**SENTIMEN ANALISIS PADA APLIKASI CEK BANSOS BERDASARKAN ULASAN PENGGUNA**

**Akhmadhetta Hafid P1, Fauzan Pradana2, Taufik Anwar3**

1,2,3 Prodi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang

12041720221@student.polinema.ac.id, 2 2041720224@student.polinema.ac.id, 3  2041720010@student.polinema.ac.id

**Abstrak**

**Pendahuluan:** Kumpulan ulasan dari pengguna google play store terhadap aplikasi *“Aplikasi Cek Bansos”* dapat dimanfaatkan menjadi sebuah analis sentiment. Data yang diperoleh melalui teknik *scraping* kemudian diolah menggunakan beberapa teknik pengolahan teks diantaranya *tokenize*, klasifikasi, *stemming*, dll. Kemudian dilakukan pemodelan dengan menggunakan dua algoritma yang berbeda untuk dilakukan perbandingan, algoritma yang digunakan yaitu *Support Verctor Machine* dan *Naïve Bayes Classifier* dengan tujuan menemukan akurasi terbaik antara keduanya. Setelah dimodelkan hasil akurasi tertinggi pada penelitian ini adalah algoritma X dengan akurasi X% dibandingkan dengan algoritma Y yang memiliki akurasi Y%

**Kata kunci :** Sentimen Analisis; *Support Vector Machine*; *Naïve Bayes*.

***Abstract***

***Introduction:*** *A collection of reviews from google play store users on the "Aplikasi Cek Bansos" application can be utilized as a sentiment analyst. The data obtained through scraping techniques is then processed using several text processing techniques including tokenization, classification, stemming, etc. Then modelling is done using two different algorithms for comparison. Then modelling is done using two different algorithms for comparison, the algorithms used are Support Vector Machine and Naïve Bayes Classifier with the aim of finding the best accuracy between the two. After modelling, the highest accuracy result in this study is algorithm X with X% accuracy compared to algorithm Y which has Y% accuracy.*

***Keyword :*** *Sentiment Analysis; Support Vector Machine; Naïve Bayes.*

1. **Pendahuluan**

Kemiskinan merupakan permasalahan yang umum bagi negara berkembang. Indonesia sebagai salah satu negara yang sedang berkembang terus melakukan upaya untuk menyelesaikan masalah kemiskinan ini dengan berbagai kebijakan (Murdiana mulyana).

Kementerian Sosial Republik Indonesia meluncurkan program bantuan sosial pada 04 januari 2021. Program tersebut meliputi Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT)/Program Sembako, dan Bantuan Sosial Tunai (BST). Program-program tersebut ditunjukan untuk mengurangi kemiskinan dengan memenuhi dan menjamin kebutuhan dasar serta meningkatkan taraf hidup bagi masyarakat penerima program bantuan sosial. Seperti yang dikutip dari Bisnis.com, mengungkapkan bahwa Kementerian Keuangan (Kemenkeu) yakin dapat menekan angka kemiskinan dengan adanya program bantuan sosial yang diberikan kepada masyarakat. Akan tetapi muncul masalah berikutnya yaitu adanya ketidaktepatan sasaran penerima bantuan sosial. Untuk mengatasi hal tersebut pemerintah memfasilitasi dengan platform berupa *website* dan aplikasi berbasis *mobile* untuk mengakses informasi terkait bantuan sosial.

Aplikasi *“Aplikasi Cek Bansos”* merupakan aplikasi berbasis *mobile* yang diluncurkan oleh Kementerian Sosial (Kemensos) pada layanan google play store. Kemensos meluncurkan aplikasi berbasis *mobile* untuk mempermudah masyarakat Indonesia dalam mendapatkan informasi terkiat bantuan sosial, mengutip dari website GoodStats, menginformasikan bahwa berdasarkan laporan dari perusahaan riset Data Reportal mengatakan bahwa jumlah pengguna perangkat seluler di Indonesia pada tahun 2022 mencapai 370,1 juta. Jumlah tersebut meningkat sebesar 13 juta atau 3.6 persen.

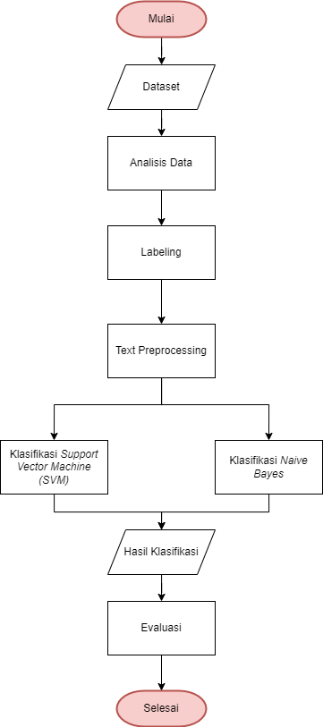
Setiap aplikasi yang di publikasi pada layanan google playstore dapat diberikan ulasan oleh pengguna. Fitur ulasan ini memiliki nilai negatif dan positif, dengan adanya ini pengguna lain akan mengetahui kualitas aplikasi tersebut sebelum memasang pada ponselnya. Kemudian dari ulasan tersebut dapat dijadikan sebagai data untuk analisis sentimen guna mengetahui opini pengguna terhadap aplikasi tersebut.

Analisis sentimen dapat diimplementasikan pada kecerdasan buatan yang ditanamkan pada perangkat lunak komputer. Dalam mengimplementasikan analisis sentimen dapat digunakan beberapa metode diantaranya yang digunakan pada peneltian ini adalah *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier*. Kedua metode ini dipilih karena memiliki akurasi yang tinggi pada jenis data teks.

Oleh karena itu, penelitian ini dilakukan untuk menganalisis sentimen terhadap *“Aplikasi Cek Bansos”* guna mengetahui kepuasan masyarakat terhadap fasilitas yang disediakan Kemensos untuk pendistribusian informasi terkait bantuan sosial.

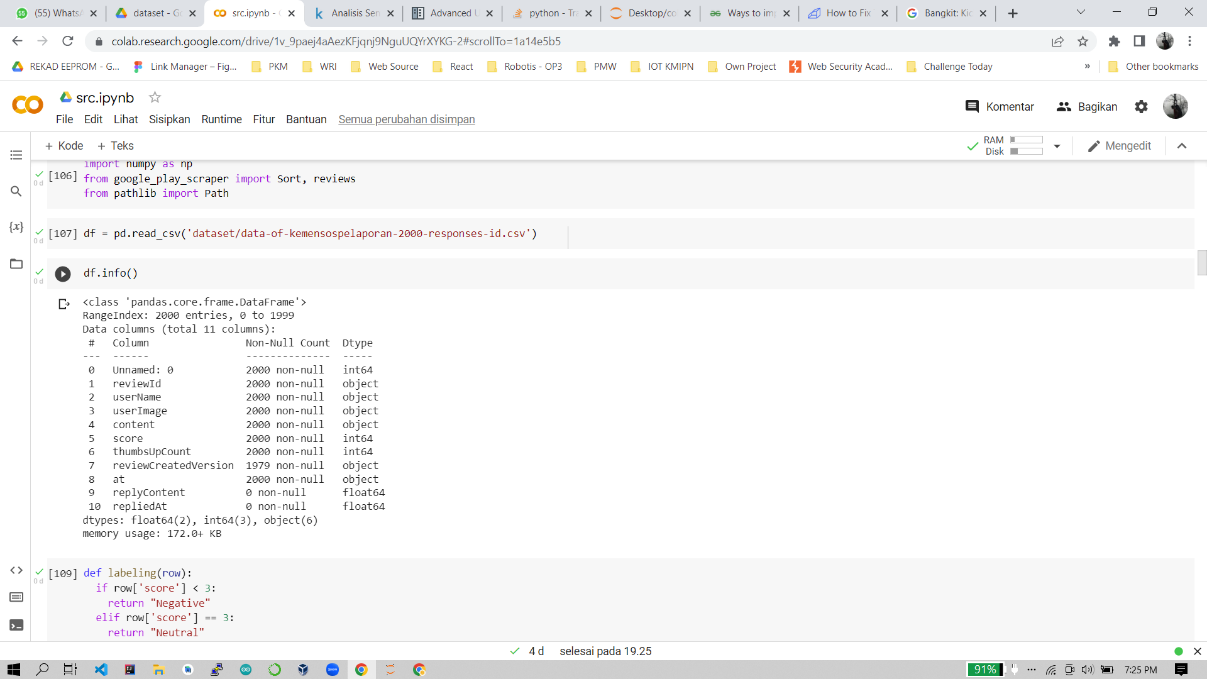
1. **Metode Penelitian**

Pada penelitian ini, penulis menggunakan *dataset* berupa ulasan yang terdapat pada *“Aplikasi Cek Bansos”* di dalam layanan google play store. Setelah *dataset* didapatkan lalu dianalisis untuk memahami struktur datanya. Lalu diikuti dengan proses pembuatan label untuk menentukan ulasan tersebut termasuk ke dalam kategori negatif, positif, atau netral. Kemudian untuk memastikan data yang akan diproses berjalan dengan baik pada model machine learning yang dibuat, maka perlu dilakukan pembersihan data *(data cleansing)*. Selanjutnya setelah struktur data terstandarisasi dilakukan pemodelan dengan menggunakan dua algoritma diantaranya, algoritma *Support Vector Machine* dan *Naïve Bayes Classifier*. Terakhir hasil dari pemodelan akan dievaluasi guna mendapatkan akurasi yang lebih baik. Tahapan metode penelitian dapat dilihat pada Gambar 1.



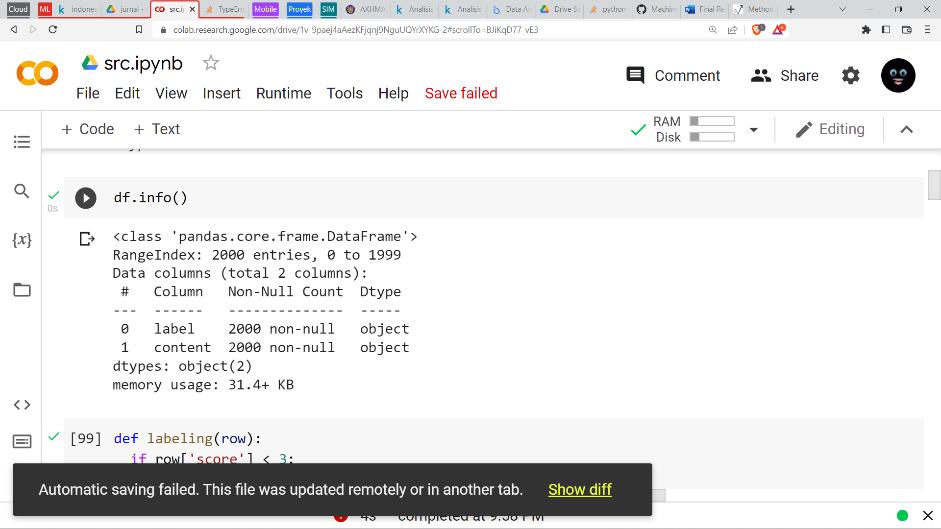
Gambar 1 Tahapan Metode Penelitian

Tahap pertama yang dilakukan adalah pengumpulan data dan memahami struktur data. Pengumpulan data pada penelitian ini menggunakan teknik *scraping data* pada ulasan *“Aplikasi Cek Bansos”* didalam layanan google play store. Data yang diperoleh sebanyak 2000 baris dan 10 kolom/ variable. Variabel dari data yang didapatkan adalah reviewId, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, repyContent, dan repliedAt. Untuk lebih jelas dapat dilihat pada Gambar 2.



Gambar 2 Variabel Dataset

Untuk melakukan analisis sentimen perlu dilakkukan pembuatan label *(labelling).* Pada tahap ini dataset diklasifikasikan menjadi tiga kategori yaitu, negatif, netral, atau positif. Selain melakukan *labelling,* variabel yang relevan untuk proses pemodelan akan dipilih terlebih dahulu. Variabel yang relevan dengan proses pemodelan bisa dilihat pada gambar 3. Setelah dilakukan *labelling* dan pemilihan variabel tahap selanjutnya adalah melakukan standarisasi struktur data yang didalamnya terdapat proses *case folding* (mengubah isi dataset menjadi huruf kecil/*lowercase*), *tokenizing* (menguraikan hasil *case folding* menjadi token-token atau kata-kata), formalisasi (mengubah kata tidak baku menjadi baku), *stop word removal* (menghaus kata yang tidak penting dari hasil formalisasi), dan *stemming* (mengubah kata berimbuhan menjadi kata dasar). Setelah dilakukan *pre-processing* pada dataset akan terlihat seperti gambar 4.

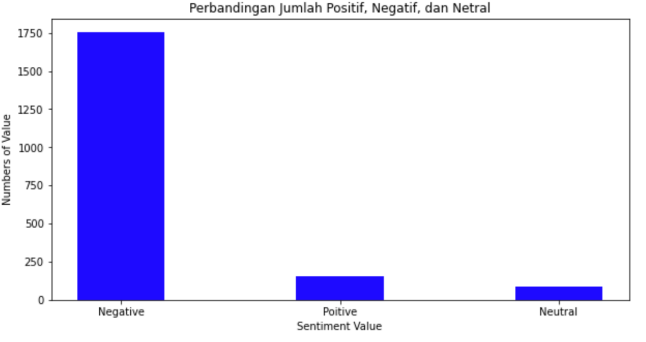


Gambar 3 Variabel Yang Berkolerasi

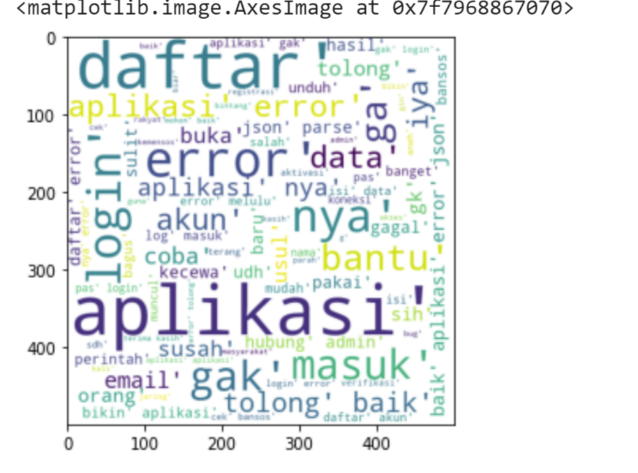


Gambar 4 Hasil Preprocessing

Sebelum melakukan pemodelan untuk memudahkan pembacaan data dilakukan visualisasi terlebih dahulu. Pada dataset yang digunakan memiliki jumlah data dengan label negatif lebih banyak daripada data netral maupun positif seperti yang digambarkan pada gambar 5. Selanjutnya untuk mengetahui kata apa saja yang sering muncul dilakukan visualisasi dengan menggunakan word cloud seperti gambar 6.

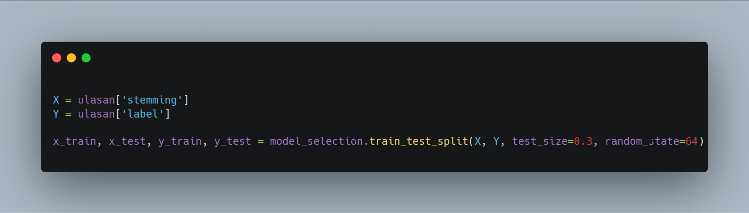


Gambar 5 Perbandingan Jumlah Positif, Negatif, dan Netral



Gambar 6 Word Cloud

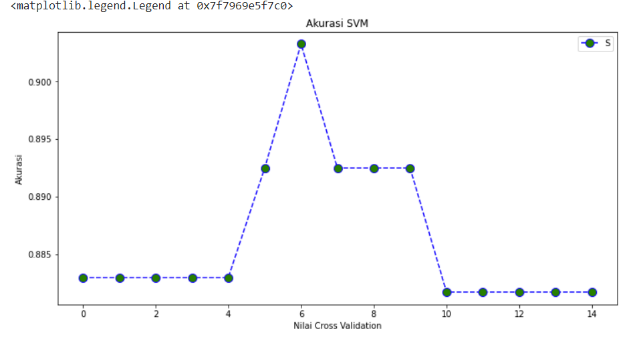
Tahap selanjutnya setelah dilakukan *pre-processing dataset* adalah pembuatan model. Sebelum membuat model, pertama-tama yan g harus dilakukan adalah splitting dataset sehingga menjadi data training dan data testing. Komposisi yang digunakan pada penelitian ini adalah 7:3, yaitu 70% untuk data training dan 30% dan ditampung dalam 4 variabel yaitu X\_train, X\_test, y\_train, y\_test. Dalam train\_test\_split berisi df[‘stemming’] yang menjadi nilai fitur dan df[‘label’] menjadi nilai label seperti yang ditunjukkan pada gambar 7.



Gambar 7 Splitting Dataset

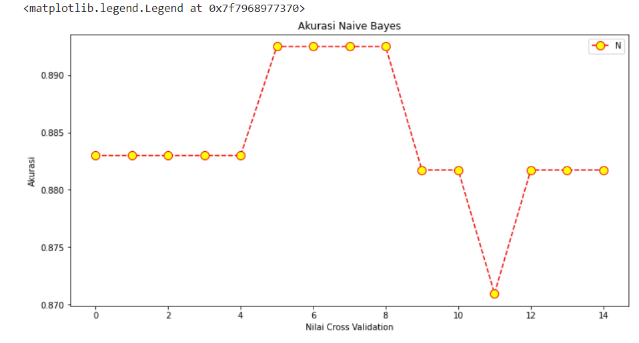
Teori-toeri sedikit, ditambah perhitungan manual biar nice.

Selanjutnya adalah pembuatan model dengan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier*. Pada model dengan menggunakan algoritma *Support Vector Machine (SVM)* menghasilkan akurasi seperti gambar 8 dan akurasi rata-rata sebesar 0.886 atau 88.6%.



Gambar 8 Akurasi SVM

Pada model dengan algoritma *Naïve Bayes Classifier* menghasilkan akurasi seperti gambar 9 dan akurasi rata-rata sebesar 0.884 atau 88.4%.



Gambar 9 Akurasi Naive Bayes Classifier

Dari perbandingan akurasi rata-rata kedua algoritma tersebut yang mana pada algoritma *Support Vector Machine (SVM)* memiliki akurasi rata-rata sebesar 88.6% dan *Naïve Bayes Classifier* sebesr 88.4%, dapat disimpulkan bahwa pada penelitian ini dengan dataset ulasan “Aplikasi Cek Bansos” algoritma *Support Vector Machine (SVM)* memiliki akurasi yang lebih baik meskipun hanya sebesar 0.2%. Namun, tidak dapat dipungkiri bahwa kedua algoritma tersebut mampu bekerja dengan cukup baik karena memiliki akurasi yang bagus yaitu sama-sama di kisaran hampir 90%.

1. **Hasil dan Diskusi**

Hasil klasifikasi dari dataset kanker paru-paru menggunakan metode *decision tree* dan *k nearest neighbor* menghasilkan akurasi yang lumayan baik yakni *decision tree* senilai 1.0 dan *k nearst neighbor* senilai 0.99. Performa klasifikasi dari hasil pengujian kepada 20% dataset menghasilkan presentase sebesar lebih dari 80%. Akurasi klasifikasi semakin baik jika data latih terus ditambah. Hasil dari penelitian ini juga menunjukkan bahwa algoritma *decision tree* dan *k nearest neighbor* handal dalam menangani kesalahan inputan dengan memperbaharui bobot.

1. **Kesimpulan**

Dari penelitian ini dapat disimpulkan bahwa pengembangan teknik klasifikasi menggunakan metode *Decision Tree* dan *K Nearest Neighbor* berhasil mengklasifikasikan pasien berdasarkan pada variable-variable yang digunakan pada dataset. Keluaran dari program klasifikasi yang dikembangkan ini dapat menggolongkan pasien apakah termasuk kedalam kategori level rendah, menengah, hingga tinggi.

1. **Daftar Pustaka**

*Indonesia*. (n.d.).