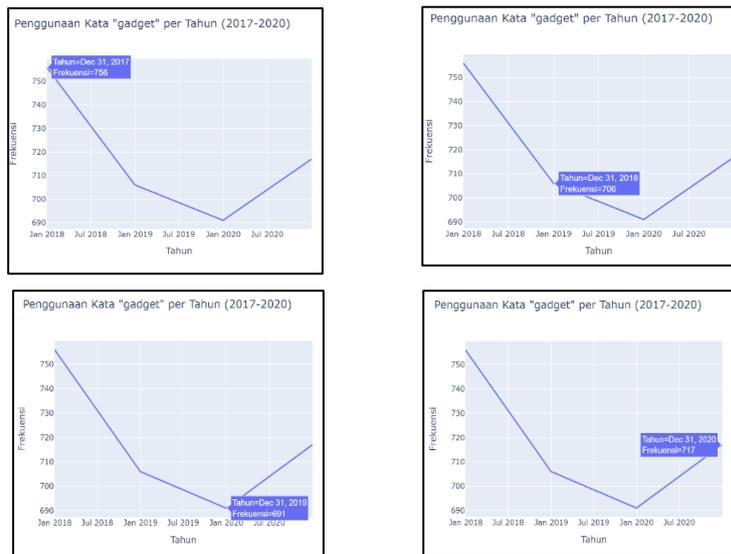
Gambar tabel tersebut merupakan tabel hasil dari membandingkan label aktual (data *test*) dengan label yang diprediksi oleh model (data *train*). Dalam setiap baris, nilai pada kolom 'aktual' menunjukkan label sebenarnya dari data tersebut, sedangkan nilai pada kolom 'prediksi' menunjukkan label yang diprediksi oleh model untuk data tersebut. Pada baris pertama, nilai 'aktual' adalah 0 (label 'gawai') dan nilai 'prediksi' adalah 1 (label 'gadget'). Ini menunjukkan bahwa model salah memprediksi data uji tersebut. Namun, pada baris kedua hingga terakhir, model berhasil memprediksi dengan benar, karena nilai 'aktual' dan 'prediksi' adalah sama (0 atau 1).

Frekuensi Penggunaan Kata "Gadget" dan "Gawai" Seiring Berkembangnya Evolusi Teknologi.

Dalam penelitian ini, ditemukan hasil eksplorasi mengenai penggunaan kata "gadget" dan "gawai" yang berubah dari era Teknologi 4.0 ke era Teknologi 5.0. Melalui analisis hasil *scrapping* twitter/X dan visualisasi grafik, dapat diperoleh pemahaman yang lebih mendalam mengenai bagaimana kata-kata ini digunakan dalam konteks masyarakat, maupun perubahan preferensi karena kemajuan teknologi.

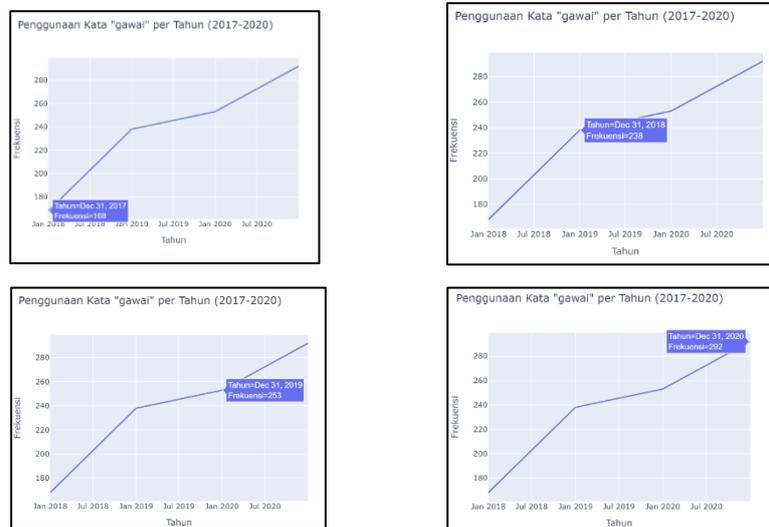
Berikut merupakan visualisasi hasil *scrapping* yang menunjukkan perbandingan frekuensi penggunaan kata "gadget" dan "gawai" dari era Teknologi 4.0 hingga 5.0.

Frekuensi Penggunaan Kata “Gadget” dari Tahun 2017-2020.



Dapat dilihat dari gambar-gambar di atas bahwa frekuensi kata “gadget” pada tanggal 31 Desember 2017 adalah sebanyak 756 *tweet*, kemudian mengalami penurunan pada tanggal 31 Desember 2018 menjadi 706 *tweet*, 691 *tweet* pada tanggal 31 Desember 2019, dan Kembali meningkat menjadi 717 *tweet* pada tanggal 31 Desember 2020. Hal ini menunjukkan adanya fluktuasi dalam penggunaan kata “gadget” selama periode tahun 2017 hingga tahun 2020.

Frekuensi Penggunaan Kata “Gawai” dari Tahun 2017-2020.



Berbeda dengan kata “gadget” yang sebelumnya, gambar-gambar di atas menunjukkan penggunaan kata “gawai” dengan tren yang cenderung meningkat dari tahun ke tahun. Pada tanggal 31 Desember 2017, frekuensi penggunaan kata “gawai” adalah 168 *tweet*, kemudian meningkat menjadi 238 *tweet* pada tanggal 31 Desember 2018, dan terus meningkat menjadi 253 *tweet* pada tanggal 31 Desember 2019.

Perbandingan Frekuensi Kata “Gawai” dan “Gadget” dalam Rentang tahun 2017-2020

Pada tahun 2017, frekuensi kata "gawai" sebesar 168, yang jelas lebih rendah jika dibandingkan dengan frekuensi kata "gadget" pada tahun yang sama yaitu sebesar 756. Di

sisi lain, pada frekuensi terendah kata "gadget" yaitu di tahun 2019 sebesar 619, perbandingannya dengan frekuensi kata "gawai" di tahun tersebut tetap lebih rendah yaitu sebesar 253, padahal pada tahun ini frekuensi kata "gawai" mencapai titik tertinggi. Hal ini menunjukkan bahwa perbedaan frekuensi penggunaan kata "gawai" tetap relatif kecil dibanding kata "gadget" meskipun mengalami peningkatan dari tahun ke tahun.

Pembahasan

Dari hasil penelitian yang sudah didapatkan sebelumnya, didapatkan bahwa algoritma *Natural Language Processing (NLP)* memainkan peran penting dalam menganalisis data linguistik terkait kata "gadget" dan "gawai" dalam konteks evolusi teknologi. Tahapan *NLP* seperti *case folding*, *tokenization*, dan *stop words removal* memungkinkan pemrosesan data teks yang lebih efisien. *Case folding* memastikan konsistensi huruf, *tokenization* memecah teks menjadi kata-kata individu untuk analisis yang lebih mudah, *stop words removal* membersihkan kata-kata yang tidak relevan. Dengan *NLP*, analisis teks akan menjadi lebih efisien dan memberikan wawasan yang lebih mendalam tentang penggunaan kata tersebut seiring dengan kemajuan teknologi.

Penggunaan algoritma *Naïve Bayes* dalam mengklasifikasi kata "gadget" dan "gawai" menunjukkan hasil yang memuaskan. Algoritma ini berhasil mencapai akurasi sebesar 94%, menunjukkan kemampuan untuk mengidentifikasi dan memprediksi kategori kata dengan baik. Meskipun ada sedikit ketidakcocokan dalam prediksi kata "gawai", kinerja model secara keseluruhan masih cukup baik, dengan nilai rata-rata dari *precision*, *recall*, dan *F1-score* sebesar 0.90. Penting untuk dicatat bahwa model cenderung lebih baik dalam memprediksi kategori "gadget" daripada "gawai", karena nilai *precision*, *recall*, *F1-score* kata "gadget" lebih tinggi dari kata "gawai". Ini menunjukkan bahwa algoritma *Naïve Bayes* efektif dalam analisis data linguistik yang berkaitan dengan evolusi teknologi serta memberikan pemahaman yang lebih baik mengenai penggunaan kata "gadget" dan "gawai" dalam konteks evolusi teknologi.

Berdasarkan hasil visualisasi penggunaan kata "gadget" dan "gawai" dari tahun 2017 hingga 2020, terdapat perbedaan tren yang menarik. Meskipun kata "gadget" memiliki frekuensi penggunaan yang jauh lebih tinggi dibandingkan dengan "gawai", kata "gawai" cenderung meningkat secara konsisten sepanjang periode, sementara "gadget" mengalami fluktuasi yang lebih signifikan.

Secara teori linguistik diakronis, fenomena ini menunjukkan bahwa preferensi bahasa masyarakat dapat berubah seiring dengan perkembangan teknologi. Kata "gawai" yang cenderung meningkat penggunaannya mengindikasikan terjadinya pergeseran leksikal, di mana masyarakat mulai lebih memilih istilah "gawai" yang dianggap lebih sesuai dengan konteks teknologi saat ini. Sementara itu, fluktuasi penggunaan kata "gadget" mencerminkan adanya variasi dan dinamika dalam pemilihan kosakata di masyarakat. Perubahan ini dapat dipandang sebagai salah satu bentuk adaptasi bahasa terhadap perkembangan zaman.

SIMPULAN

Penggunaan kata "gadget" dan "gawai" menunjukkan tren yang berbeda dari tahun 2017 hingga 2020, dengan "gawai" mengalami peningkatan konsisten dan "gadget" menunjukkan fluktuasi yang signifikan. Algoritma *Natural Language Processing (NLP)* memainkan peran penting dalam menganalisis data linguistik terkait kedua kata tersebut. Teknik-teknik *NLP* seperti *case folding*, *tokenization*, dan *stop words removal* meningkatkan efisiensi analisis teks. Algoritma *Naïve Bayes* digunakan untuk mengklasifikasi kata "gadget" dan "gawai", mencapai akurasi 94%, dengan performa lebih baik pada kata "gadget" dibanding "gawai". Ini menunjukkan efektivitas *Naïve Bayes* dalam analisis data linguistik terkait evolusi teknologi dan memberikan wawasan yang lebih mendalam tentang penggunaan kedua kata tersebut.

DAFTAR PUSTAKA

- Awan, A. A. (2023, September 22). *What is tokenization? types, use cases, implementation*. DataCamp. Bab 2 landasan teori - elibrary unikom. (n.d.).
- Derry. (2013). *Bila Si Kecil Bermain Gadget: Panduan Bagi Orang Tua Untuk Memahami Faktor-Faktor Penyebab Anak Kecanduan Gadget*. Jakarta: Bisakimia.
- Dr. Suherman, S.Kom., M.M., Dr. Musnaini, S.E., M.M. , Hadion Wijoyo, S.E.,S.H.,S.Sos.,S.Pd.,M.H.,M.M.,Ak.,CA.,QWP®, Dr (C). Irjus Indrawan,S.Pd.I.,M.Pd.I. (2020). *INDUSTRY 4.0 vs SOCIETY 5.0*. Jawa Tengah:Pena Persada.
- Fadilah, N., & Azisi. (n.d.-a). Sinkronis Dan Diakronis Linguistik. <https://ejournal.stainh.ac.id/index.php/kafaah/article/download/76/25/325>
- Faruqi, Umar Al. 2019. "Future Service in Industry 5.0." *Jurnal Sistem Cerdas* 2 (1): 67–79.
- Hasibuan, A. (2015, July 2). PERBEDAAN TEORI LINGUISTIK. <https://ejournal.bbg.ac.id/metamorfosa/article/download/127/100/>
- Kamus Besar Bahasa Indonesia (KBBI). (2023). Badan Bahasa Kemendikbudristek.
- Nordquist, R. (2019, July 7). *What is diachronic linguistics in grammar?*. ThoughtCo. <https://www.thoughtco.com/diachronic-linguistics-term-1690385>
- Nugroho, K. S. (2020, June 4). *Dasar text preprocessing Dengan Python*. Medium.
- Oktriwina, A. S. (2021, February 2). *NLP: Kecerdasan Buatan Yang bantu Komputer Pahami Bahasa manusia*. Glints Blog.
- Schwab, K. (2017). *The Fourth Industrial Revolution*. Portfolio Penguin.
- Sunaryo. (2000). *Metode Riset Sosial, Suatu Pengantar*. Malang: IKIP Malang
- Tetesanpena. (n.d.). Pengertian Linguistik Sinkronik, Sinkronis Dan Diakronik. Belajar Bahasa. <https://funcass.blogspot.com/2018/02/pengertian-linguistik-sinkronik.html?m=1>
- Verihubs. (2023, October 23). *NLP (Natural Language Processing): Contoh & Cara Kerjanya*.