

Analisis Sentimen Produk Amazon Menggunakan Algoritma Naïve Bayes Pada Data Review Pelanggan

Jesica Yolanda Br.Sibarani¹, Agus Junaidi^{2*}, Mochammad Abdul Azis³, Yoseph Tajul Arifin⁴, Diah Puspitasari⁵

^{1,2,3,4,5} Universitas Bina Sarana Informatika, Indonesia

¹e-mail: jesikasibarani123@gmail.com

^{2*}e-mail: agus.asj@bsi.ac.id

³e-mail: mochamad.mmz@bsi.ac.id

⁴e-mail: yoseph.ypa@bsi.ac.id

⁵e-mail: diah.puspitasari@bsi.ac.id

Diterima	Direvisi	Disetujui
02-11-2023	20-11-2023	05-12-2023

Abstrak - Banyaknya data yang tersebar di internet dapat diolah untuk berbagai kepentingan baik untuk bisnis, sekolah, politik dan lainnya. Dan peningkatan teknologi juga berdampak ke pasar produksi. Dari banyaknya e-commerce yang muncul dari berbagai negara dengan penggunaan yang relative mudah dan *efficient*. Yang membuat pelanggan tidak harus datang langsung ke toko dari produk yang ingin dibeli. Produk-produk yang dijual di e-commerce juga sangat variatif dan lengkap. Pelanggan juga dapat membeli barang yang dijual di toko luar negeri dengan sekali klik maka barang akan sampai kerumah. Salah satu e-commerce terbesar dan dapat dijangkau dari tiap-tiap negara yang berbeda adalah Amazon. Amazon menyediakan lebih dari miliaran produk dan di akses banyak pengguna setiap harinya. Maka dengan banyaknya persaingan toko dan produk di Amazon maka pembeli harus dapat menilai produk tidak hanya dari deskripsi dan foto yang disediakan toko tetapi juga harus melihat ulasan yang diberikan pembeli yang telah lebih dulu membeli barang tersebut dan seberapa banyak barang itu telah dibeli. Data review amazon yang digunakan ada 60.888 dan setelah dibersihkan dan dibuang kalimat duplikat maka data bersihnya ada 43.126 data. Dengan data sentiment positif sebesar 32318 data, negative 7.478 data dan neutral 3.330 data dengan tingkat akurasi menggunakan metode naïve bayes yaitu sebesar 100%.

Kata Kunci: Analisis sentimen, Produk amazon, Algoritma Naïve Bayes

Abstract - The abundance of data available on the internet can be processed for various purposes, including business, education, politics, and more. Technological advancements have also impacted the production market. With the emergence of numerous e-commerce platforms from different countries, the ease of use and efficiency have increased. This eliminates the need for customers to physically visit stores to purchase desired products. The products available on e-commerce platforms are diverse and comprehensive. Customers can even buy items sold by overseas stores with just a click, and the goods will be delivered to their doorstep. One of the largest and globally accessible e-commerce platforms is Amazon. Amazon offers billions of products and is accessed by a large number of users every day. Due to the intense competition among stores and products on Amazon, buyers need to evaluate products not only based on the store's description and photos but also by considering reviews from previous customers who have purchased the same item, as well as the quantity of items sold. The dataset used for this analysis consists of 60,888 Amazon reviews. After cleaning the data and removing duplicate sentences, the clean dataset consists of 43,126 records. The dataset is further categorized into positive sentiment (32,318 records), negative sentiment (7,478 records), and neutral sentiment (3,330 records) and the accuracy rate achieved using the Naïve Bayes method is 100%.

Keywords: Sentiment analysis, Amazon product, Naive Bayes Algorithm

PENDAHULUAN

Salah satu E-commerce yang paling besar didunia adalah Amazon.com. Amazon.com Inc merupakan perusahaan multinasional teknologi Perusahaan Amerika yang berfokus pada e-niaga, komputasi awan, streaming digital dan kecerdasan

buatan. Ini adalah salah satu dari Lima Besar perusahaan di industri teknologi informasi Amerika Serikat, bersama dengan Google (Alphabet), Apple, Meta (Facebook) dan Microsoft (Act Exchange, 2020).

Perusahaan ini memiliki kantor yang berpusat di Seattle, Washington, Amerika Serikat ini pada



awalnya hanya menjual buku namun sekarang telah bervariasi dengan berbagai macam kategori penjualan baik elektronik, non elektronik, sandang dan pangan, aksesoris dan banyak lagi. Selain itu Amazon juga merancang dan membuat perangkat elektronik berupa alat untuk membaca buku elektronik Amazon Kindle dan Komputer tablet Kindle fire, selain itu Amazon juga menyediakan jasa layanan komputasi dan cloud. Hasil yang dihasilkan oleh mesin pencari Amazon sebagian besar ditentukan oleh biaya promosi. Lokalisasi toko lokal Amazon, yang berbeda dalam pemilihan dan harga, dibedakan berdasarkan domain tingkat atas dan kode dari tiap negara (Clinton, 2021).

Data Metode analisis sentiment dengan menggunakan algoritma Naive Bayes dipilih karena metode ini telah terbukti efektif dalam melakukan klasifikasi teks berdasarkan sentimennya. Algoritma ini juga relatif mudah diimplementasikan dan memerlukan sedikit waktu untuk melatih modelnya. Algoritma *Naive Bayes* merupakan salah satu algoritma klasifikasi yang banyak digunakan pada Data Mining ataupun Text Mining. Algoritma *Naive Bayes* berdasarkan teorema *Bayes* bahwa semua kegiatan memberikan sebuah kontribusi yang sama penting atau saling bebas pada pemilihan kelas tertentu. Salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam Text Mining adalah metode *Naive Bayes* yang sering disebut dengan *Naive Bayes Classifier* (Darwis et al., 2021). Metode *Naive Bayes Classifier* adalah metode yang dapat digunakan dalam pengambilan keputusan untuk mendapatkan hasil yang lebih baik pada suatu masalah klasifikasi (Alita et al., 2021). Algoritma *Naive Bayes Classifier* digunakan untuk mendapatkan hasil akurasi terbaik sehingga dapat digunakan untuk dataset Gagal Jantung (Subarkah et al., 2022).

Naive Bayes Classifier adalah sekumpulan algoritma yang didasarkan pada Teorema Bayes. Dengan kata lain, algoritma ini bukan algoritma tunggal melainkan satu grup algoritma dimana masing-masing memiliki prinsip kerja yang mirip. Algoritma ini bekerja berdasarkan prinsip probabilitas bersyarat, seperti yang diberikan oleh Teorema Bayes (Berrar, 2018). Teorema Bayes menemukan probabilitas atau kemungkinan suatu peristiwa akan terjadi dengan memberikan probabilitas peristiwa lain yang telah terjadi. Dalam istilah yang lebih sederhana, Teorema Bayes adalah metode untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya (Trivusi, 2022). Pengklasifikasi *Naive Bayes* (NB) adalah keluarga pengklasifikasi probabilistik sederhana berdasarkan asumsi umum bahwa semua fitur tidak bergantung satu sama lain (Xu, 2018) Pengklasifikasi *Naive Bayes* adalah salah satu algoritma yang banyak digunakan untuk aplikasi penambangan data (Foo et al., 2022). Penelitian ini menggunakan metode *Naive Bayes Classifier* (NBC) untuk menganalisis tingkat sentimen terhadap data yang dilakukan dengan crawling di Twitter (Normawati & Prayogi, 2021)

Mendapatkan kepuasan dan kepercayaan pelanggan telah menjadi tantangan utama dalam mencapai kesuksesan di dunia bisnis. Pebisnis perlu mengidentifikasi masalah yang muncul dari ulasan yang diberikan oleh pelanggan (Sihombing et al., 2021). Dalam sebuah e-commerce Shopee, proses jual beli terus berjalan setiap harinya, dan komentar yang diberikan konsumen akan meningkat. Dengan menggunakan teknik Text Mining metode klasifikasi, akan diketahui suatu sentimen bernilai positif, netral atau negatif (Sari & Wibowo, 2019).

Komentar yang diberikan oleh konsumen akan menjadi acuan/review dari suatu produk yang telah dibeli oleh konsumen (Hariguna et al., 2019). Aktivitas online yang paling populer di Indonesia adalah penggunaan media sosial (khotimah & Utami, 2022). Dengan meluasnya penggunaan jejaring sosial, forum, dan blog, ulasan pelanggan muncul sebagai faktor penting untuk keputusan pembelian pelanggan (Catal & Nangir, 2017). Penelitian ini menilai apakah istilah yang terkait dengan pengalaman tamu dapat digunakan untuk mengidentifikasi cara untuk meningkatkan layanan perhotelan. Sebuah studi dilakukan untuk mengidentifikasi fitur yang relevan secara empiris untuk mengklasifikasikan kepuasan pelanggan (Sánchez-Franco et al., 2019). Hasilnya menunjukkan bahwa pengklasifikasi Bayes naif untuk data fungsional berlaku untuk masalah klasifikasi multi-kategori dan memiliki kinerja sampel terbatas yang lebih disukai daripada pesaing (Zhang & Sakhanenko, 2019).

METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen dengan menggunakan bantuan perangkat lunak Rapidminer yang sering digunakan dalam membuat analisis *data mining* dengan tahapan sebagai berikut:

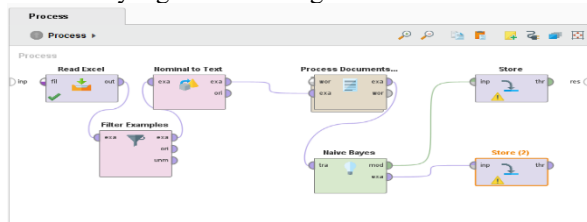
1. Pengumpulan Data: Dalam penelitian ini pengumpulan data dilakukan dengan mengambil data review produk Amazon dari tahun 2013-2019. Data yang dimiliki meliputi unique ID, category, review header, review text, rating, own_rating. Data diperoleh melalui data sekunder pada laman www.kaggle.com/datasets.
2. Pra-pemrosesan Data: Data yang telah dimiliki kemudian akan diproses dan menghapus data yang tidak relevan seperti tanda baca, angka, dan simbol. Selain itu, data juga akan dilakukan penghapusan kata-kata stopword dan dilakukan normalisasi kata-kata. Cara kerja atau metode yang digunakan stopword removal dengan mencari kata yang sama pada data yang akan dianalisis dan menghapus kata tersebut termasuk tanda baca dan karakter khusus.
3. Metode *Machine Learning*, yaitu merupakan metode *Naive Bayes Classifier* (NBC) digunakan untuk melakukan klasifikasi ulasan yang berbentuk positif, netral dan negative.
4. Analisis Hasil dan penarikan kesimpulan: Hasil dari evaluasi model akan dianalisis untuk menentukan

tingkat akurasi dan performa model yang telah dibuat. Hasil analisis ini dapat digunakan untuk mengevaluasi keefektifan algoritma *Naive Bayes* dalam melakukan sentimen analisis terhadap review negative dan positif produk di Amazon. Kemudian menarik kesimpulan dari hasil penelitian yang telah dilakukan.

HASIL DAN PEMBAHASAN

Proses Data Latih Sentimen

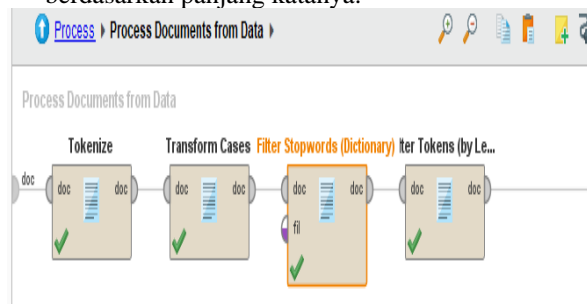
Melakukan *process* pada data ulasan (*review*) yang sudah memiliki sentiment yang data tersebut digunakan sebagai data uji untuk menentukan sentiment yang masih kosong.



Gambar 1. Operator untuk process awal

Process Documents from Data: didalam process ini memiliki *sub-process* yang digunakan untuk menghilangkan kata-kata atau kalimat yang tidak sesuai dengan '*Natural Language Processing*' (NLP) dan mendapatkan hasil kalimat-kalimat dengan kata kunci yang lebih penting dan terstruktur:

1. *Tokenize*, untuk merubah kalimat menjadi perkata.
2. *Transform Cases*, untuk merubah huruf besar menjadi huruf kecil semua.
3. *Filter Stopwords (Dictionary)*, menghilangkan kata-kata yang tidak penting berdasarkan file *stopword* yang digunakan.
4. *Filter Tokens (by Length)*, menentukan kalimat berdasarkan panjang katanya.



Gambar 2. Sub-Process Documents from Data

3.1 Comparative Judgement

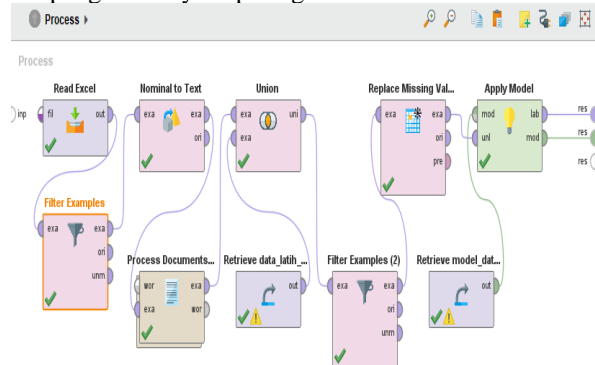
Process Sentiment adalah *process* yang digunakan untuk menentukan nilai sentiment yang berupa positif, negative dan netral dari sebuah data ulasan produk yang ditinggalkan pengguna produk tersebut.

Penjelasan operator yang digunakan pada Gambar 3 adalah:

1. *Read Excel*: file yang dimiliki dimasukkan dengan menghilangkan beberapa kolom yang tidak

diperlukan, setelahnya kolom *Own_Rating* di ubah menjadi label.

2. *Filter Examples*: digunakan untuk memisahkan *own_rating* yang belum berisi sentiment (positif, negative dan netral) dari yang sudah memiliki sentiment, data yang diambil adalah data yang tidak memiliki sentiment di *Own_Rating* (*is missing*). Dipengaturan *add filters* kita harus merubah pengaturannya seperti gambar 3.



Gambar 3. Operator untuk process lanjutan

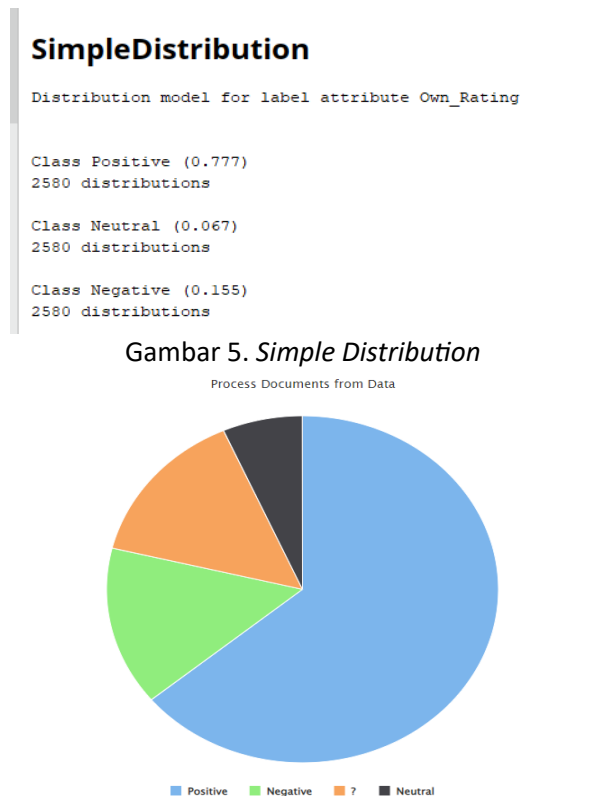
3. *Nominal to Text*: digunakan untuk mengubah nilai nominal didalam didalam file ke bentuk text agar dapat dibaca saat melakukan process dokuments from data. Data tidak akan bisa diprocess bila nilai nominal dalam data tidak diubah ke bentuk text.
4. *Process Documents from Data*: pengaturan yang digunakan masih sama dengan sebelumnya dengan parameters menggunakan vector TF-IDF dan juga sub-process.
5. *Retrieve_data_latih_amazon*: adalah data yang sebelumnya sudah disimpan diprocess awal.
6. *Union*: digunakan untuk menggabungkan data latih amazon yang sebelumnya sudah tersimpan dengan data baru yang sedang diproses.
7. *Filter Examples*: akan digunakan kembali untuk memproses data baru.
8. *Replase missing value*: digunakan untuk mengubah tanda tanya "?" di data yang kosong menjadi angka zero "0".
9. *Retrieve_model_data_amazon*: adalah data yang sebelumnya sudah disimpan diprocess awal.
10. *Apply model*: digunakan untuk mengisi model sentiment yang kosong dengan menggunakan sampel model yang sebelumnya sudah di isi.

Own_Ra...	predictedD...	confid...	confid...	confid...	text	Rating	assamazing	awful	awful	avg
1	Negative	0	0	1	gone within month	1	0	0	0	0
1	Positive	1	0	0	price quality built quality this price amazing mo...	5	0	0	0	0
1	Positive	1	0	0	there special old memory card after have so...	5	0	0	0	0
1	Positive	1	0	0	super	5	0	0	0	0
1	Positive	1	0	0	good phone this price range segment front ca...	4	0	0	0	0
1	Positive	1	0	0	complaints whatsoever they then make sure h...	5	0	0	0	0
1	Negative	0	0	1	very good item	5	0	0	0	0
1	Positive	1	0	0	this earphone awesome must	5	0	0	0	0
1	Positive	1	0	0	this best earphones used perfect combination	5	0	0	0	0
1	Negative	0	0	1	worth value money never down	5	0	0	0	0
1	Positive	1	0	0	person habitual customized android face like	5	0	0	0	0
1	Positive	1	0	0	awesome product picture quality brilliant	5	0	0	0	0
1	Negative	0	0	1	display problem occurs within months	1	0	0	0	0
1	Positive	1	0	0	product good only thing given days useful dur...	4	0	0	0	0

Gambar 4. Hasil Prediction sentiment

Gambar 4 adalah hasil yang didapat dari process analisis sentiment dengan algoritma *naïve bayes*

menggunakan Rapidminer. Pada gambar 4 dapat dilihat Own_Rating (sentiment) yang sebelumnya masih tanda tanya (?) atau kosong sekarang sudah berisi sentiment prediction dari hasil pemrocessing menggunakan metode *naive bayes*.



Gambar 6. Original data Own_rating

Gambar diatas adalah *visualizations* dari data asli yang dimiliki. Dapat dilihat bahwa responden lebih banyak yang memberi sentiment positive. Dari *visualizations* diatas data yang tidak memiliki sentiment adalah berwarna orange. Dan data sentiment yang kosong tidak lebih banyak dari data yang sudah memiliki sentiment. Hal tersebut akan membuat tingkat akurasi prediksi lebih besar dan akurat.

Attribute	Parameter	Positive	Neutral	Negative
Rating	mean	4.717	3	1.259
Rating	standard deviation	0.451	0.001	0.440
sat	mean	0.000	0	0
sat	standard deviation	0.008	0.001	0.001
saifa	mean	0.000	0	0
saifa	standard deviation	0.013	0.001	0.001
saiga	mean	0	0	0.004
saiga	standard deviation	0.001	0.001	0.048
saibly	mean	0.000	0	0
saibly	standard deviation	0.007	0.001	0.001
saib	mean	0.000	0	0
saib	standard deviation	0.011	0.001	0.001
saie	mean	0.001	0.000	0.001
saie	standard deviation	0.018	0.029	0.009
saibut	mean	0.004	0.002	0.005

Gambar 7. Distribution table

Data yang dimiliki di proses dan dibersihkan hingga mendapatkan hasil seperti gambar-gambar diatas. Tahap awal proses yang dilakukan adalah pembersihan dan penyederhanaan kata untuk

mendapatkan kata-kata yang jauh lebih penting. Hasil masing-masing langkah dari tahap preprocessing dapat dilihat dalam Tabel.1 sampai Tabel.4.

Tabel 1. *Cleansing* dalam tahapan *preprocessing* data

Sebelum <i>Cleansing</i>	Sesudah <i>Cleansing</i>
<i>Back camer is good in full light.Pubg is very smooth battery life is also very good in 9000 you can go with this phone it has every feture</i>	<i>Back camer is good in full light Pubg is very smooth battery life is also very good in 9000 you can go with this phone it has every feture</i>
<i>Awesome book at reasonable price, must buy</i>	<i>Awesome book at reasonable price must buy</i>
.....
<i>The book is fine,not bad,contains nice concepts and nicely explained. Likes its concepts.</i>	<i>The book is fine,not bad contains nice concepts and nicely explained Likes its concepts</i>
<i>Nice tv and pic quality .good custmer srrvice my tv install 3 hours Isme home theater kaise lagega</i>	<i>Nice tv and pic quality good custmer srrvice my tv install 3 hours Isme home theater kaise lagega</i>
<i>Iâ€™m a fan of Alexa Fire Stic. Its a value for money thing and your TV has whole lot of good options to spend time on. Definitely recommended. Easy user interface and voice control. Thereâ€™s a small lag in voice command, that can be fixed.</i>	<i>a fan of Alexa Fire Stic Its a value for money thing and your TV has whole lot of good options to spend time on. Definitely recommended Easy user interface and voice control a small lag in voice command that can be fixed</i>

Tabel 2. *Case Folding* dalam tahapan *preprocessing* data

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
<i>Back camer is good in full light Pubg is very smooth battery life is also very good in 9000 you can go with this phone it has every feture</i>	<i>back camer is good in full light pubg is very smooth battery life is also very good in 9000 you can go with this phone it has every feture</i>
<i>Awesome book at reasonable price, must buy</i>	<i>awesome book at reasonable price must buy</i>
<i>The book is fine,not bad contains nice concepts and nicely explained Likes its concepts</i>	<i>the book is fine,not bad contains nice concepts and nicely explained likes its concepts</i>
<i>Nice tv and pic quality good custmer srrvice my tv install 3 hours Isme home theater kaise lagega</i>	<i>nice tv and pic quality good custmer srrvice my tv install 3 hours isme home theater kaise lagega</i>
<i>a fan of Alexa Fire Stic Its a value for money thing and your TV has whole lot of good options to spend time on. Definitely recommended. Easy user interface and voice control. Thereâ€™s a small lag in voice command, that can be fixed</i>	<i>a fan of alexa fire stic its a value for money thing and your tv has whole lot of good options to spend time on. definitely recommended easy user interface and voice control a small lag in voice command that can be fixed</i>

Tabel 3. *Stemming* dan *filtering* dalam tahapan *preprocessing* data

Sebelum <i>Stemming</i> dan <i>Filtering</i>	Sesudah <i>Stemming</i> dan <i>Filtering</i>
back camer is good in full light pubg is very smooth battery life is also very good in 9000 you can go with this phone it has every feture awesome book at reasonable price must buy the book is fine,not bad contains nice concepts and nicely explained likes its concepts nice tv and pic quality good custmer srrvice my tv install 3 hours isme home theater kaise lagega a fan of alexa fire stic its a value for money thing and your tv has whole lot of good options to spend time on definitely recommended easy user interface and voice control thereâ voice command that can be fixed	camer good full light pubg smooth battery life good phone feture awesome book reasonable price book fine contains nice concepts nicely explained likes concepts nice quality good custmer srrvice install hours isme theater kaise lagega alexa fire stick value money thing good options spend time definitely recommended easy user interface voice control thereâ voice command fixed

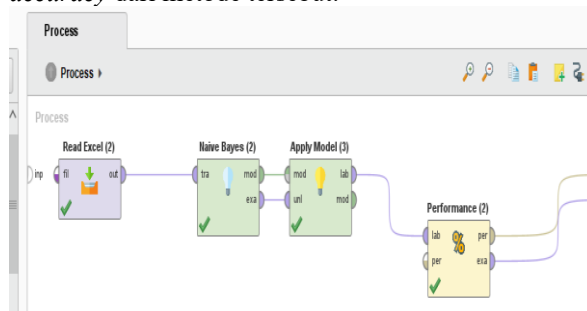
Tabel 4. *Tokenize* dalam tahapan *preprocessing data*

Sebelum <i>Tokenize</i>	Sesudah <i>Tokenize</i>
camer good full light pubg smooth battery life good phone feture	['camer', 'good', 'full', 'light', 'pubg', 'smooth', 'battery', 'life', 'good', 'phone', 'feture']
awesome book reasonable price	['awesome', 'book', 'reasonable', 'price']
book fine contains nice concepts nicely explained likes concepts	['book', 'fine', 'contains', 'nice', 'concepts', 'nicely', 'explained', 'likes', 'concepts']
nice quality good custmer srrvice install hours isme theater kaise lagega	['nice', 'quality', 'good', 'custmer', 'srrvice', 'install', 'hours', 'isme', 'theater', 'kaise', 'lagega']
alexa fire stick value money thing good options spend time definitely recommended easy user interface voice control thereâ voice command fixed	['alexa', 'fire', 'stick', 'value', 'money', 'thing', 'good', 'options', 'spend', 'time', 'definitely', 'recommended', 'easy', 'user', 'interface', 'voice', 'control', 'thereâ', 'voice', 'command', 'fixed']

3.2 Performance

Performance adalah operator yang melihat dan menilai seberapa baik dan besar tingkat *accuracy* dari data yang sedang diuji. Data yang sudah memiliki

sentiment kemudian disimpan dan di uji kembali dengan algoritma *naive bayes* untuk melihat tingkat *accuracy* dari metode tersebut:



Gambar 7. *Performance Naive Bayes*

Data yang sudah di *process* dan sudah mendapatkan nilai sentiment nya disemua data review kemudian disimpan dan di draf kembali untuk melihat nilai *accuracy* dari data yang telah berisi sentiment dengan metode *naive bayes*, maka didapat hasil nilai *accuracy* dari metode *naive bayes* untuk analisis sentiment review produk di *e-commerce* Amazon yang sangat tinggi yaitu mencapai 100.00%. Hal ini membuktikan metode *naive bayes* adalah metode yang dapat memprediksi sentiment dengan tingkat akurasi sangat baik.

	true Positive	true Neutral	true Negative	class precision
pred. Positive	32318	0	0	100.00%
pred. Neutral	0	3330	0	100.00%
pred. Negative	0	0	7478	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 8. Hasil *Accuracy Naive Bayes*

Tabel 5. Hasil *Accuracy*

	true Positive	true Neutral	true Negative	class precision
pred. Positive	32318	0	0	100.00%
pred. Neutral	0	3330	0	100.00%
pred. Negative	0	0	7478	100.00%
class recall	100.00%	100.00%	100.00%	

1. Menghitung Nilai Akurasi

$$\begin{aligned}
 \text{Accurarcy} &= \frac{TP+TNR+TN}{TP+TNR+TN+PP+PNR+PN} \times 100\% \\
 &= \frac{32318+3330+7478}{32318+3330+7478+0+0+0} \times 100\% \\
 &= \frac{43126}{43126} \times 100\% = 100\%
 \end{aligned}$$

2. Menghitung Class Recall

$$\text{Class recall positive} = \frac{TP}{TP+PNR+PN} \times 100\%$$

$$= \frac{32318}{32318 + 0 + 0} \times 100\% = 100\%$$

$$\text{Class Recall Neutral} = \frac{\text{TNR}}{\text{TNR} + \text{FP} + \text{FN}} \times 100\%$$

$$= \frac{3330}{3330 + 0 + 0} \times 100\% = 100\%$$

$$\text{Class Recall Negative} = \frac{\text{TN}}{\text{TN} + \text{PP} + \text{PN}} \times 100\%$$

$$= \frac{7478}{7478 + 0 + 0} \times 100\% = 100\%$$

3. Menghitung *Class Precision*

$$\text{Class Precision Positive}$$

$$= \frac{\text{TP}}{\text{TP} + \text{PP}} \times 100\%$$

$$= \frac{32318}{32318 + 0} \times 100\% = 100\%$$

$$\text{Class Precision Neutral}$$

$$= \frac{\text{TNR}}{\text{TNR} + \text{PNR}} \times 100\%$$

$$= \frac{3330}{3330 + 0} \times 100\% = 100\%$$

$$\text{Class Precision Negative}$$

$$= \frac{\text{TN}}{\text{TN} + \text{PN}} \times 100\%$$

$$= \frac{7478}{7478 + 0} \times 100\% = 100\%$$

Keterangan:

TP = True Positive

TNR = True Neutral

TN = True Negative

PP = Pred Positive

PNR = Pred Neutral

PN = Pred Negative

Performance		PerformanceVector			
Performance		PerformanceVector:			
		accuracy: 100.00%			
		ConfusionMatrix:			
		True:	Positive	Neutral	Negative
Description		Positive:	32318	0	0
		Neutral:	0	3330	0
		Negative:	0	0	7478
		kappa: 1.000			
		ConfusionMatrix:			
		True:	Positive	Neutral	Negative
Annotations		Positive:	32318	0	0
		Neutral:	0	3330	0
		Negative:	0	0	7478

Gambar 9. *Performance Vector*

Gambar 9 adalah tampilan hasil *performance* dalam bentuk *vector*



Gambar 10. *word cloud*

Gambar 10 adalah *word cloud* untuk visualisasi yang menampilkan kata-kata yang paling sering digunakan dalam data ulasan yang dibuat menggunakan *jupyter notebook*.

KESIMPULAN

Berdasarkan kesimpulan yang didapat peneliti dalam menggunakan metode *naïve bayes* terlihat bahwa tingkat akurasi dan ketepatan dalam menentukan sentiment sangat baik yang terlihat dalam penelitian ini, secara detail kesimpulannya adalah: 1) Analisis sentiment menggunakan metode *naïve bayes* sangat efektif untuk menfrediksi sentiment dari sebuah data review pelanggan. Metode ini sangat membantu untuk menentukan sentiment positif, negative, dan netral dari sebuah data review yang belum memiliki data sentiment. 2) Dari hasil process sentiment data untuk sentiment positif lebih tinggi dari data sentiment negative yang membuktikan bahwa lebih banyak pelanggan yang puas berbelanja di amazon daripada pelanggan yang merasa tidak puas. 3) Nilai *accuracy* atau tinggi ketepatan dalam menggunakan *naïve bayes* sangat tinggi dan baik dengan tinggkat *accuracy* sebesar 100.00%. Data dengan sentiment positif ada 37767 data, netral ada 3950 data dan negative ada 8766 data sentiment. 4) Dari data sentiment yang memiliki sentiment positif terbanyak diharapkan calon pelanggan untuk dapat memilih produk dengan review terbaik dan menjadikannya sebagai referensi sehingga waktu yang digunakan dalam mencari produk yang bagus dan berguna lebih cepat.

REFERENSI

Act Exchange. (2020). *AMAZON.COM, INC.* 2020. <https://www.sec.gov/ix?doc=/Archives/edgar/data/1018724/000101872421000004/amzn-20201231.htm>

Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK PENDUKUNG KEPUTUSAN PENERIMA BEASISWA. *Jurnal Data Mining Dan Sistem Informasi*, 2(1). <https://doi.org/10.33365/jdmsi.v2i1.1028>

Berrar, D. (2018). Bayes' theorem and naive bayes classifier. *Encyclopedia of Bioinformatics and*

- Computational Biology: ABC of Bioinformatics*, 1–3, 403–412. <https://doi.org/10.1016/B978-0-12-809633-8.20473-1>
- Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing Journal*, 50. <https://doi.org/10.1016/j.asoc.2016.11.022>
- Clinton, B. (2021). *Sejarah Amazon, Toko Buku Online yang Berubah Jadi Raksasa Marketplace*. Kompas.Com. <https://tekno.kompas.com/read/2021/05/05/20210037/sejarah-amazon-toko-buku-online-yang-berubah-jadi-raksasa-marketplace?page=all>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. 15(1), 131–145.
- Foo, L. K., Chua, S. L., & Ibrahim, N. (2022). Attribute weighted naïve bayes classifier. *Computers, Materials and Continua*, 71(1). <https://doi.org/10.32604/cmc.2022.022011>
- Hariguna, T., Baihaqi, W. M., & Nurwanti, A. (2019). Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes Classifier Algorithm. *IJIS: International Journal of Informatics and Information Systems*, 2(2). <https://doi.org/10.47738/ijis.v2i2.13>
- Khotimah, Aristin Chusnul, & Utami, E. (2022). Comparison Naïve Bayes Classifier, K-Nearest Neighbor and Support Vector Machine in the Classification of Individual on Twitter Account. *Jurnal Teknik Informatika (Jutif)*, 3(3).
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2).
- Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, 101. <https://doi.org/10.1016/j.jbusres.2018.12.051>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2).
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233–242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Subarkah, P., Damayanti, W. R., & Permana, R. A. (2022). Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure. *ILKOM Jurnal Ilmiah*, 14(2). <https://doi.org/10.33096/ilkom.v14i2.1148.120-125>
- Trivusi. (2022). *Pengertian dan Contoh Algoritma Naive Bayes Classifier*. Trivusi.Com. <https://www.trivusi.web.id/2022/07/algoritma-naive-bayes.html>
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1). <https://doi.org/10.1177/0165551516677946>
- Zhang, Y. C., & Sakhanenko, L. (2019). The naive Bayes classifier for functional data. *Statistics and Probability Letters*, 152. <https://doi.org/10.1016/j.spl.2019.04.017>