

Analisis Sentimen Churn Pelanggan dalam Layanan Streaming NETFLIX di X Menggunakan Metode IndoBERT

Farencia Levis¹, Cindy Chuwardi², Yoshe Wuvanka³, Eveleen Huandra⁴

^{1,2,3,4}Sistem Informasi, Universitas Pelita Harapan, Medan, Indonesia

Email: ¹03081220029@student.uph.edu, ²03081220003@student.uph.edu, ³03081220020@student.uph.edu,

⁴03081220013@student.uph.edu

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap layanan *streaming* Netflix yang diungkapkan melalui media sosial X (sebelumnya Twitter), guna mengidentifikasi potensi *churn* atau berhentinya pelanggan berlangganan. Metode yang digunakan dalam penelitian ini adalah pendekatan *Text Mining* dan *Sentiment Analysis*, dengan dukungan model *Natural Language Processing* (NLP) berbasis *IndoBERT*. Data dikumpulkan melalui proses web scraping dengan kata kunci terkait keluhan atau keputusan berhenti berlangganan Netflix, kemudian dilakukan tahap preprocessing teks berupa *case folding*, *cleaning*, *lemmatization*, dan *tokenization*. Hasil klasifikasi sentimen menunjukkan mayoritas ulasan bersentimen negatif, yang mengindikasikan risiko *churn* cukup tinggi. Faktor utama yang memicu sentimen negatif antara lain harga langganan, kebijakan login lintas perangkat, dan pembatalan konten favorit. Hasil dari penelitian ini dapat dimanfaatkan oleh tim pemasaran dan pengembangan produk Netflix dalam menyusun strategi retensi pelanggan berbasis opini pengguna. Penelitian ini juga membuktikan bahwa model *IndoBERT* efektif dalam mengklasifikasikan opini berbahasa Indonesia dari media sosial ke dalam kategori sentimen positif, netral, dan negatif.

Kata Kunci: Sentimen, *Churn*, *IndoBERT*, Netflix, Media Sosial, Text Mining

ABSTRACT

This study aims to analyze customer sentiment toward Netflix's streaming service as expressed on social media platform X (formerly Twitter), in order to identify potential churn. The research employs a combination of Text Mining and Sentiment Analysis methods, utilizing the IndoBERT-based Natural Language Processing (NLP) model. Data was collected using web scraping techniques with keywords indicating complaints or cancellation of Netflix subscriptions. The text data underwent preprocessing steps including case folding, cleaning, lemmatization, and tokenization. Sentiment classification results showed that most tweets expressed negative sentiment, suggesting a high risk of customer churn. Key factors driving negative sentiment include subscription pricing, login policy restrictions, and the cancellation of popular content. These findings can assist Netflix's marketing and product development teams in creating data-driven retention strategies. Furthermore, the study demonstrates that the IndoBERT model is effective in classifying Indonesian-language social media opinions into positive, neutral, and negative sentiment categories.

Keywords: Sentiment, *Churn*, *IndoBERT*, Netflix, Social Media, Text Mining

Penulis Korespondensi:

Farencia Levis

Email: farencialevis@gmail.com

Article Info

Diterima: 27 April 2025

Direvisi: 30 April 2025

Disetujui: 20 Mei 2025

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Layanan *streaming* seperti Netflix telah menjadi bagian penting dalam industri hiburan digital, menawarkan berbagai konten yang dapat diakses kapan saja dan di mana saja. [1] Namun, dalam beberapa tahun terakhir, Netflix menghadapi tantangan dalam mempertahankan pelanggan akibat persaingan ketat dengan layanan *streaming* lain seperti Disney+, HBO Max, dan Amazon Prime Video. [2]

Fenomena *churn* pelanggan (berhenti berlangganan) menjadi isu yang signifikan bagi perusahaan, karena dapat berdampak langsung pada penurunan pendapatan, loyalitas pelanggan, serta pertumbuhan bisnis. [3] Beberapa faktor yang mempengaruhi *churn* pelanggan meliputi harga langganan, kualitas konten, pengalaman pengguna, serta kebijakan perusahaan seperti berbagi akun dan iklan. [4] Sejumlah studi sebelumnya telah menunjukkan bahwa faktor-faktor seperti harga langganan, kualitas konten, kemudahan penggunaan, serta kebijakan berbagi akun dan iklan dapat memengaruhi keputusan pelanggan untuk berhenti menggunakan layanan. [5] [6] Oleh karena itu, memprediksi pelanggan yang berpotensi berhenti berlangganan menjadi langkah strategis bagi Netflix untuk mengoptimalkan layanannya dan meningkatkan retensi pelanggan. [7]

Dalam penelitian ini, digunakan pendekatan NLP dengan metode *IndoBERT*, yaitu salah satu model deep learning berbasis transformer. [8] Model ini digunakan untuk melakukan analisis sentimen terhadap ulasan dan opini pelanggan terkait layanan Netflix. Hasil klasifikasi sentimen apakah positif, netral, atau negatif dapat dijadikan sebagai indikator awal untuk mendeteksi potensi *churn*.

Metode *IndoBERT* unggul dalam memahami konteks bahasa alami yang kompleks, sehingga lebih akurat dalam menilai nuansa sentimen dalam teks pelanggan. [9] [10] Penelitian ini diharapkan dapat memberikan manfaat dalam bentuk *insight* berharga tentang persepsi pelanggan terhadap layanan Netflix di media sosial, membantu perusahaan mengenali indikasi *churn* lebih awal melalui analisis opini, serta menyediakan dasar data yang dapat dimanfaatkan untuk menyusun strategi peningkatan layanan dan retensi pelanggan. Selain itu, penelitian ini juga menunjukkan efektivitas penggunaan model *IndoBERT* dalam analisis opini di media sosial X, khususnya dalam konteks isu *churn* pelanggan.

Dengan memahami pola *churn* pelanggan, Netflix dapat mengembangkan strategi pemasaran yang lebih efektif, meningkatkan pengalaman pengguna, serta mengurangi tingkat *churn*. Hasil penelitian ini diharapkan dapat memberikan wawasan berharga bagi perusahaan dalam mengelola retensi pelanggan secara lebih optimal.

2. METODE PENELITIAN

2.1. JENIS PENELITIAN

Penelitian ini merupakan penelitian kuantitatif eksperimental yang menggunakan pendekatan *data-driven*. Metode yang digunakan adalah *Text Mining* untuk menggali informasi dari teks tidak terstruktur, dan *Sentiment Analysis* untuk mengklasifikasikan opini pelanggan menjadi sentimen positif, netral, atau negatif. [11] Analisis dilakukan menggunakan Teknik *Natural Language Processing* (NLP) menggunakan model *IndoBERT* untuk klasifikasi sentimen. Fokus utama penelitian ini adalah mendeteksi opini negatif yang berpotensi menjadi indikator *churn* pelanggan. Tahapan penelitian akan meliputi:

1. Pengumpulan data dari media sosial menggunakan keyword terkait *churn* pelanggan.
2. Pra-pemrosesan data teks (*preprocessing*) untuk membersihkan dan menormalkan data.
3. Ekstraksi fitur dan tokenisasi menggunakan metode NLP.
4. Klasifikasi sentimen menggunakan model *IndoBERT*.
5. Analisis hasil untuk mengidentifikasi opini negatif sebagai indikator *churn*.

Hasil dari model dan analisis akan digunakan untuk memberikan rekomendasi strategis dalam mengurangi *churn* di layanan *streaming* seperti Netflix.

2.2. ALAT & TEKNOLOGI

1. *Google Colab* untuk mengeksekusi semua proses analisis
2. *Python libraries: pandas, nltk, spacy, transformers, matplotlib, emoji*
3. *Pretrained IndoBERT model* untuk mengklasifikasi sentimen
4. *Excel* agar menyimpan hasil *preprocessing & output* klasifikasi

2.3. TAHAPAN PENELITIAN

2.3.1. PENGUMPULAN DATA

Data pada penelitian ini dikumpulkan melalui platform media sosial X dengan menggunakan *tools* pihak ketiga (*tweet-harvest*) menggunakan *Google Colab*. Kata kunci pencarian yang digunakan disesuaikan dengan konteks *churn* pelanggan Netflix, yaitu "*netflix cancel*", "*berhenti langganan*", "*netflix not worth it*", "*unsub Netflix*", "*gak langganan Netflix lagi*"

dengan rentang waktu antara tahun 2023 hingga 2025. Dalam tahap pengumpulan data ini, dihasilkan kumpulan data mentah berupa tweet yang mengandung opini atau keluhan pelanggan terhadap layanan Netflix.

Berikut ini merupakan salah satu cuplikan kode yang digunakan dalam proses ekstraksi data sosial media untuk mengumpulkan opini dan keluhan pelanggan yang relevan terhadap *churn* layanan Netflix:

```
data = 'netflix_churn.csv'
search_keyword = 'netflix cancel OR berhenti langganan OR netflix not worth it OR unsub
netflix OR netflix mahal OR gak langganan netflix lagi since:2023-01-01 until:2025-01-01'
limit = 100

!npx --yes tweet-harvest@latest -o "{data}" -s "{search_keyword}" -l {limit} --token
"TOKEN_TWITTER"
```

Gambar 1 Kode Pengumpulan Data di Google Colab

2.3.2.EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) adalah proses awal dalam analisis data yang bertujuan untuk memahami struktur, pola, dan karakteristik data sebelum dilakukan pemodelan lebih lanjut. [12] Pada tahap awal EDA, dilakukan proses pemuatan data dari file “*netflix_churn.csv*”. Dataset ini berisi kumpulan teks atau opini pelanggan yang diperoleh dari media sosial X terkait layanan Netflix. Dari keseluruhan kolom yang tersedia, hanya kolom *full_text* yang dipilih untuk dianalisis lebih lanjut karena berisi isi ulasan utama yang akan digunakan dalam proses klasifikasi dan prediksi *churn*. Berikut adalah tautan [Google Colab](#) untuk pre-processing dan klasifikasi dengan metode *IndoBERT* terhadap opini pelanggan terhadap layanan Netflix.

2.3.3.PRE-PROCESSING

1. Case Folding, Data Cleaning, Truncating

Pada tahap ini, dilakukan *preprocessing* awal terhadap data ulasan pelanggan yang diperoleh dari platform X untuk mendukung analisis prediksi *churn* pelanggan layanan Netflix. Langkah-langkah *preprocessing* yang diterapkan adalah sebagai berikut: [13]

a. Case Folding

Case folding adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*). Tujuannya adalah untuk mengurangi keragaman kata akibat penggunaan huruf kapital. Semua huruf diubah menjadi huruf kecil menggunakan fungsi *lower()* agar kata-kata seperti “Netflix” dan “netflix” dianggap sama oleh sistem.

b. Data Cleaning

Menghapus elemen-elemen yang tidak relevan untuk analisis teks, antara lain:

- Baris baru (*n*)
- Emoji dan emotikon
- *Link* (tautan)
- *Username* (@...)
- *Hashtag* (#...)
- Angka dan simbol

Proses ini dilakukan dengan bantuan *library* dan beberapa ekspresi reguler (*regex*).

c. Truncating (Pemotongan Teks)

Untuk menjaga efisiensi pemrosesan dan konsistensi panjang input teks, dilakukan pemotongan terhadap ulasan hingga 50 karakter pertama. Semua hasil *preprocessing* disimpan dalam struktur *dictionary* bernama *churn_preprocess*, dengan kolom-kolom:

- *original* : isi tweet asli
- *lower* : hasil case folding
- *clean* : hasil pembersihan teks
- *truncate* : hasil pemotongan

2. Lemmatization

Lemmatization adalah proses dalam pemrosesan bahasa alami (NLP) yang bertujuan untuk mengubah setiap kata dalam teks menjadi bentuk dasarnya (*lemma*), tanpa mengubah makna aslinya. [14] Pada tahap ini, dilakukan proses *lemmatization* menggunakan *library spaCy* dengan model bahasa Inggris *en_core_web_sm*. *Lemmatization* bertujuan untuk mengubah setiap kata dalam teks ulasan pelanggan menjadi bentuk dasarnya (*lemma*), seperti “*running*” menjadi “*run*”. Hal ini dilakukan untuk menyederhanakan bentuk kata, sehingga proses analisis menjadi lebih efisien dan akurat.

```
import pandas as pd
import spacy

# Load model bahasa Inggris spaCy
nlp = spacy.load("en_core_web_sm")

# Load data
df = pd.read_excel("Preprocessing_Netflix.xlsx")

# Fungsi untuk lemmatization
def lemmatized_text(text):
    doc = nlp(text)
    return ' '.join([token.lemma_ for token in doc if not token.is_punct and not
token.is_space])

# Terapkan lemmatization ke kolom 'clean'
df['lemmatized'] = df['clean'].astype(str).map(lemmatized_text)

# Tampilkan hasil
print(df[['clean', 'lemmatized']].head())

# Simpan ke Excel (hasil akan disimpan sebagai file yang bisa diunduh)
df.to_excel("Preprocessing_Lemmatized_Netflix.xlsx", index=False)
```

Gambar 2 Kode Lemmatization di Google Colab

3. Tokenization

Tokenization adalah proses memecah teks menjadi satuan terkecil, yaitu token atau kata. [15] Dalam tahapan ini digunakan library NLTK (*Natural Language Toolkit*) dengan fungsi *word_tokenize*. *Input* untuk proses ini berasal dari hasil *lemmatization* sebelumnya (kolom *lemmatized*). Hasil dari tokenisasi disimpan dalam bentuk list kata di kolom '*tokenized*', lalu dikonversi menjadi *string* (dipisah dengan spasi) dan disimpan ke dalam kolom '*tokenized_str*'.

```
import pandas as pd
import numpy as np
from nltk.tokenize import word_tokenize
import nltk
nltk.download('punkt') # pastikan ini sudah diunduh

# Load hasil preprocessing + stemming
df = pd.read_excel("Preprocessing_Lemmatized_Netflix.xlsx")

# Gunakan kolom 'lemmatized' sebagai input tokenization
def tokenize_text(text):
    return word_tokenize(text)

# Terapkan tokenisasi (hasil jadi list token, bukan string lagi)
df['tokenized'] = df['lemmatized'].astype(str).map(tokenize_text)

# Bersihkan baris kosong (kalau ada)
df['tokenized'].replace('', np.nan, inplace=True)
df['tokenized'].replace(' ', np.nan, inplace=True)
```

```
df.dropna(subset=['tokenized'], inplace=True)

# Cek hasil
print(df[['lemmatized', 'tokenized']].head())

# Simpan hasil final token ke Excel
df['tokenized_str'] = df['tokenized'].apply(lambda tokens: ' '.join(tokens))
df.to_excel("Preprocessing Tokenized Netflix.xlsx", index=False)
```

Gambar 3 Kode Tokenization di Google Colab

2.3.4. PENGELOMPOKKAN DATA

Pada tahap ini dilakukan analisis sentimen menggunakan model *BERT Multilingual* yang disediakan oleh *Hugging Face*, yaitu model *nlptown/bert-base-multilingual-uncased-sentiment*. Model ini mampu mengklasifikasikan sentimen berdasarkan skala bintang dari 1 hingga 5. Setiap prediksi yang diberikan oleh model berupa label "*X stars*", kemudian dikonversi ke bentuk numerik. Setelah itu, dilakukan pemetaan untuk mengelompokkan skor bintang menjadi tiga kategori sentimen:

- 1–2 bintang: *Negative*
- 3 bintang: *Neutral*
- 4–5 bintang: *Positive*

Tujuan dari proses ini adalah untuk mengetahui persepsi atau kepuasan pelanggan berdasarkan ulasan mereka, yang kemudian bisa digunakan untuk menganalisis kemungkinan pelanggan melakukan *churn* (berhenti berlangganan).

2.3.5. VISUALISASI ANALISIS SENTIMEN

Setelah itu dilakukan analisis distribusi sentimen terhadap ulasan pelanggan dengan menggunakan visualisasi berupa grafik batang, kami dapat melihat proporsi ulasan yang positif, netral, dan negatif. Hasil dari analisis distribusi ini akan memberikan gambaran tentang apakah pelanggan umumnya merasa puas atau tidak dengan layanan yang disediakan.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap layanan *streaming* Netflix yang diekspresikan melalui media sosial X. Metode yang digunakan menggabungkan pendekatan *IndoBERT* untuk analisis sentimen. Langkah-langkah yang dilakukan meliputi:

3.1. PENGUMPULAN DATA

Hasil dari proses crawling ini menghasilkan 78 *tweets* yang mengandung potensi indikator *churn*, dan telah disimpan dalam tautan [crawl_netflix_churn.csv](#) untuk dianalisis lebih lanjut pada tahap transformasi dan klasifikasi data.

Tabel 1 Data Crawl Netflix

conversation id str	created at	favorite count	full text	id str	lang
1.71E+18	Tue Oct 03 17:00:34 +0000 2023	0	Why I cut and bailed. I basically will only renew when the last season of Stranger Things comes out and then cancel right after. Netflix is currently \$15.99 and you can bet it's going right up to \$19.99. Not worth it.	1.71E+18	en
1.62E+18	Thu Feb 02 01:48:49 +0000 2023	1084	Well folks its time to cancel your subscriptions and let Netflix suffer. Absolutely baffling they trying this and keep increasing their subscription prices. https://t.co/yGK7WhTiuX	1.62E+18	Australia
1.84E+18	Wed Oct 09 10:43:32 +0000 2024	5	I had to cancel my Netflix to budget and such - but when I canceled I did put one of the reasons as other: you guys keep canceling my favorite shows and it's not worth putting time and effort into using your app Cuz it's true - tired of getting invested and then they are gone	1.84E+18	en
1.66E+18	Sat May 13 05:04:43 +0000 2023	13	Looks like I was right. Netflix have decided to cancel #LockwoodandCo after one season. They've been doing this far too much recently & I'm starting to wonder if it's worth cancelling my subscription! https://t.co/nuWkrG9TUe	1.66E+18	Somewhere
1.62E+18	Wed Feb 01 16:22:58 +0000 2023	55	Netflix is rly about to learn how disposable it is. If you prevent me from signing in across multiple households I will simply cancel. You're not worth buying more subscriptions for.	1.62E+18	en

Tujuan dari pengumpulan data ini untuk mengetahui opini dan keluhan pengguna terkait layanan Netflix yang dapat mengindikasikan potensi *churn* pelanggan.

3.2. PRE-PROCESING TEKS

Melakukan *case folding*, *cleaning*, *truncating*, *lemmatization*, dan tokenisasi untuk mempersiapkan teks yang bersih dan siap dianalisis. *Preprocessing* bertujuan untuk membersihkan dan menormalkan teks agar lebih siap digunakan untuk tahapan analisis.

1. Case Folding, Data Cleaning, Truncate

Seluruh teks diubah menjadi huruf kecil (*lowercase*) agar tidak terjadi perbedaan interpretasi antara kata yang sama tetapi dengan format kapitalisasi berbeda, seperti “Netflix” dan “netflix”. Teks dibersihkan dari berbagai elemen yang tidak relevan dengan analisis sentiment. Teks yang telah dibersihkan kemudian dipotong (*truncate*) hingga maksimal 50 karakter pertama.

Tabel 2 Hasil Case Folding, Data Cleaning, dan Truncate

original	Lower	clean	truncate
Why I cut and bailed. I basically will only renew when the last season of Stranger Things comes out and then cancel right after. Netflix is currently \$15.99 and you can bet it s going right up to \$19.99. Not worth it.	why i cut and bailed. i basically will only renew when the last season of stranger things comes out and then cancel right after. netflix is currently \$15.99 and you can bet it s going right up to \$19.99. not worth it.	why i cut and bailed i basically will only renew when the last season of stranger things comes out and then cancel right after netflix is currently and you can bet it s going right up to not worth it	why i cut and bailed i basically will only renew w
Well folks its time to cancel your subscriptions and let Netflix suffer. Absolutely baffling they trying this and keep increasing their subscription prices. https://t.co/yGK7WhTiuX	well folks its time to cancel your subscriptions and let netflix suffer. absolutely baffling they trying this and keep increasing their subscription prices. https://t.co/ygk7whtiux	well folks its time to cancel your subscriptions and let netflix suffer absolutely baffling they trying this and keep increasing their subscription prices	well folks its time to cancel your subscriptions a
I had to cancel my Netflix to budget and such - but when I canceled I did put one of the reasons as other: you guys keep canceling my favorite shows and it's not worth putting time and effort into using your app Cuz it's true - tired of getting invested and then they are gone	i had to cancel my netflix to budget and such - but when i canceled i did put one of the reasons as other: you guys keep canceling my favorite shows and it's not worth putting time and effort into using your app cuz it's true - tired of getting invested and then they are gone	i had to cancel my netflix to budget and such but when i canceled i did put one of the reasons as other you guys keep canceling my favorite shows and it s not worth putting time and effort into using your app cuz it s true tired of getting invested and then they are gone	i had to cancel my netflix to budget and such but
Looks like I was right. Netflix have decided to cancel #LockwoodandCo after one season. They've been doing this far too much recently & i'm starting to wonder if it's worth cancelling my subscription! https://t.co/nuWkrG9TUE	looks like i was right. netflix have decided to cancel #lockwoodandco after one season. they've been doing this far too much recently & i'm starting to wonder if it's worth cancelling my subscription! https://t.co/nuwkrG9tue	looks like i was right netflix have decided to cancel lockwoodandco after one season they ve been doing this far too much recently amp i m starting to wonder if it s worth cancelling my subscription	looks like i was right netflix have decided to can
Netflix is rllly about to learn how disposable it is. If you prevent me from signing in across multiple households I will simply cancel. You're not worth buying more subscriptions for.	netflix is rllly about to learn how disposable it is. if you prevent me from signing in across multiple households i will simply cancel. you're not worth buying more subscriptions for.	netflix is rllly about to learn how disposable it is if you prevent me from signing in across multiple households i will simply cancel you re not worth buying more subscriptions for	netflix is rllly about to learn how disposable it i

Hasil preprocessing dikonversi ke dalam bentuk *DataFrame* dan disimpan ke file Excel [Preprocessing CDT Netflix.xlsx](#).

2. Lemmetization

Lemmatization adalah proses mengubah setiap kata ke dalam bentuk dasar (*lemma*) tanpa mengubah makna dasarnya. Misalnya, kata “*watching*”, “*watches*”, dan “*watched*” semuanya akan diubah menjadi bentuk dasarnya “*watch*”. Data yang digunakan berasal dari file *Preprocessing_Netflix.xlsx*, dan hasil *lemmatization* disimpan ke dalam kolom baru bernama *lemmatized*. File hasil *lemmatization* ini disimpan ke file baru [Preprocessing Lemmatized Netflix.xlsx](#).

Tabel 3 Hasil Lemmetization

truncate	lemmetized
why i cut and bailed i basically will only renew w	why I cut and bail I basically will only renew when the last season of stranger thing come out and then cancel right after netflix be currently and you can bet it s go right up to not worth it
well folks its time to cancel your subscriptions a	well folk its time to cancel your subscription and let netflix suffer absolutely baffle they try this and keep increase their subscription price
i had to cancel my netflix to budget and such but	I have to cancel my netflix to budget and such but when I cancel I do put one of the reason as other you guy keep cancel my favorite show and it s not worth put time and effort into use your app cuz it s true tired of getting invest and then they be go
looks like i was right netflix have decided to can	look like I be right netflix have decide to cancel lockwoodandco after one season they ve be do this

<i>netflix is rlly about to learn how disposable it i</i>	<i>far too much recently amp I m start to wonder if it s worth cancel my subscription netflix be rlly about to learn how disposable it be if you prevent I from sign in across multiple household I will simply cancel you re not worth buy more subscription for</i>
---	---

3. Tokenization

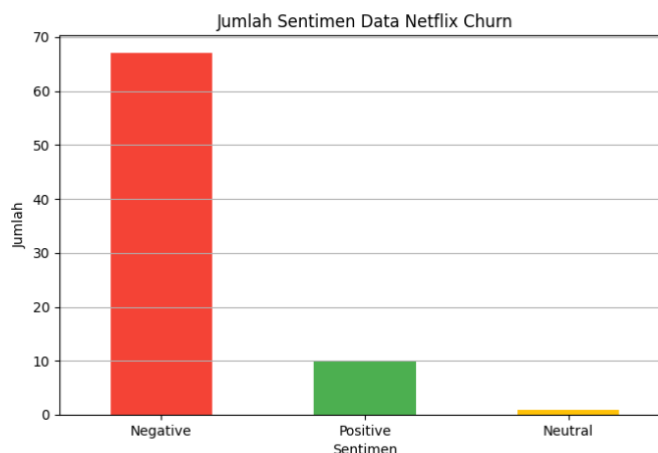
Tokenization adalah proses memecah kalimat atau teks menjadi unit terkecil yang disebut token, biasanya berupa kata. Tujuan dari tahap ini adalah untuk mempersiapkan data teks agar bisa digunakan pada model machine learning, khususnya untuk tahap analisis sentimen. Hasil dari *tokenization* bisa dilihat di file Excel [Preprocessing Tokenized Netflix.xlsx](#).

Tabel 4 Hasil Tokenization

<i>truncate</i>	<i>lemmetized</i>	<i>tokenized</i>	<i>tokenized str</i>
<i>why i cut and bailed i basically will only renew w</i>	<i>why I cut and bail I basically will only renew when the last season of stranger thing come out and then cancel right after netflix be currently and you can bet it s go right up to not worth it</i>	<i>['why', 'I', 'cut', 'and', 'bail', 'I', 'basically', 'will', 'only', 'renew', 'when', 'the', 'last', 'season', 'of', 'stranger', 'thing', 'come', 'out', 'and', 'then', 'cancel', 'right', 'after', 'netflix', 'be', 'currently', 'and', 'you', 'can', 'bet', 'it', 's', 'go', 'right', 'up', 'to', 'not', 'worth', 'it']</i>	<i>why I cut and bail I basically will only renew when the last season of stranger thing come out and then cancel right after netflix be currently and you can bet it s go right up to not worth it</i>
<i>well folks its time to cancel your subscriptions a</i>	<i>well folk its time to cancel your subscription and let netflix suffer absolutely baffle they try this and keep increase their subscription price</i>	<i>['well', 'folk', 'its', 'time', 'to', 'cancel', 'your', 'subscription', 'and', 'let', 'netflix', 'suffer', 'absolutely', 'baffle', 'they', 'try', 'this', 'and', 'keep', 'increase', 'their', 'subscription', 'price']</i>	<i>well folk its time to cancel your subscription and let netflix suffer absolutely baffle they try this and keep increase their subscription price</i>
<i>i had to cancel my netflix to budget and such but</i>	<i>I have to cancel my netflix to budget and such but when I cancel I do put one of the reason as other you guy keep cancel my favorite show and it s not worth put time and effort into use your app cuz it s true tired of getting invest and then they be go</i>	<i>['I', 'have', 'to', 'cancel', 'my', 'netflix', 'to', 'budget', 'and', 'such', 'but', 'when', 'I', 'cancel', 'I', 'do', 'put', 'one', 'of', 'the', 'reason', 'as', 'other', 'you', 'guy', 'keep', 'cancel', 'my', 'favorite', 'show', 'and', 'it', 's', 'not', 'worth', 'put', 'time', 'and', 'effort', 'into', 'use', 'your', 'app', 'cuz', 'it', 's', 'true', 'tired', 'of', 'getting', 'invest', 'and', 'then', 'they', 'be', 'go']</i>	<i>I have to cancel my netflix to budget and such but when I cancel I do put one of the reason as other you guy keep cancel my favorite show and it s not worth put time and effort into use your app cuz it s true tired of getting invest and then they be go</i>
<i>looks like i was right netflix have decided to can</i>	<i>look like I be right netflix have decide to cancel lockwoodandco after one season they ve be do this far too much recently amp I m start to wonder if it s worth cancel my subscription</i>	<i>['look', 'like', 'I', 'be', 'right', 'netflix', 'have', 'decide', 'to', 'cancel', 'lockwoodandco', 'after', 'one', 'season', 'they', 've', 'be', 'do', 'this', 'far', 'too', 'much', 'recently', 'amp', 'I', 'm', 'start', 'to', 'wonder', 'if', 'it', 's', 'worth', 'cancel', 'my', 'subscription']</i>	<i>look like I be right netflix have decide to cancel lockwoodandco after one season they ve be do this far too much recently amp I m start to wonder if it s worth cancel my subscription</i>
<i>netflix is rlly about to learn how disposable it i</i>	<i>netflix be rlly about to learn how disposable it be if you prevent I from sign in across multiple household I will simply cancel you re not worth buy more subscription for</i>	<i>['netflix', 'be', 'rlly', 'about', 'to', 'learn', 'how', 'disposable', 'it', 'be', 'if', 'you', 'prevent', 'I', 'from', 'sign', 'in', 'across', 'multiple', 'household', 'I', 'will', 'simply', 'cancel', 'you', 're', 'not', 'worth', 'buy', 'more', 'subscription', 'for']</i>	<i>netflix be rlly about to learn how disposable it be if you prevent I from sign in across multiple household I will simply cancel you re not worth buy more subscription for</i>

3.3. KLASIFIKASI SENTIMEN DENGAN INDOBERT

Model *IndoBERT* berhasil mengklasifikasikan tweet yang telah diproses ke dalam tiga kategori sentimen. Pemrosesan ini dilakukan sepenuhnya di *Google Colab* menggunakan pipeline dari *library Transformers*. Hasil dari tahapnya bisa dilihat dari file Excel berikut [Hasil_3Label Sentimen NetflixChurn.xlsx](#).



Gambar 4 Jumlah Sentimen Data Netflix Churn

3.4. ANALISIS DAN INTERPRETASI SENTIMEN CHURN

Setelah dilakukan *preprocessing* dan klasifikasi sentimen menggunakan model *IndoBERT*, dilakukan analisis lebih lanjut terhadap hasil klasifikasi untuk menjawab rumusan masalah dan tujuan penelitian.

3.4.1.DISTRIBUSI SENTIMEN PELANGGAN

Berdasarkan *output* dari klasifikasi sentimen, data dibagi ke dalam tiga kategori:

- Positif: mencerminkan kepuasan pelanggan
- Netral: tidak menunjukkan emosi kuat terhadap layanan
- Negatif: menunjukkan ketidakpuasan dan menjadi indikator potensi *churn*

Tabel 5 Persentase Sentimen

Sentimen	Jumlah Tweet	Persentase
Positif	10	12.8%
Netral	1	1.3%
Negatif	67	85.9%

Dari hasil persentase klasifikasi sentimen model *IndoBERT* bisa diasumsikan, mayoritas pelanggan mengungkapkan ketidakpuasan terhadap layanan Netflix, tingginya persentase tweet negatif menandakan bahwa tingkat risiko *churn* sangat tinggi dalam periode waktu pengamatan. Hasil ini memperkuat kebutuhan untuk mengevaluasi harga layanan, konten, dan kebijakan platform. Distribusi ini menjadi indikator kuat bagi perusahaan untuk segera mengambil tindakan strategis, baik dari sisi pemasaran, pelayanan pelanggan, maupun pengembangan produk.

3.4.2.INDIKATOR CHURN BERDASARKAN ANALISIS SENTIMEN

Dari hasil klasifikasi, dilakukan pemetaan kata kunci dominan yang sering muncul pada *tweet* dengan sentimen negatif. Beberapa tema utama yang muncul antara lain:

- Harga mahal: “Netflix mahal”, “*not worth*”, “*price increase*”
- Konten dikurangi atau dibatalkan: “*cancel show*”, “*favourite show gone*”
- Kebijakan login: “*multiple households*”, “*sharing password*”
- Kompetitor lebih menarik: “Disney”, “Prime”, “*cheaper option*”

Keluhan ini mengarah pada alasan konkret pelanggan berhenti berlangganan. Dengan mengidentifikasi pola ini, perusahaan bisa mengambil langkah pencegahan yang lebih tepat.

3.4.3.VALIDASI MANFAAT INDOBERT

Penggunaan *IndoBERT* terbukti efisien dalam:

- Mengklasifikasikan sentimen secara akurat.
- Menangkap konteks *churn* dari opini pendek dan informal seperti *tweet*.
- Mengurangi ambiguitas teks media sosial, berkat *pretraining* yang luas.

Model ini mendukung tujuan penelitian ke-2 dan ke-3 secara signifikan, dan memberikan dasar analitik yang kuat dalam menyusun strategi retensi pelanggan.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap layanan *streaming* Netflix yang diekspresikan melalui media sosial X (sebelumnya Twitter), serta mengidentifikasi opini negatif yang berpotensi menjadi indikator *churn*. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa tujuan penelitian sebagaimana dijabarkan dalam Bab Pendahuluan telah tercapai, dengan didukung oleh temuan dan analisis pada Bab Hasil dan Pembahasan.

Kesimpulan utama dari penelitian ini adalah sebagian besar tweet pelanggan menunjukkan sentimen negatif, yang menjadi indikator kuat adanya potensi *churn* jika tidak segera ditindaklanjuti. Sentimen negatif ini umumnya berkaitan dengan harga layanan yang meningkat, konten yang tidak memuaskan, dan kebijakan login lintas perangkat yang dianggap membatasi kenyamanan pelanggan. Metode *IndoBERT* terbukti efektif dalam memetakan opini pelanggan ke dalam tiga kategori sentimen utama: positif, netral, dan negatif. Analisis sentimen berhasil mengidentifikasi faktor-faktor utama penyebab *churn*, yang dapat dijadikan dasar untuk pengambilan keputusan bisnis secara strategis oleh pihak Netflix.

Temuan ini dapat dimanfaatkan oleh tim bisnis Netflix untuk mengembangkan strategi promosi ulang (re-marketing) kepada pelanggan yang menunjukkan potensi *churn*. Hasil klasifikasi sentimen juga dapat digunakan untuk menyesuaikan kebijakan dan fitur layanan, berdasarkan sentimen pelanggan terhadap aspek-aspek tertentu. Selain itu, perusahaan dapat meningkatkan *customer experience* melalui perbaikan yang bersifat responsif terhadap opini dan keluhan nyata dari pengguna.

Saran untuk penelitian selanjutnya adalah penggunaan dataset dengan volume yang lebih besar dan lebih bervariasi secara temporal agar hasilnya lebih general dan menggabungkan analisis sentimen dengan analisis prediktif lainnya seperti *clustering churn behavior* atau *predictive modeling* berbasis atribut pengguna.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan laporan penelitian yang berjudul "Analisis Sentimen Churn Pelanggan dalam Layanan Streaming Netflix di X Menggunakan Metode IndoBERT" dengan baik. Tim ini juga ingin menyampaikan terima kasih yang sebesar-besarnya kepada dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan yang sangat berharga selama proses penelitian ini. Laporan ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan laporan ini di masa yang akan datang.

REFERENSI

- [1] Y. Anjani, M. Wicaksana, and A. Kuswanti, "Penggunaan aplikasi streaming Netflix pada generasi Z," *Ikon--Jurnal Ilmiah Ilmu Komunikasi*, vol. 28, no. 1, pp. 88–96, 2023.
- [2] P. S. Sitanggang, "Strategi pemasaran global terhadap Netflix," *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, vol. 1, no. 9, pp. 3026–3035, 2022.
- [3] K. S. R. Palluvi, N. Syaada, and B. Intan, "Komparasi Metode Decision Tree dan K-Nearest Neighbor (KNN) dalam Memprediksi Costumer Churn Pada Perusahaan Telekomunikasi," *Bulletin of Information System Research*, vol. 3, no. 1, pp. 39–45, 2024.
- [4] S. T. Sjukun and M. M. SM, *Pemasaran Di Era Digital*. CV. AZKA PUSTAKA, 2024.
- [5] P. Simaniburuk *et al.*, *MEMAHAMI PERILAKU KONSUMEN: Strategi Pemasaran yang Efektif pada Era Digital*. PT. Sonpedia Publishing Indonesia, 2023.
- [6] S. E. Yoyo Sudaryo, M. MM, S. P. Nunung Ayu Sofiaty Efi, S. E. Mohamad Arfiman Yosep, S. T. Budi Nurdiansyah, and M. H. SE, *Digital Marketing dan fintech di Indonesia*. Penerbit Andi, 2020.
- [7] R. S. Y. Zebua *et al.*, *BISNIS DIGITAL: Strategi Administrasi Bisnis Digital Untuk Menghadapi Masa Depan*. PT. Sonpedia Publishing Indonesia, 2023.
- [8] D. Sebastian, H. D. Purnomo, and I. Sembiring, "Bert for natural language processing in bahasa Indonesia," in *2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, IEEE, 2022, pp. 204–209.
- [9] R. Merdiansah, S. Siska, and A. A. Ridha, "Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 221–228, 2024.
- [10] M. P. Firdaus and D. Trisnawarman, "Analisis Sentimen Publik terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube: Public Sentiment Analysis of the Public Housing Savings Program Using the IndoBERT Lite Model on YouTube Comments," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 359–368, 2025.
- [11] A. Onan, "Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 572–589, 2021.
- [12] F. Elfaladonna, I. G. T. Isa, D. Sartika, and A. M. Putra, *Buku Ajar Dasar Exploratory Data Analysis (EDA)*. Penerbit NEM, 2024.
- [13] M. Kunilovskaya and A. Plum, "Text preprocessing and its implications in a digital humanities project," in *Proceedings of the Student Research Workshop Associated with RANLP 2021*, 2021, pp. 85–93.
- [14] R. Pramana, J. J. Subroto, and A. A. S. Gunawan, "Systematic literature review of stemming and lemmatization performance for sentence similarity," in *2022 IEEE 7th international conference on information technology and digital applications (ICITDA)*, IEEE, 2022, pp. 1–6.
- [15] S. U. Royan, N. Suarna, I. Ali, and D. Solihudin, "ANALISIS SENTIMEN ULASAN PRODUK SKINCARE DI SHOPEE UNTUK MENINGKATKAN KUALITAS PRODUK MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *Jurnal Informasi dan Komputer*, vol. 13, no. 01, pp. 96–105, 2025.