

ANALISIS SENTIMEN TERHADAP XIAOMI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES

Ranita Wahyu Utami¹, Ahmad Jazuli², Tutik Khotimah³

¹²³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus
Email: ¹201851242@std.umk.ac.id, ²ahmad.jazuli@umk.ac.id, ³tutik.khotimah@umk.ac.id

(Naskah masuk: 29 Desember 2021, diterima untuk diterbitkan: 31 Desember 2021)

Abstrak

Seiring berkembangnya teknologi, hampir seluruh aspek kehidupan manusia menggunakan teknologi untuk membantu tugasnya. Peran teknologi pada pekerjaan, dunia Pendidikan, maupun sebagai hiburan. Salah satu perusahaan elektronik yang digandrungi masyarakat adalah brand Xiaomi. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat terhadap brand Xiaomi menggunakan metode *Naïve Bayes Classifier* (NBC). Penelitian ini mengambil data dari media sosial *twitter* menggunakan teknik *scrapping* sebanyak 2.078 data *tweet*. Dalam proses analisis ini menggunakan library *text blob* untuk mendapatkan *labelling* otomatis dari data *tweet*. Berdasarkan hasil penelitian, analisis sentimen terhadap brand Xiaomi mendapatkan akurasi sebesar 71.88% dengan data polaritas positif 39%, polaritas netral 51%, dan polaritas negatif sebesar 10%.

Kata kunci: *xiaomi, twitter, text blob, naïve bayes classifier*

ANALYSIS OF SENTIMEN ON XIAOMI INDONESIA USING NAÏVE BAYES METHOD INGGRIS

Abstract

As technology develops, almost all aspects of human life use technology to help with their tasks. The role of technology in work, the world of education, and as entertainment. One of the electronic companies loved by the public is the Xiaomi brand. This study aims to determine public sentiment towards the Xiaomi brand using the Naïve Bayes Classifier (NBC) method. This study took data from social media twitter using scrapping techniques as much as 2,078 tweet data. In this analysis process, we use the text blob library to get automatic labeling of tweet data. Based on the results of the study, sentiment analysis on the Xiaomi brand obtained an accuracy of 71.88% with 39% positive polarity data, 51% neutral polarity, and 10% negative polarity.

Keywords: *xiaomi, twitter, text blob, naïve bayes classifier*

1. PENDAHULUAN

Mengungkapkan pendapat merupakan aktivitas yang sering dilakukan. Dewasa ini, mengungkapkan pendapat dapat dilakukan melalui platform media sosial. Dalam kamus KBBI, pendapat atau pandangan merupakan sesuatu yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu yang bertentangan dan pertimbangan pada pikiran. Seiring dengan perkembangan teknologi, saat ini masyarakat menggunakan media sosial untuk mengemukakan pendapat, sentimen, maupun pandangannya kepada

masyarakat luas. Media sosial merupakan sebuah wadah yang digunakan oleh masyarakat untuk melakukan interaksi dengan manusia lain untuk menciptakan pengalaman baru, membagikan momen, dan bertukar informasi dalam sebuah wadah virtual. Media sosial yang saat ini digandrungi masyarakat adalah, *Tik Tok, Facebook, Instagram, Youtube*, dan *Twitter*.

Twitter merupakan salah satu platform media sosial yang digunakan masyarakat sebagai wadah untuk menyampaikan pendapat, gagasan, maupun sebuah

sentimen. *Twitter* adalah jejaring sosial paling besar yang didirikan oleh Jack Dorsey serta diluncurkan untuk umum pada tanggal 15 Juli 2006. Berdasarkan data yang diperoleh dari PT. Bakrie Telecom, media sosial *Twitter* kini memiliki 19,5 juta pengguna di Indonesia dari total seluruh pengguna global sebanyak 500 juta. *Twitter* menjadi salah satu jejaring sosial paling besar di dunia sehingga mampu memperoleh laba mencapai USD 145 juta (kominfo.go.id, 2013).

Semakin berkembangnya jaman, hampir seluruh aspek kehidupan manusia dibantu alat. Alat untuk membantu pekerjaan, sebagai pendukung media pembelajaran, bahkan sebagai hiburan. Salah satu perusahaan elektronik yang di gandrungi masyarakat adalah brand *Xiaomi*. Perusahaan tersebut berasal dari China. Brand tersebut mengeluarkan berbagai macam produk seperti gawai, laptop, *smartwatch*, televisi dan lain sebagainya. Dalam setiap produk yang diluncurkan pasti mendapat berbagaimacam pendapat / sentimen dari masyarakat.

Sehingga berdasarkan latar belakang tersebut penelitian ini dibuat untuk menganalisis sentimen konsumen terhadap produk *Xiaomi* di media sosial *Twitter* dengan menggunakan metode *Naïve Bayes Classifier* (NBC).

1.1. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas, maka permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana sentimen masyarakat terhadap produk brand *Xiaomi* di *Twitter* dan mengklasifikasikan dengan metode *Naïve Bayes Classifier* (NBC).

1.2. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Pengambilan data hanya pada media sosial *Twitter*.
2. Data yang dianalisis hanya *tweet* yang menggunakan Bahasa Indonesia.
3. Data yang dianalisis hanya *tweet* terdapat kata kunci *xiaomi*.
4. *Tweet* yang digunakan hanya *tweet* yang berupa *text*.

1.3. Tujuan Penelitian

Adapun tujuan yang akan dicapai dari penelitian ini adalah untuk mengetahui sentimen masyarakat terhadap produk brand *Xiaomi* dengan mengimplementasikan metode *Naïve Bayes Classifier* (NBC) pada sistem analisis sentiment produk brand *Xiaomi* pada media sosial *Twitter*.

2. TINJAUAN PUSTAKA

2.1. *Xiaomi*

Xiaomi adalah perusahaan elektronik swasta milik pribadi dari Tiongkok yang didirikan pada tahun 2010. *Xiaomi* merupakan distributor gawai terbesar ke-3 di dunia yang merancang,

mengembangkan, dan menjual gawai, aplikasi seluler, dan elektronik seluler. Tak hanya gawai, *xiaomi* juga mempunyai produk lain yang diminati masyarakat. Produk elektronik yang dikeluarkan oleh *xiaomi* seperti laptop, televisi, *smartwatch*, *cctv* dan masih banyak lagi (Apriyani dan Suharti 2017).

2.2. *Twitter*

Twitter adalah salah satu *platform* media sosial yang terkenal dan telah memperoleh perhatian yang spesial karena pengguna memperoleh kemudahan dalam menyalurkan informasi tentang pendapat / opini mereka tentang topik tertentu melalui pesan yang ditujukan untuk umum, yang disebut sebagai *tweet*. Selain informasi yang dibagikan secara sukarela oleh pengguna, *tweet* juga dapat menyimpan berbagai berita yang berkaitan dengan lokasi pengguna serta memungkinkan berisi informasi seperti profil pengguna, lokasi, emoji, tautan, dan tagar yang bisa membantu pengguna mengekspresikan perasaannya dengan lebih baik dan menjadikannya sumber informasi yang berharga (D'Andrea et al. 2019).

2.3. Analisis Sentimen

Analisis sentimen adalah sebuah metode dalam mengumpulkan komentar beserta ulasan orang lain tentang suatu hal seperti informasi pada suatu media soial yang berbasis web (Wardani dan Erfina 2021).

Sedangkan menurut D'Andrea (2019) analisis sentimen adalah salah satu penelitian yang cakupannya luas seperti sebagai penelitian tentang opini sentimen, emosi, dan suasana hati yang diungkapkan dalam teks serta dapat mendeteksi sentimen seperti positif, negatif, dan netral.

3. METODOLOGI PENELITIAN

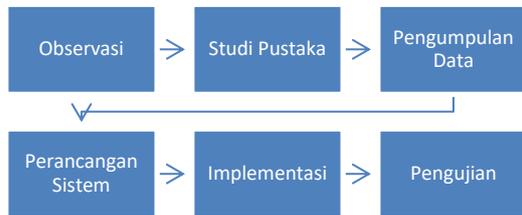
3.1. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data berupa *tweet* yang berbahasa Indonesia. *Tweet* dikumpulkan dari media sosial *Twitter*. Data dikumpulkan menggunakan *Twitter* API. Data yang dikumpulkan adalah *tweet* dengan kata kunci "*xiaomi*". Data yang diambil sebanyak 2.078 data dari tanggal 25 Januari – 3 Februari 2022. Data yang dikumpulkan dibagi menjadi dua jenis data yaitu data latih dan data uji dengan perbandingan 80:20 yang akan digunakan untuk menguji algoritma *naive bayes*.

3.2. Metode Penelitian

Metode penelitian terdiri dari langkah-langkah yang peneliti lakukan untuk mengumpulkan informasi atau data dan melakukan penelitian terhadap data yang didapatkan. Metode penelitian yang digunakan penulis dalam

perancangan sistem analisis dapat dilihat pada gambar 3.1.



Gambar 3.1. Diagram Proses Penelitian

3.2.1 Observasi

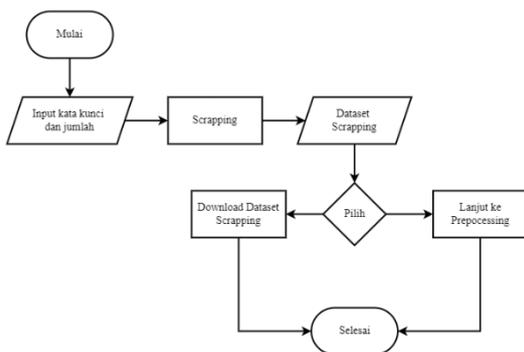
Observasi merupakan kegiatan untuk mempelajari dan mengamati komentar / *tweet* pada *Twitter* untuk memperoleh gambaran mengenai data yang akan digunakan dalam penelitian.

3.2.2 Studi Pustaka

Studi pustaka merupakan kegiatan mengumpulkan informasi yang dibutuhkan untuk melakukan suatu penelitian. Sumber informasi bisa didapatkan dari jurnal, tesis, dan sumber lain yang mengusung tema tentang analisis sentimen.

3.2.3 Pengumpulan Data

Pengumpulan data *twitter* dalam penelitian ini menggunakan metode *scrapping*. *Scrapping* adalah proses pengambilan data *twitter* menggunakan *library tweepy* yang memberikan akses API *Twitter*. Lalu data tersebut disimpan dalam format CSV untuk digunakan dalam proses selanjutnya. Proses pengumpulan data dengan metode *scrapping* dapat dilihat pada gambar 3.2.

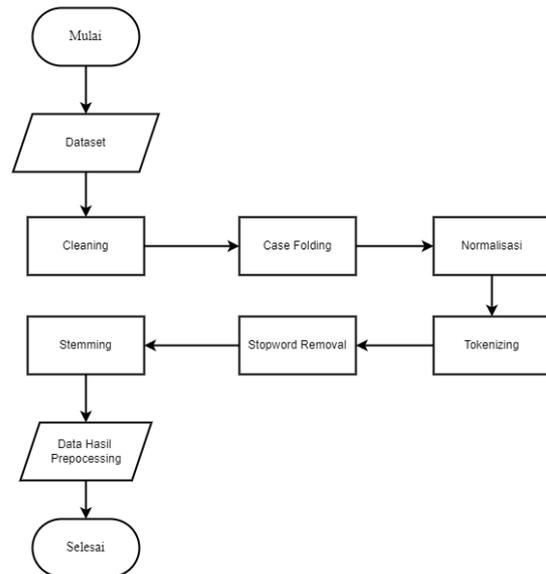


Gambar 3.2. Diagram Alir Pengumpulan Data

3.2.4 Perancangan Sistem

Setelah proses *scrapping*, proses selanjutnya adalah *preprocessing*. *Preprocessing* merupakan proses pembersihan data sebelum dilakukan proses pengolahan data lebih lanjut. Proses ini bertujuan untuk membersihkan data, menyeragamkan data, mengurangi *noise* data, dan membentuk data yang konsisten. Dalam *preprocessing* ada beberapa proses yaitu *cleaning*, *tokenizing*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Kemudian data *preprocessing* disimpan dalam

format CSV. Proses perancangan sistem dengan *preprocessing* data dapat dilihat pada gambar 3.3.



Gambar 3.3. Diagram Alir Perancangan Sistem

Adapun tahapan dalam *preprocessing* data adalah sebagai berikut:

- 1) *Cleaning*
Cleaning adalah tahapan dimana karakter dan tanda baca yang tidak dibutuhkan dihilangkan dari teks. Hal tersebut berfungsi untuk mengurangi *noise* pada dataset. Contoh karakter yang dihilangkan dalam proses ini seperti URL, tag (#), tanda baca seperti titik (.), koma (,) dan tanda baca lainnya (Yulita et al. 2021). Tidak hanya tanda baca, *cleansing* juga dapat menghilangkan angka, *whitespace*, dan emoji yang terdapat pada sebuah *tweet*.
- 2) *Case folding*
Case folding adalah proses penyeragaman / pengubahan huruf baik huruf kecil maupun huruf besar. Karena dalam penulisan sebuah *tweet* pasti ada perbedaan bentuk huruf (Yulita et al. 2021).
- 3) *Normalisasi*
Normalisasi adalah proses mengubah kata yang tidak lazim, singkatan, dan tidak baku menjadi sebuah kata yang baku sesuai dengan pedoman pada KBBI. Contohnya seperti *bngt* menjadi *banget*, *mkn* menjadi *makan*, *ak* menjadi *aku* dan lain sebagainya (Rozi et al. 2019).
- 4) *Tokenizing*
Tokenizing adalah proses pemotongan string *input* berdasarkan tiap kata yang menyusunnya. Data *tweet* yang diambil akan dipotong menjadi sebuah token yang dipisahkan dengan tanda petik dan koma (Rozi et al. 2018).
- 5) *Stopword removal*
Stopword removal adalah proses penghapusan pada sebuah kalimat jika ada kata / *term* yang tidak dianggap tidak terlalu penting. Oleh karena

itu tujuan *stopword removal* ini untuk menghilangkan / menghapus kata / *term* yang dianggap tidak penting seperti waktu, penghubung, dan lain sebagainya agar memudahkan dalam proses klasifikasi (Fatayat dan Nugroho 2021).

6) *Stemming*

Stemming adalah proses untuk menghilangkan imbuhan pada masing-masing kata yang telah diproses sehingga menjadi kata dasar, dan dalam tahapan ini juga mempunyai tujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat (Wardani dan Erfina 2021).

7) Pelabelan Data / *labelling*

Text blob merupakan salah satu *library* dari *python* untuk melakukan pemrosesan bahasa alami (NLP). *Library text blob* akan menghitung nilai polaritas dari sebuah data untuk mendapatkan label atau sentimen pada data tersebut. Nilai polaritas merupakan suatu fungsi untuk melihat kecenderungan sentiment pada sebuah data. Hasil dari nilai polaritas akan digolongkan menjadi kelas positif, kelas netral, dan kelas negatif (Baita et al. 2021).

3.2.5 Implementasi

Naive Bayes adalah metode yang diterapkan untuk memprediksi karena mengandung probabilistik sederhana yang digunakan pada teorema bayes dengan ketergantungan, yang kuat (Sinaga, Sawaluddin, & Suwilo, 2020). Proses klasifikasi dengan metode *Naive Bayes* menggunakan dataset *labelling* yang sudah dilakukan oleh sistem. Proses klasifikasi menggunakan *library sklearn*.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{1}$$

Keterangan:

- A : Hipotesis data pada kelas yang spesifik.
- B : Data kelas yang belum diketahui.
- P(A|B) : Probabilitas hipotesis A berdasarkan pada kondisi B.
- P(A) : Probabilitas A.
- P(B|A) : Probabilitas B berdasarkan pada hipotesis H.
- P(B) : Probabilitas B.

3.2.6 Pengujian

Dalam mengevaluasi performance algoritma dari *Machine Learning* (ML) kita menggunakan *Confusion Matrix*. *Confusion Matrix* adalah merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma (Winahyu dan Suharjo 2021). Dengan menggunakan *Confusion Matrix*, kita bisa menentukan *Accuracy*, dengan yang perhitungan sebagai berikut:

- 1) *Accuracy* merupakan perhitungan untuk menghilangkan label prediksi yang sesuai pada label aktual. Jika hasil akurasi yang didapatkan pada perhitungan lebih tinggi dari prediksinya, maka algoritma yang digunakan baik.

$$accuracy = \frac{TP + TNN}{TP + FP + TN + FN} \times 100\% \tag{2}$$

- 2) *Precision* merupakan perhitungan perbandingan data yang dikenal didalam sistem dengan jumlah dalam dokumen yang dikenal sistem.

$$precision = \frac{TP}{TP + FP} \tag{3}$$

- 3) *Recall* merupakan perhitungan perbandingan total data berkaitan dengan semua total data yang berkaitan.

$$recall = \frac{TP}{TP + FN} \tag{4}$$

- 4) Nilai *f1-score* merupakan suatu nilai yang digunakan untuk mengetahui nilai kinerja dari algoritma *Naive Bayes*.

$$f1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

4. HASIL DAN PEMBAHASAN

4.1 Pengambilan Data

Tahap paling awal dalam penelitian ini ialah proses mengumpulkan data tweet menggunakan *tweepy* dengan pemrograman *pyhon*. Dataset yang diambil dalam proses *scrapping* memiliki keterbatasan dalam mengambil data. Data yang diambil dalam penelitian ini adalah *tweet* terbaru dalam 7 hari kebelakang saat melakukan *scrapping* data. Syarat dalam melakukan *scrapping* adalah perangkat yang digunakan harus terhubung dengan internet dan mempunyai API *Twitter* dengan *key* dan token sebagai autentikasi untuk menghubungkan suatu program. *Scrapping* dilakukan dengan batasan menggunakan kata kunci *xiaomi*. Setelah kata kunci ditetapkan maka selanjutnya adalah mengisi berapa banyak data yang akan diambil didalam penelitian ini. saat proses *scrapping* selesai maka data mentah tersebut akan disimpan dalam bentuk file .CSV untuk dapat dilanjutkan pada tahap selanjutnya.

Tabel 4.1. Data Hasil Proses Scrapping

No	Tanggal	User name	Tweet
1	03/02/2022 07:52:03	clbair	@ssefnum ini aja aku xiaomi note 5 pro awet dari 2016 sampe skrg 2022 sekitar 2,5jt dulu skrg udh 1,5jt keknya

2	03/02/20 22 05:16:32	zainal 0403	@MIUI_Indonesia kapan diperbarui. Bug nya meresahkan @Xiaomi #RN9P https://t.co/WDW9ec5uvb
3	03/02/20 22 04:20:12	komp asco m	Kamera Xiaomi 11T dipakai untuk mengambil foto dengan berbagai suasana yang berbeda. Bagaimana kemampuan kamera ponsel flagship Xiaomi ini? https://t.co/AK9A1YuHDv

4.2 Preprocessing Data

Tahap setelah *scrapping* data *tweet* adalah *preprocessing text* pada *tweet*. Dalam proses ini data *tweet* akan dibersihkan dari angka, simbol, tagar, tautan, dan emoji. Lalu kata tidak penting dalam *tweet* akan dihilangkan, serta *tweet* akan mengalami proses pengubahan kata tidak baku menjadi kata baku.

1) Hasil *cleaning*

Proses *cleaning* diperlukan untuk menghilangkan atau menghapus tanda baca, tautan, tagar, *mention*, dan angka agar data yang akan digunakan dalam keadaan bersih dan data menjadi efektif. Sistem akan mencari karakter khusus yang telah ditentukan, lalu karakter tersebut akan dihilangkan dari *tweet* tersebut. Contoh dari proses *cleaning* dapat dilihat pada tabel 4.2.

Tabel 4.2. Data hasil proses *cleaning*

No.	<i>Tweet</i>	<i>Cleaning</i>
1.	@sfeffnum ini aja aku xiaomi note 5 pro awet dari 2016 sampe skrg 2022 sekitar 2,5jt dulu skrg udh 1,5jt keknya	ini aja aku xiaomi note pro awet dari sampe skrg sekitar jt dulu skrg udh jt keknya
2.	@MIUI_Indonesia kapan diperbarui. Bug nya meresahkan @Xiaomi #RN9P https://t.co/WDW9ec5uvb	Indonesia kapan diperbarui Bug nya meresahkan RNP
3.	Kamera Xiaomi 11T dipakai untuk mengambil foto dengan berbagai suasana yang berbeda. Bagaimana	Kamera Xiaomi dipakai untuk mengambil foto dengan berbagai suasana yang berbeda

No.	<i>Tweet</i>	<i>Cleaning</i>
	kemampuan kamera ponsel flagship Xiaomi ini? https://t.co/AK9A1YuHDv	Bagaimana kemampuan kamera ponsel flagship Xiaomi ini

2) Hasil *casefolding*

Proses *case folding* dilakukan untuk menyeragamkan huruf menjadi kecil semua (*lower case*) agar huruf data menjadi seragam. Pada proses ini sistem akan menelusuri semua *tweet* dari awal sampai akhir dan jika ditemukan huruf kapital akan diubah menjadi huruf kecil (*lower case*). Contoh dari proses *case folding* dapat dilihat pada tabel 4.3.

Tabel 4.3. Data hasil proses *casefolding*

No.	<i>Cleaning</i>	<i>Casefolding</i>
1.	ini aja aku xiaomi note pro awet dari sampe skrg sekitar jt dulu skrg udh jt keknya	ini aja aku xiaomi note pro awet dari sampe skrg sekitar jt dulu skrg udh jt keknya
2.	Indonesia kapan diperbarui Bug nya meresahkan RNP	indonesia kapan diperbarui bug nya meresahkan rnp
3.	Kamera Xiaomi dipakai untuk mengambil foto dengan berbagai suasana yang berbeda Bagaimana kemampuan kamera ponsel flagship Xiaomi ini	kamera xiaomi dipakai untuk mengambil foto dengan berbagai suasana yang berbeda bagaimana kemampuan kamera ponsel flagship xiaomi ini

3) Hasil Normalisasi

Proses normalisasi merupakan untuk mengubah kata singakatan, kata gaul, dan kata yang tidak baku menjadi baku. Normalisasi dapat dilakukan dengan menambahkan kamus bahasa tidak baku.csv pada sistem. Pada proses ini sistem akan menelusuri semua *tweet* dari awal sampai akhir dan jika ditemukan kata yang tidak sesuai KBBI akan diubah menjadi kata baku, seperti mkn menjadi makan, bngt menjadi banget, dan skrg menjadi sekarang. Contoh dari proses *case folding* dapat dilihat pada tabel 4.4.

Tabel 4.4. Data hasil proses normalisasi

No.	Casefolding	Normalisasi
1.	ini aja aku xiaomi note pro awet dari sampe skrg sekitar jt dulu skrg udh jt keknya	ini saja aku xiaomi note pro awet dari sampai sekarang sekitar juta dulu sekarang sudah juta kayaknya
2.	Indonesia kapan diperbarui Bug nya meresahkan RNP	indonesia kapan diperbarui bug nya meresahkan rnp
3.	Kamera Xiaomi dipakai untuk mengambil foto dengan berbagai suasana yang berbeda Bagaimana kemampuan kamera ponsel flagship Xiaomi ini	kamera xiaomi dipakai untuk mengambil foto dengan berbagai suasana yang berbeda bagaimana kemampuan kamera ponsel flagship xiaomi ini

- 4) Hasil tokenizing
 Proses *tokenizing* merupakan tahap pertama dalam *Natural Language Processing* (NLP). *Tokenizing* dilakukan untuk membagi string menjadi token dengan menggunakan *library* NTLK. Pada proses ini sistem akan menelusuri semua *tweet* dari awal sampai akhir dan akan mengubah string menjadi token dari keseluruhan *tweet* dan diapit dengan tanda petik dan dipisahkan dengan tanda koma. Contoh dari proses *tokenizing* dapat dilihat pada tabel 4.5.

Tabel 4.5. Data hasil proses tokenizing

No.	Normalisasi	Tokenizing
1.	ini saja aku xiaomi note pro awet dari sampai sekarang sekitar juta dulu sekarang sudah juta kayaknya	['ini', 'saja', 'aku', 'xiaomi', 'note', 'pro', 'awet', 'dari', 'sampai', 'sekarang', 'sekitar', 'juta', 'dulu', 'sekarang', 'sudah', 'juta', 'kayaknya']
2.	indonesia kapan diperbarui bug nya meresahkan rnp	['indonesia', 'kapan', 'diperbarui', 'bug', 'nya', 'meresahkan', 'rnp']
3.	kamera xiaomi dipakai untuk mengambil foto dengan berbagai	['kamera', 'xiaomi', 'dipakai', 'untuk', 'mengambil', 'foto', 'dengan', 'berbagai', 'suasana', 'yang', 'berbeda', 'bagaimana', 'kemampuan', 'kamera', 'ponsel', 'flagship', 'xiaomi', 'ini']

No.	Normalisasi	Tokenizing
	suasana yang berbeda bagaimana kemampuan kamera ponsel flagship xiaomi ini	'suasana', 'yang', 'berbeda', 'bagaimana', 'kemampuan', 'kamera', 'ponsel', 'flagship', 'xiaomi', 'ini']

- 5) Hasil *stopword removal*
 Proses *stopword removal* merupakan proses untuk menghilangkan atau kata yang tidak penting dalam suatu *tweet*. Kata tersebut dihapus karena tidak mengandung arti yang mempresentasikan dari *tweet* tersebut. Pada proses ini sistem akan menelusuri semua *tweet* dari awal sampai akhir dan jika ditemukan kata yang tidak penting, maka akan dihapus. Proses ini menggunakan data *stopword* dari *library* dan juga dapat ditambahkan dengan data *stopword* milik sendiri. Contoh dari proses *stopword removal* dapat dilihat pada tabel 4.6.

Tabel 4.6. Data hasil proses tokenizing

No.	Tokenizing	Stopword removal
1.	['ini', 'saja', 'aku', 'xiaomi', 'note', 'pro', 'awet', 'dari', 'sampai', 'sekitar', 'juta', 'dulu', 'sekarang', 'sudah', 'juta', 'kayaknya']	['aja', 'aku', 'xiaomi', 'note', 'pro', 'awet', 'sampe', 'skrg', 'jt', 'dulu', 'skrg', 'udh', 'jt', 'keknya']
2.	['indonesia', 'kapan', 'diperbarui', 'bug', 'nya', 'meresahkan', 'rnp']	['indonesia', 'kapan', 'diperbarui', 'bug', 'meresahkan', 'rnp']
3.	['kamera', 'xiaomi', 'dipakai', 'untuk', 'mengambil', 'foto', 'dengan', 'berbagai', 'suasana', 'yang', 'berbeda', 'bagaimana', 'kemampuan', 'kamera', 'ponsel', 'flagship', 'xiaomi', 'ini']	['kamera', 'xiaomi', 'dipakai', 'mengambil', 'foto', 'berbagai', 'suasana', 'berbeda', 'bagaimana', 'kemampuan', 'kamera', 'ponsel', 'flagship', 'xiaomi']

- 6) Hasil *stemming*
 Proses *stemming* merupakan proses untuk menghapus imbuhan dari sebuah *tweet* agar menjadi kata dasar dengan menggunakan *library* sastrawi yang berbahasa Indonesia. Pada proses ini sistem akan menelusuri

semua *tweet* dari awal sampai akhir dan jika ditemukan kata berimbuhan, maka kata berimbuhan tersebut akan diubah menjadi kata dasar. Contoh dari proses *stemming* dapat dilihat pada tabel 4.7.

Tabel 4.7. Data hasil proses *stemming*

No.	Stopword removal	Stemming
1.	['ini', 'saja', 'aku', 'xiaomi', 'note', 'pro', 'awet', 'dari', 'sampai', 'sekarang', 'sekitar', 'juta', 'dulu', 'sekarang', 'sudah', 'juta', 'kayaknya']	aja aku xiaomi note pro awet sampe skrg jt dulu skrg udh jt kek
2.	['indonesia', 'kapan', 'diperbarui', 'bug', 'nya', 'meresahkan', 'mp']	indonesia kapan baru bug resah mp
3.	['kamera', 'xiaomi', 'dipakai', 'untuk', 'mengambil', 'foto', 'dengan', 'berbagai', 'suasana', 'yang', 'berbeda', 'bagaimana', 'kemampuan', 'kamera', 'ponsel', 'flagship', 'xiaomi', 'ini']	kamera xiaomi pakai ambil foto bagai suasana beda bagaimana mampu kamera ponsel flagship xiaomi

4.3 Labelling Menggunakan Text Bloob

Proses *labelling* merupakan proses mengetahui sentimen dari sebuah *tweet* yang dapat diberikan label positif, negatif, dan netral. Pada proses *labelling* ini menggunakan *library text bloob*. *Text bloob* merupakan salah satu *library* yang disediakan oleh python untuk melakukan pemrosesan pada bidang *Natural Language Processing (NLP)*. Namun *library text bloob* belum tersedia untuk bahasa Indonesia, jadi dataset dari hasil *stemming* yang berbahasa Indonesia diterjemahkan kedalam bahasa inggris untuk dilakukan pemrosesan dengan *text bloob*. Penentuan *labelling* sebuah *tweet* dengan menggunakan nilai polaritas. *Tweet* dengan nilai 1 merupakan kelas nilai positif, 0 merupakan kelas nilai netral, sedangkan -1 menunjukan kelas nilai negatif. Contoh dari proses *labelling* dapat dilihat pada tabel 4.7.

Tabel 4.7. Data hasil proses *stemming*

No.	Tweet bersih	Hasil labelling
1.	aja aku xiaomi note pro awet sampe skrg jt dulu skrg udh jt kek	Netral

No.	Tweet bersih	Hasil labelling
2.	indonesia kapan baru bug resah mp	Positif
3.	kamera xiaomi pakai ambil foto bagai suasana beda bagaimana mampu kamera ponsel flagship xiaomi	Netral

4.4 Klasifikasi

Dataset *tweet* yang telah melalui proses *preprocessing* dan *labelling* menggunakan *text bloob* akan diproses pada tahap klasifikasi menggunakan metode *Naïve Bayes*. Pada sistem ini proses klasifikasi dapat menggunakan dua cara. yang pertama cara otomatis dengan klik Lanjutkan pada proses *labelling* maka data akan langsung dilanjutkan pada proses klasifikasi. Yang kedua yaitu dengan cara *upload file labelling.csv* lalu data akan di klasifikasikan. Adapun *Interface* dari halaman klasifikasi dapat dilihat pada gambar 4.1.



Gambar 4.1 Interface Halman Upload Klasifikasi

Proses selanjutnya yaitu data hasil *labelling* menggunakan *text bloob* akan dilakukan pengujian menggunakan *confusion matrix*. Perhitungan yang dilakukan akan menghasilkan akurasi dari metode *naïve bayes*, *recall*, *precision*, *f1-score*, dan *support*. Hasil perhitungan dari *confusion matrix* pada gambar 4.17 adalah sebagai berikut:

a) *Precision*

Precision merupakan hasil perhitungan dari nilai *predicted true* positif dibagi dengan total nilai *predicted* positif. Hasil perhitungan dari *precision* adalah positif 0,66, netral 0,77, dan negatif 0,0. Adapun penjabaran perhitungan dari *precision* adalah sebagai berikut:

$$Precision \text{ positif} = \frac{130}{130+43+23} = \frac{130}{196} = 0,66$$

$$Precision \text{ netral} = \frac{169}{31+169+20} = \frac{169}{220} = 0,77$$

$$Precision \text{ negative} = \frac{0}{0} = 0$$

b) *Recall*

Recall merupakan hasil perhitungan dari nilai *predicted true* positif dibagi dengan total nilai positif. Hasil perhitungan dari *recall* adalah positif 0,80, netral 0,79, dan

negatif 0,0. Adapun penjabaran perhitungan dari *recall* adalah sebagai berikut:

$$Recall\ positif = \frac{130}{130+31+0} = \frac{130}{161} = 0,80$$

$$Recall\ netral = \frac{169}{43+169+0} = \frac{169}{212} = 0,79$$

$$Recall\ negative = \frac{0}{23+20+0} = 0$$

c) Nilai *f1-score*

F1-score merupakan *average* (rata-rata) dari hasil perhitungan *precision* dan *recall*. Hasil perhitungan dari *recall* adalah positif 0,72, netral 0,78, dan negatif 0,0. Adapun penjabaran perhitungan dari *F1-score* adalah sebagai berikut:

$$F1\text{-score}\ positif = 2 \times \frac{(0,66 * 0,80)}{(0,66 + 0,80)} = 2 \times \frac{0,53}{1,47} = 0,72$$

$$F1\text{-score}\ netral = 2 \times \frac{(0,77 * 0,79)}{(0,77 + 0,79)} = 2 \times \frac{0,61}{1,57} = 0,78$$

$$F1\text{-score}\ negatif = 2 \times \frac{(0 * 0)}{(0 + 0)} = 0$$

d) Akurasi

Akurasi merupakan nilai *predicted* dari hasil perhitungan *true* positif dan *true* negative dari data yang diambil. Hasil perhitungan dari akurasi adalah 71,88. Adapun penjabaran perhitungan dari akurasi adalah sebagai berikut:

Akurasi

=

$$\frac{(130 + 169 + 0)}{(130 + 31 + 0 + 43 + 169 + 0 + 23 + 20 + 0)} \times 100\%$$

$$= 71,88$$

Adapun interface dari halaman hasil klasifikasi dapat dilihat pada gambar 4.20.

Klasifikasi Naive Bayes

	PRECISION	RECALL	F1-SCORE	SUPPORT
Negatif	0.000000	0.000000	0.000000	43.000000
Netral	0.768182	0.797700	0.782407	212.000000
Positif	0.663265	0.807453	0.726291	161.000000
accuracy	0.718750	0.718750	0.718750	0.71875
macro avg	0.477149	0.534874	0.503566	416.000000
weighted avg	0.648774	0.718750	0.680590	416.000000

Akurasi Naive Bayes 71.88 %

Confusion Matrix

	NEGATIF	NETRAL	POSITIF
Negatif	0	20	23
Netral	0	169	43
Positif	0	31	130

Gambar 4.2 Interface Hasil Klasifikasi

4.5 Visualisasi

Hasil akhir dari klasifikasi akan ditampilkan dalam bentuk diagram batang, diagram lingkaran, dan visualisasi dengan *wordcloud*. Hasil visualisasi dalam sistem ini dapat dilihat pada gambar 4.21.



Gambar 4.3 Interface Halaman Visualisasi

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari hasil penelitian analisis sentimen terhadap produk brand xiaomi menggunakan metode naïve bayes dengan menggunakan data sebanyak 2.078 tweet adalah sebagai berikut:

1. Penggunaan algoritma *Naïve Bayes Classifier* dalam klasifikasi dapat menghasilkan akurasi sebesar 71,88% dengan nilai polaritas positif 39%, netral 51%, dan negative sebesar 10%.
2. Sistem analisis sentimen dapat menampilkan hasil klasifikasi dengan diagram batang, diagram lingkaran, dan visualisasi dengan *wordcloud*.

5.2 Saran

Adapun saran untuk pengembangan penelitian analisis sentimen pada penelitian mendatang adalah sebagai berikut:

1. Penelitian ini menggunakan metode *Naïve Bayes Classifier*, pada penelitian selanjutnya diharapkan dapat menggunakan metode lain untuk membandingkan akurasi dari algoritma.
2. Dataset pada penelitian ini tidak seimbang, sehingga menyebabkan akurasi rendah, pada penelitian selanjutnya diharapkan dapat menambahkan metode smoot untuk mengatasi dataset yang tidak seimbang.
3. Penelitian ini diimplementasikan pada *platform* web, diharapkan pada penelitian selanjutnya dapat menggunakan *platform* yang lain.
4. Penelitian ini menggunakan media sosial *Twitter* untuk pengambilan data atau *scrapping*, pada penelitian selanjutnya diharapkan dapat menggunakan media sosial yang lain.

DAFTAR PUSTAKA

- APRIYANI, N. and SUHARTI, S., 2017. Analisis Pengaruh Persepsi Kebermanfaatan, Persepsi Kemudahan dan Kepercayaan Terhadap Minat Beli Ulang Pengguna Smartphone Xiaomi. *Manajemen Dewantara*, 1(1), pp.21-34.
- BAITA, A., PRISTYANTO, Y. and CAHYONO, N., 2021. Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). *Information System Journal*, 4(2), pp.42-46.
- D'ANDREA, E., DUCANGE, P., BECHINI, A., RENDA, A. and MARCELLONI, F., 2019. Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116, pp.209-226.
- D'ANDREA, E., DUCANGE, P., BECHINI, A., RENDA, A. and MARCELLONI, F., 2019. Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116, pp.209-226.
- FATAYAT, F. AND NUGROHO, R.A., 2021. Analisa Penentuan Dosen Pembimbing Tugas Akhir Mahasiswa Menggunakan Naive Bayes Classifier. *Simitika*, 4(3), pp.1-7.
- Kominfo. (Rmg). 2018. dilansir pada 07 November 2013 from https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+Orang/0/berita_satker.
- ROZI, I.F., ARDIANSYAH, R. AND REBEKA, N., 2019. Penerapan Normalisasi Kata Tidak Baku Menggunakan Levenshtein Distance pada Analisa Sentimen Layanan PT. KAI di Twitter. In *Seminar Informatika Aplikatif Polinema* (pp. 106-111).
- ROZI, I.F., HAMDANA, E.N. and ALFAHMI, M.B.I., 2018. Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus SAMSAT Kota Malang). *Jurnal Informatika Polinema*, 4(2), pp.149-149.
- SINAGA, L.M. and SUWILO, S., 2020. Analysis of classification and Naïve Bayes algorithm k-nearest neighbor in data mining. In *IOP Conference Series: Materials Science and Engineering* (Vol. 725, No. 1, p. 012106). IOP Publishing.
- WARDANI, N.R. and ERFINA, A., 2021, September. Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes. In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 11-18).
- WINAHYU, J. and SUHARJO, I., 2021. Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, 10(2), pp.206-214.
- YULITA, W., 2021. Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), pp.1-9.