



## Sentimen Analisis Review Aplikasi Cek Bansos Pada Google Play Store Menggunakan Metode *Naïve Bayes*

Ary Ardiansyah<sup>1</sup>, Pareza Alam Jusia<sup>2</sup>, Rudolf Sinaga<sup>3</sup>, Clarisa Putri Valentina<sup>4</sup>, Nadia Pardede<sup>5</sup>

<sup>1-5</sup> Sistem Informasi, Ilmu Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: [ryuzen190301@gmail.com](mailto:ryuzen190301@gmail.com)<sup>1</sup>, [parezaalam@gmail.com](mailto:parezaalam@gmail.com)<sup>2</sup>, [rudolfverdinan@gmail.com](mailto:rudolfverdinan@gmail.com)<sup>3</sup>,

[Cputrivalentina@gmail.com](mailto:Cputrivalentina@gmail.com)<sup>4</sup>, [Nadiapardede282@gmail.com](mailto:Nadiapardede282@gmail.com)<sup>5</sup>

Alamat: Jalan Jendral Sudirman, Thehok, Jambi 36138.

\* Penulis Korespondensi: [ryuzen190301@gmail.com](mailto:ryuzen190301@gmail.com)

**Abstract.** *The Ministry of Social Affairs has made a new breakthrough in facilitating the public in checking social assistance recipients, namely the social assistance check application. User reviews can be used to find out whether the application provides benefits to the community or not. However, these reviews need to be processed using sentiment analysis. Then to do sentiment analysis requires machine learning. One method that includes machine learning is Naïve Bayes. The purpose of this research is to implement the Naïve Bayes method in conducting sentiment analysis and find out whether the social assistance check application is beneficial to society based on the results of sentiment analysis. In this study, two categories of sentiment are used, namely positive and negative. The author collects by crawling using the Google Play Scraper library. The results of crawling data obtained as many as 4000 data. The results showed that the actual data that had been labeled using Textblob resulted in 987 negative label reviews and 628 positive label reviews. Meanwhile, the Naïve Bayes method is able to analyze the review sentiment of the social assistance check application with the results of 1181 negative sentiments and 434 positive sentiments. The Naïve Bayes model has a good accuracy rate of 0.77 or 77% in analyzing sentiment for social assistance check application reviews.*

**Keywords:** *analysis sentiment; machine learning; social assistance check; google play; naïve bayes.*

**Abstrak**–Kementerian sosial membuat penerobosan baru dalam memudahkan masyarakat dalam melakukan pengecekan penerima bantuan sosial yaitu pada aplikasi cek bansos. Ulasan – ulasan pengguna dapat digunakan untuk mengetahui apakah aplikasi memberikan manfaat untuk masyarakat atau tidak. Namun ulasan tersebut perlu diolah menggunakan analisis sentimen. Lalu untuk melakukan analisis sentimen membutuhkan *Machine learning*. Salah satu metode yang termasuk machine learning adalah *Naïve Bayes*. Tujuan penelitian ini adalah untuk mengimplementasikan metode *Naïve Bayes* dalam melakukan sentimen analisis dan mengetahui apakah aplikasi cek bansos bermanfaat untuk masyarakat berdasarkan hasil analisis sentimen. Pada penelitian ini menggunakan dua kategori sentimen yakni positif dan negatif. Penulis melakukan pengumpulan dengan cara *crawling* menggunakan *library google play scrapper*. Hasil dari *crawling data* didapatkan sebanyak 4000 data. Hasil penelitian data aktual yang telah dilakukan analisis sentimen menggunakan metode *Naïve Bayes* mampu menganalisis sentimen review aplikasi cek bansos dengan hasil hasil 1214 positif dan 2478 negatif dengan akurasi yang baik sebesar 0.77 atau 77% . pada aplikasi cek bansos.

**Kata Kunci:** *sentimen analisis; machine learning; cek bansos; google play; naïve bayes*

### 1. LATAR BELAKANG

Pandemi Covid-19 awal muncul pada tahun 2019 dan menyebar ke seluruh negara termasuk Indonesia. Semenjak virus tersebut diberitakan masuk di Indonesia pada awal bulan Maret 2020 menyerang banyak korban hingga menyebabkan efek domino pada semua sektor, terutama pada sektor ekonomi (Noerkaisar, 2021). Pada masa Covid-19 berdampak pada interaksi sosial masyarakat. Seluruh aktivitas ekonomi terhambat mulai dari tidak bisa dilakukannya pekerjaan harian, gelombang PHK (Pemutusan Hubungan Kerja) yang meningkat yang dilakukan oleh perusahaan karena tidak bisa mempertahankan karyawan pada

kondisi pandemi. Selain itu adanya penurunan penghasilan (Rahmansyah & I., 2020). Pemerintah membuat solusi untuk menghadapi situasi tersebut dengan memberikan bantuan sosial kepada masyarakat. Bantuan sosial (Bansos) berasal dari Anggaran Pendapatan dan Belanja Negara (APBN).

Pengeluaran bantuan sosial yang diselenggarakan pemerintah pusat atau daerah berupa transfer uang, barang, atau jasa untuk melindungi masyarakat dari kemungkinan terjadinya risiko sosial serta meningkatkan kemampuan ekonomi (Pratama & S., 2022). Pemerintah daerah tidak memberikan bantuan sosial kepada seluruh masyarakat, namun Pemerintah Daerah hanya diperbolehkan memberi bantuan sosial kepada penduduk sesuai dengan kemampuan keuangan pemerintahan daerah. Penduduk yang dapat menerima bantuan sosial adalah penduduk yang mengalami keadaan yang tidak stabil sebagai akibat dari pandemi (Latif & P., 2022). Kementerian sosial juga membuat penerobosan baru dalam memudahkan masyarakat dalam melakukan pengecekan penerima bantuan sosial. Pada aplikasi cek bansos juga memiliki menu usul sanggah yang memudahkan masyarakat mengusulkan nama yang berhak dan tidak berhak mendapatkan bantuan sosial (Latif & P., 2022). Ulasan – ulasan pengguna dapat digunakan untuk mengetahui apakah aplikasi memberikan manfaat untuk masyarakat atau tidak. Namun ulasan tersebut perlu diolah menggunakan analisis sentimen. Dengan melakukan analisis sentimen, dapat mengetahui apakah aplikasi cek bansos memiliki sentimen positif atau negatif.

Sentimen analisis (*sentiment analysis*) adalah teknik pemrosesan bahasa alami (*Natural Language Processing*) yang bertujuan untuk mengidentifikasi dan mengekstrak sentimen atau opini yang terkandung dalam teks, baik itu positif, negatif, atau netral. Tujuan dari sentimen analisis adalah untuk memahami dan menganalisis bagaimana pengguna atau konsumen merespons suatu produk, layanan, atau topik tertentu, sehingga dapat membantu perusahaan atau organisasi dalam pengambilan keputusan bisnis.

Analisis sentimen dapat melakukan penyelidikan *review* suatu produk di internet untuk menentukan opini atau perasaan terhadap produk secara keseluruhan. Informasi yang diperoleh dalam bentuk teks yang banyak tersedia di internet dalam beberapa format untuk forum, media sosial, dan situs yang berisi ulasan. Dengan analisis sentimen, kita dapat mengolah data ini menjadi data terstruktur (Oktaviani & S., 2020). Analisis sentimen memiliki tiga tingkatan yaitu tingkat dokumen, kalimat dan aspek. Pada penelitian ini akan menggunakan tingkat kalimat. Selain itu analisis sentimen dapat menentukan ulasan yang dikeluarkan termasuk dalam kategori sentimen positif atau negatif (Surohman & W., 2020).

Lalu untuk melakukan analisis sentimen membutuhkan *Machine learning*. *Machine learning* mempelajari cara menjalankan tugas atau memecahkan masalah secara otomatis yang merupakan pengaplikasian dari *artificial intelligence* (Mansour & A., 2022). Salah satu metode yang termasuk machine learning adalah *Naïve Bayes*. *Naïve Bayes* adalah metode klasifikasi untuk text mining yang digunakan dalam analisis sentimen. Secara teoritis baik dalam hal konsistensi data dan klasifikasi perhitungan. *Naïve Bayes* umumnya digunakan dalam teknik klasifikasi, khususnya analisis sentimen pada Twitter, dengan menggunakan berbagai metode seperti *Unigram Naïve Bayes*, *Multinomial Nave Bayes*, dan *Maximum Entropy Classification*. Fitur utama dari klasifikasi *Naïve Bayes* adalah untuk mendapatkan hipotesis yang kuat dari setiap kondisi (Pristiyono & R., 2020).

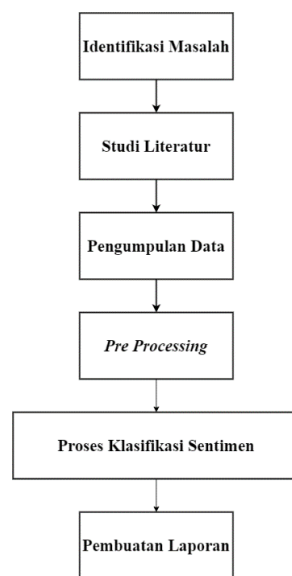
Beberapa penelitian terkait analisis sentimen diantaranya adalah penelitian yang dilakukan oleh Ramdhani, et al terkait analisis review produk sebelum dan sesudah melakukan *endorsement* menggunakan metode *Naïve Bayes*. Berdasarkan hasil penelitian, metode *Naïve Bayes* menghasilkan akurasi lebih dari 80%. Akurasi yang dihasilkan untuk analisis sentimen sebelum *endorsement* adalah 80.37% pada data training dan 85.84% pada data *testing*. Sedangkan akurasi yang dihasilkan analisis sentimen setelah *endorsement* adalah 86.03% pada data training dan 86.15% pada data *testing* (Ramdhani & H., 2018). Kemudian penelitian yang dilakukan Surohman, et al., terkait analisis sentimen pada ulasan *Fintech* menggunakan metode *Naïve Bayes* dan KNN. Berdasarkan hasil penelitian, metode *Naïve Bayes* lebih unggul dari pada metode KNN dalam melakukan analisis sentimen. Nilai akurasi yang dihasilkan metode *Naïve Bayes* mencapai 84,85%, sedangkan metode KNN menghasilkan nilai akurasi 82,96% (Surohman & W., 2020).

Metode *Naïve Bayes* dipilih untuk penelitian tugas akhir ini. Metode *Naïve Bayes* dapat mengkategorikan data berdasarkan perhitungan probabilitas masing-masing kelas terhadap item - item dalam sampel data. *Naïve Bayes* memperkirakan probabilitas kelas berdasarkan kualitasnya dan memilih kelas dengan probabilitas tertinggi (Samsir & W., 2022). Selain itu, *Naïve Bayes* memiliki kelebihan yakni dapat mempekirakan parameter yang dibutuhkan untuk klasifikasi dengan akurasi, presisi dan recall yang baik meskipun menggunakan jumlah data yang kecil (Jaya & Y., 2021). Pada penelitian sebelumnya dapat diketahui bahwa metode *Naïve Bayes* memiliki kemampuan untuk menganalisis sentimen dengan akurasi yang tinggi jika dibandingkan menggunakan metode KNN.

## 2. METODE PENELITIAN

### Tahapan Penelitian

Tahapan penelitian adalah serangkaian langkah atau proses yang dilakukan untuk merancang dan melakukan penelitian secara sistematis dan terstruktur, mulai dari penentuan topik penelitian, studi literatur, perancangan penelitian, pengumpulan data, analisis data, hingga penulisan laporan dan diseminasi hasil penelitian. Tahapan penelitian bertujuan untuk memastikan bahwa penelitian dilakukan dengan tepat dan menghasilkan hasil yang valid dan akurat. Berikut merupakan tahapan penelitian pada penelitian ini:



**Gambar 1.** Tahapan Penelitian

### Identifikasi Masalah

Hal pertama yang dilakukan pada pelaksanaan penelitian tugas akhir adalah indentifikasi masalah, dimana pada tahap ini mengidentifikasi masalah yang akan diangkat pada diteliti. Dalam mengidentifikasi masalah, penulis memanfaatkan informasi dari google play store untuk mengetahui ulasan – ulasan pada aplikasi, serta memanfaatkan jurnal terkini untuk menemukan isu terkini. Sehingga didapatkan masalah untuk menganalisis sentimen ulasan pada aplikasi cek bansos. Penjabaran terkait masalah yang diangkat telah dijelaskan pada sub bab latar belakang

### Studi Literatur

Setelah mengidentifikasi masalah, dilakukan studi literatur, penulis memanfaatkan artikel, jurnal dan *e-book*. Dengan melakukan studi literatur, penulis dapat memahami teori terkait analisis sentimen, metode *Naïve Bayes* dan teori lainnya untuk mendukung pelaksanaan penelitian tugas akhir

## Pengumpulan Data

Pada tahap pengumpulan data, penulis menggunakan cara *crawling* menggunakan bahasa pemrograman python. Data diambil dari situs resmi google play store terkait review atau ulasan aplikasi cek bansos (<https://play.google.com/store/apps/details?id=id.go.kemensos.pelaporan&hl=id&gl=US&pli=1>) dengan jumlah 4000 data. Analisis sentimen dengan 4000 data dapat memberikan hasil yang cukup representatif untuk mengetahui sentimen pengguna terhadap aplikasi di Google Play Store. Namun, untuk memastikan keakuratan hasil analisis, ada beberapa faktor yang perlu diperhatikan, seperti:

- a. Kualitas data: Pastikan data yang diambil berkualitas dan sesuai dengan tujuan analisis. Hal ini dapat dilakukan dengan melakukan filter dan pemrosesan data untuk membuang data yang tidak relevan atau mengandung noise.
- b. Metode analisis: Pilih metode analisis yang sesuai dengan tujuan analisis dan data yang diambil. Beberapa metode analisis yang umum digunakan untuk analisis sentimen antara lain *Naive Bayes*, *Support Vector Machine (SVM)*, dan *Deep Learning*.
- c. Evaluasi hasil analisis: Lakukan evaluasi hasil analisis untuk memastikan keakuratan dan validitas hasil analisis. Hal ini dapat dilakukan dengan membandingkan hasil analisis dengan data yang telah diverifikasi manual atau dengan menggunakan metrik evaluasi seperti akurasi, presisi, dan recall.

## Pre Processing

Setelah melakukan pengumpulan data, data perlu melalui tahap *pre-processing*, agar data dapat siap untuk diklasifikasi. Menurut Hendriyanto, *pre-processing* terdiri dari beberapa proses yang dijelaskan sebagai berikut (Hendriyanto, 2022):

- a. *Data cleansing*, melakukan penghilangan noise seperti menghilangkan tanda baca, simbol, angka, link dan tag.
- b. *Case folding*, mengubah teks yang memiliki huruf kapital menjadi huruf kecil.
- c. *Tokenizing*, melakukan pemisahan kata berdasarkan spasi pada kata yang ditemukan.
- d. *Stopword removal*, mengambil kata – kata penting dari hasil tokenizing dengan menghapus kata yang tidak memiliki arti.
- e. *Stemming*, melakukan pencarian kata dasar pada setiap kata setelah melakukan *stopword removal*. Pada tahap *stemming* juga menghilangkan kata imbuhan.
- f. Labeling Data

Pada tahap labeling data atau pelabelan data akan diproses otomatis dengan mealukan perhitungan nilai polaritas menggunakan library TextBlob. *Library TextBlob* memiliki

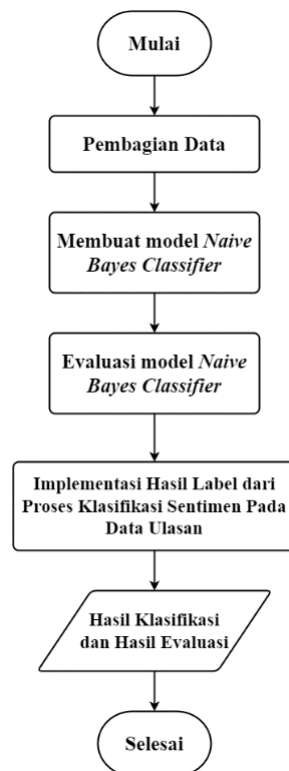
kelemahan dimana hanya dapat bekerja optimal pada teks berbahasa Inggris, sehingga sebelum dilakukannya labeling menggunakan TextBlob, data akan diterjemahkan dari bahasa Indonesia ke bahasa Inggris.

Pada penelitian ini menggunakan dua kategori sentimen yakni positif dan negatif. Library TextBlob akan melakukan labeling menggunakan nilai polaritas. Kisaran nilai polaritas untuk sentimen positif berada pada rentang (0,1). Sedangkan kisaran nilai polaritas untuk sentimen negatif berada pada rentang (-1,0).

- g. Setelah melakukan pelabelan data, dilakukannya *feature extraction* menggunakan TF-IDF. Dengan menggunakan TF-IDF dilakukannya pemberian bobot pada kata untuk mengklasifikasikan dokumen dan memudahkan untuk mengekstrak informasi pada ulasan pengguna. Pada proses TF-IDF menggunakan *library Scikit Learn* yakni from `sklearn.feature_extraction.text import TfidfVectorizer`.

### Klasifikasi Sentimen

Setelah data melalui tahap pre-processing, data siap untuk diklasifikasi. Pada proses klasifikasi terdiri dari beberapa tahap utama dimulai dari pembagian data, pembuatan model klasifikasi *Naïve Bayes*, evaluasi dan melakukan implementasi hasil label. Proses klasifikasi digambarkan pada diagram alur gambar 1.2 dan dijelaskan sebagai berikut:



Gambar 1. Klasifikasi Sentimen

## **Pembagian Data**

Setelah melakukan *feature extraction*, dilakukannya pembagian data menjadi data training atau data pelatihan dan data testing atau data pengujian. Rasio yang akan dilakukan uji coba oleh penulis ada 80:20.

## **Membuat Model Klasifikasi Naïve Bayes**

Pada tahap klasifikasi sentimen dilakukannya pembuatan model *Naïve Bayes* menggunakan *library Scikit Learn* yakni `sklearn.naive_bayes.MultinomialNB`. Setelah membuat model, data akan diklasifikasikan dengan model yang telah dibuat dan dihasilkannya sentimen pada setiap ulasan.

## **Evaluasi Model Naïve Bayes**

Tahap evaluasi dilakukan untuk mengetahui bagaimana kinerja model klasifikasi Naïve Bayes dalam mengklasifikasi sentimen. Evaluasi model diukur menggunakan confusion matrix yang terdiri dari nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Perhitungan nilai evaluasi menggunakan *library Scikit Learn* yakni `from sklearn.metrics import confusion_matrix`. Setelah melalui tahap ini akan diperoleh hasil klasifikasi dan hasil evaluasi model Naïve Bayes.

## **Implementasi Hasil Label**

Hasil label yang didapatkan dari proses klasifikasi sentimen menggunakan metode Naïve Bayes akan diimplementasi atau diterapkan pada data ulasan aplikasi cek bansos. Sehingga pada setiap ulasan akan diketahui kategori sentimennya.

## **3. HASIL DAN PEMBAHASAN**

Hasil dan pembahasan bertujuan untuk menjawab pertanyaan penelitian, memberikan penjelasan atas hasil yang diperoleh, dan menunjukkan arti penting dari penelitian yang dilakukan. Berikut merupakan bagian dari hasil dan pembahasan yang telah dilakukan pada penelitian ini:

### **Hasil Pengumpulan Data**

Pada penelitian ini, penulis memilih ulasan dengan score 1 hingga 5, dengan urutan ulasan MOST\_RELEVAN dan MOST\_NEWEST. Setiap score dan urutan diambil 400 ulasan. Sehingga hasil dari crawling data didapatkan sebanyak 4000 data. Data yang didapatkan memiliki beberapa atribut yang dituliskan pada Tabel 1. berikut ini:

**Tabel 1.** Atribut Data

<b>Nama Atribut</b>	<b>Keterangan</b>
reviewId	Berisi no Id review
username	Berisi username pengguna aplikasi cek bansos
userImage	Berisi link gambar profil pengguna
content	Berisi komentar pengguna terkait aplikasi cek bansos
score	Berisi angka penilaian yang diberikan untuk aplikasi cek bansos
thumbsUpCount	Berisi jumlah penyuka komentar
reviewCreatedVersion	Berisi seri atau versi aplikasi yang digunakan oleh pengguna
at	Berisi waktu pengguna mengunggah komentar
replyContent	Berisi balasan komentar antar pengguna
repliedAt	Berisi waktu balasan komentar antar pengguna
sortOrder	Berisi kategori komentar diurutkan berdasarkan most relevant dan newest
appId	Berisi Id aplikasi cek bansos

Tetapi tidak semua atribut akan dipakai dipenelitian ini. Penulis hanya menggunakan komentar atau ulasan pengguna aplikasi cek bansos yang berada pada atribut *content*. Agar penulis bisa mendapatkan atribut *content* saja, penulis melakukan penghapusan kolom atau atribut yang tidak digunakan seperti atribut *reviewId*, *username*, *userImage*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *At*, *replyContent*, *repliedAt*, *sortOrder*, dan *appId*.

**Tabel 2.** Review Pengguna

<b>Ulasan Pengguna</b>
Tiap bikin akun slalu eror data g bsa masuk, kalo EMG g bsa buat dftr lagi mnding dhpus aj aplikasi ny, dripda tiap download trs slalu GGL registrasi...
Bagaimana cara pendaftaran PKH saya tidak pernah Dapat bantuan apapun kecuali bantuan waktu musim korona dulu, di daerah saya yang dapat bantuan PKH termasuk orang-orang yang orang kaya
Ini aplikasi kenapa Udah daftar katanya udah berhasil daftar silahkan login Pas login username dan paswod tidak ditemukan Sumpah ga jelas
Saya sudah mencoba mendaftar sesuai KK dan KTP namun selalu muncul ""Alamat tidak sesuai"" padahal dari no KK, NIK, alamat, serta unggahan foto sudah sesuai dengan sebenar-benarnya. Untuk admin gimana solusinya
Setelah sekian lama akhir nya berhasil, dari mulai daftar akun, aktivasi sehari langsung bisa log in, sanggah dan usulan data juga langsung berhasil, moga diterima usulan bantuan saya, Aamiin
sudah buat usulan bantuan pbi di aplikasi dari bulan desember, tp sampai skarang 3 bulan lebih belum diapa2in, dicek aja tdk sama sekali. percuma bikin aplikasi kalo penerapan di lapangannya nol, ujungnya buang2 anngaran

### Hasil Pre Processing

Setelah didapatkannya data ulasan pengguna aplikasi cek bansos, data perlu melalui pre-processing sebelum memasuki tahapan klasifikasi sentimen. Hasil dari tahap pre-processing dituliskan sebagai berikut:

## Data Cleaning

**Tabel 3.** Data Cleaning

<i>Content</i>	<i>Data Cleaning</i>
Mantapp saya beri bintang 5 langsung karena ashoy cara krja tipu2 nya, saya rakyat miskin, trdampak covid19, ppkm, dtambah lagi dbikin ribet gini..mna hati nurani kalian.. Kasihan yg sdh daftar dan kasih data mereka,itu mau di apain NIK mereka slanjutx karena ini aplikasi ngga bisa login juga katanya..saya tdk mau ikutan..tdk semudah itu ngasih bantuan..	Mantap saya beri bintang langsung karena ashoy cara krja tipu nya, saya rakyat miskin trdampak covid ppkm dtambah lagi dbikin ribet gini mna hati nurani kalian Kasihan yg sdh daftar dan kasih data mereka itu mau di apain NIK mereka slanjutx karena ini aplikasi ngga bisa login juga katanya saya tdk mau ikutan tdk semudah itu ngasih bantuan

## Case Folding

**Tabel 4.** Case Folding

<i>Remove_Symbol</i>	<i>Lower</i>
Mantapp saya beri bintang langsung karena ashoy cara krja tipu nyasaya rakyat miskin trdampak covid ppkm dtambah lagi dbikin ribet ginimna hati nurani kalian Kasihan yg sdh daftar dan kasih data merekaitu mau di apain NIK mereka slanjutx karena ini aplikasi ngga bisa login juga katanyasaya tdk mau ikutantdk semudah itu ngasih bantuan	mantapp saya beri bintang langsung karena ashoy cara krja tipu nyasaya rakyat miskin trdampak covid ppkm dtambah lagi dbikin ribet ginimna hati nurani kalian kasihan yg sdh daftar dan kasih data merekaitu mau di apain nik mereka slanjutx karena ini aplikasi ngga bisa login juga katanyasaya tdk mau ikutantdk semudah itu ngasih bantuan

## Stopword Removal, Stemming, Tokenizing

**Tabel 5.** Stopword Removal, Stemming, Tokenizing

<i>Sebelum</i>	<i>Sesudah</i>
secara aplikasi ini mudah digunakan tetapi masa iya orang miskin bisa menggunakan aplikasi sprt ini jangankan menggunakan aplikasi ini hp pun kadang tidak punya aneh banget ini tidak ada cara lain apa	['aplikasi', 'mudah', 'iya', 'miskin', 'aplikasi', 'sprt', 'aplikasi', 'hp', 'kadang', 'aneh', 'banget']

## Hasil Labeling

Setelah melalui proses data *cleansing* hingga tokenizing, dilanjutkannya dengan tahap labeling. Namun sebelum melakukan labeling, penulis menerjemahkan data dari bahasa Indonesia ke bahasa Inggris menggunakan *function Translator()* dari *library googletrans*.

**Tabel 6.** Translate

<i>Hasil Terjemahan Menggunakan Googletrans</i>
<i>Translated (src=id, dest=en, text=Every time I create an account, there's always a data error, I can't enter my family card, but I can't even make a list, so I'm just going to dhpus, the application just drops off every time I download it, but it still doesn't register, pronunciation=Every time I create</i>

*an account, there's always a data error, I can't enter my family card, but I can't even make a list, so I'm just going to dhpus, the application just drops off every time I download it, but it still doesn't register, extra\_data="{\"translat...\"})*

Apabila dilihat dari hasil terjemahan data perlu dibersihkan kembali dikarenakan hasil terjemahan terdapat kata - kata yang tidak diperlukan seperti Translate, src, id, dest, en, text, pronunciation, extra\_data, dan translat. Selain itu juga terdapat simbol yang tidak diperlukan seperti (, {, = dan adanya duplikasi hasil terjemahan. Sehingga kata – kata, simbol – simbol, dan hasil terjemahan yang duplikat perlu di hapus. Penulis menggunakan Microsoft excel untuk menghapus kata-kata yang tidak diperlukan, lalu menggunakan python untuk menghapus simbol - simbol. Hasil teks terjemahan yang telah di bersihkan, dituliskan pada tabel 4.10.

**Tabel 7.** Translate yang sudah dibersihkan

Hasil Terjemahan Yang Telah Dibersihkan

*Every time I create an account, there's always a data error, I can't enter my family card, but I can't even make a list, so I'm just going to dhpus, the application just drops off every time I download it, but it still doesn't register*

Setelah melalui proses penerjemahan, dilanjutkan dengan *labeling* menggunakan TextBlob. TextBlob melakukan labeling atau pemberian label pada teks menggunakan nilai polaritas. Pada penelitian ini, apabila nilai polaritas lebih dari 0 akan diberi label sentimen positif. Sedangkan apabila nilai polaritas adalah kurang dari sama dengan 0 akan diberi label sentimen negatif.

**Tabel 8.** Labeling Texblob

<b>Teks</b>	<b>Polaritas</b>	<b>Label</b>
<i>Every time I create an account, there's always a data error, I can't enter my family card, but I can't even make a list, so I'm just going to dhpus, the application just drops off every time I download it, but it still doesn't register</i>	0.0	Negatif
<i>try to register according to the family card, the sitting card, the address always appears according to the address, even though the family card number is the address of the neighborhood association, upload the photo according to the correct admin, what is the solution</i>	0.124	Negatif
<i>it's cool, I never got social assistance, just php"</i>	0.19	Positif

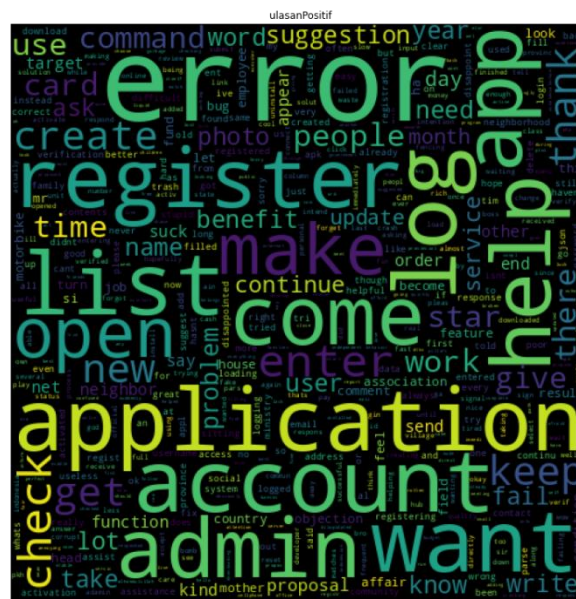
**Hasil Feature Extraction**

Setelah melalui tahap labeling dilakukannya tahap feature extraction yang digunakan untuk menilai relevansi kata pada dokumen teks. Penulis menggunakan TF-IDF yang tersedia pada library scikit-learn untuk feature extraction. Saat melakukan TF-IDF perlu mencantumkan

stopword. Agar kata - kata yang tidak bermakna tidak diperhitungkan pada TF-IDF. TF-IDF memiliki parameter min\_df yang berfungsi memfilter kata, penulis menggunakan min\_df= 5.

	able	accept	access	accidentally	account	accounts	accurate	activate	activated	activation	...	weird	well	whats	while	wifi	work	worthy	write	wrong	year	
0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
1	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
2	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
3	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
4	0.0	0.0	0.000000	0.0	0.278179	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.424509	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1610	0.0	0.0	0.937817	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
1611	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
1612	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
1613	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
1614	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	

Gambar 2. TF IDF



Gambar 3. Ulasan Positif



Gambar 4. Ulasan Negatif

## Hasil Klasifikasi Sentimen

Model yang telah dibuat perlu dilakukannya evaluasi. Penulis menggunakan evaluasi menggunakan *confusion matrix* yang tersedia pada *library* sklearn. Hasil dari *confusion matrix* ditunjukkan pada Gambar 1.6 berikut:

```

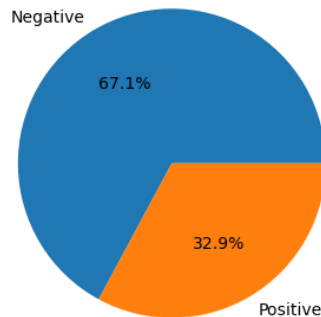
Classification Report

```

	precision	recall	f1-score	support
0	0.75	0.98	0.85	496
1	0.89	0.33	0.49	243
accuracy			0.77	739
macro avg	0.82	0.66	0.67	739
weighted avg	0.80	0.77	0.73	739
[[486 10]				
[162 81]]				

**Gambar 5.** Confusion Matrix

Kemudian hasil implementasi di simpan dalam bentuk excel dan dirapikan secara manual menggunakan Microsoft excel agar mudah untuk mengetahui mana teks dengan sentimen negatif dan positif. Hasil klasifikasi sentimen menggunakan *Naïve Bayes* dari 3692 data, didapatkannya hasil 1214 positif dan 2478 negatif. Banyaknya sentimen di visualisasikan menggunakan diagram batang pada Gambar 1.7.



**Gambar 6.** Klasifikasi Naïve Bayes

Hasil klasifikasi sentimen menggunakan metode *Naïve Bayes* pada setiap teks ulasan dituliskan pada Tabel 1.9.

**Tabel 9.** Hasil Klasifikasi Sentimen NB Pada Setiap Teks Ulasan

Sentimen	Teks Ulasan Bahasa Inggris
Negatif	<i>Every time I create an account, there's always a data error, I can't enter my family card, but I can't even make a list, so I'm just going to dhpus, the application just drops off every time I download it, but it still doesn't register</i>
Negatif	<i>try to register according to the family card, the sitting card, the address always appears according to the address, even though the family card number is the address of the neighborhood association, upload the photo according to the correct admin, what is the solution</i>

Positif	<i>it's cool, I never got social assistance, just php"</i>
Negatif	<i>it's been so long, the results start registering account activation, the day immediately logs in, the objection to the data proposal is direct, the results are expected to be accepted, the proposal for help, aamiin</i>
Positif	<i>Register as an application to receive direct cash assistance, BBM, then the post office disburses my father's name, register to receive direct cash assistance</i>

#### 4. KESIMPULAN DAN SARAN

Semua tahapan penelitian telah dilakukan, sehingga ditarik kesimpulan sebagai berikut:

1. Pada data aktual yang telah dilakukan analisis sentimen menggunakan metode Naïve Bayes mampu menganalisis sentimen review aplikasi cek bansos dengan hasil 1214 positif dan 2478 negatif dengan akurasi yang baik sebesar 0.77 atau 77%.
2. Dengan melakukan penelitian penulis menambah pengetahuan terkait implementasi metode Naïve Bayes pada analisis sentimen. Berdasarkan hasil analisis sentimen, aplikasi cek bansos mendapatkan banyak sekali sentimen negatif dari masyarakat, sehingga dapat disimpulkan bahwa aplikasi cek bansos kurang memberikan manfaat bagi masyarakat.

#### DAFTAR REFERENSI

- Hendriyanto, M. D. (2022). Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax. *JURNAL ILMIAH INFORMATIKA*, 10(02).
- Jaya, T. S., & Y., M. (2021). Predicting the Quality of Pineapple Using the Naive Bayes Classifier Method. *2nd International Conference on Agriculture and Applied Science (ICoAAS 2021)*.
- Latif, I. S., & P., I. A. (2022). Problematika Penyalahgunaan Bantuan Sosial Pada Masa Pandemi. *JUSTISI*, 8(2), 95–107.
- Mansour, N. A., & A., A. I. S. B. H. A. (2022). Accurate detection of Covid 19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. *Journal of Ambient Intelligence and Humanized Computing*, 13, 41–73.
- Noerkaisar, N. (2021). EFEKTIVITAS PENYALURAN BANTUAN SOSIAL PEMERINTAH UNTUK MENGATASI DAMPAK COVID-19 DI INDONESIA. *Jurnal Manajemen Perbendaharaan*, 2(1), 83–101.
- Oktaviani, V., & S., B. W. H. Y. R. S. (2020). Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method. *ISNPINSA 2020 Journal of Physics: Conference Series IOP Publishing*.
- Pratama, R. A., & S., M. I. F. (2022). Sosialisasi Penyaluran Bantuan Sosial (Bansos) Pada Era New Normal Di Desa Ciamis Kecamatan Sungkai Utara Kabupaten Lampung Utara. *JPM: Jurnal Pengabdian Kepada Masyarakat*, 3(1), 107–118.

- Pristiyono, & R., M. R. M. A. A. I. A. A. F. H. (2020). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. *Annual Conference on Computer Science and Engineering Technology (AC2SET) 2020*.
- Rahmansyah, W., & I., R. A. Q. R. A. S. (2020). PEMETAAN PERMASALAHAN PENYALURAN BANTUAN SOSIAL UNTUK PENANGANAN COVID-19 DI INDONESIA. *Jurnal Pajak Dan Keuangan Negara*, 2(1), 90–102.
- Ramdhani, S. L., & H., R. A. M. A. (2018). Sentiment Analysis of Product Reviews using Naive Bayes Algorithm: A Case Study. *The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) 2018*.
- Samsir, & W., K. A. H. D. R. A. R. (2022). Implementation Naïve Bayes Classification for Sentiment Analysis on Internet Movie Database. *Building of Informatics, Technology and Science (BITS)*, 4(1), 1–6.
- Surohman, & W., S. A. R. F. F. (2020). Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor. *Jurnal Sains Dan Manajemen*, 8(1), 93–105.