

Analisis Sentimen Ulasan Aplikasi KLC2 pada Google Play Store menggunakan Algoritma Naïve Bayes Classifier (NBC)

Agus Hekso Pramudijono¹, Jamila Lestyowati²

^{1,2} *Pusdiklat keuangan Umum, BPPK, Kementerian Keuangan
Jl Pancoran Timur II No 1 Jakarta Selatan, DKI Jaakrta, Indonesia*

¹ gushekso@kemenkeu.go.id

² jlestyowati@kemenkeu.go.id

Received on 31-10-2023, revised on 07-11-2023, accepted on 16-11-2023

Abstract

Learning in the digital era through various platforms is widely used in companies, private organizations, and government. KLC2 was developed into a knowledge management system platform at the Ministry of Finance. KLC2 presents learning information through training organized by the Education and Training Center at BPPK. Various activities can be attended through KLC2. KLC2 can be downloaded via the Google Play Store. One of the essential features of applications on the Google Play Store is reviews, which can be used by users to rate and provide opinions in the form of review text about the application they are using. A popular method that is often used to convey ideas and feelings about service applications is reviewed. Sentiment analysis is the process of classifying text into several classes, such as positive, negative, or neither. This research aims to analyze the sentiment of KLC2 application reviews on the Google Play Store. The data source uses primary data from the Google Play Store and is analyzed using NBC. The research results show that the accuracy rate is 94%. Meanwhile, 83% of reviews were positive, 13% were negative, and 4% were neutral.

Keywords: Sentiment Analysis, Google Play Store, Naïve Bayes Classifier, KLC2

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Jamila Lestyowati
Pusdiklat Keuangan Umum
Jl Pancoran Timur II no 1 Jakarta Selatan, Indonesia
Email: jlestyowati@kemenkeu.go.id

I. INTRODUCTION

Teknologi Informasi dan Komunikasi (TIK) telah menjadi bagian penting dari kehidupan sehari-hari kita di zaman sekarang. TIK telah mengubah cara kita berkomunikasi, bekerja, dan bahkan berinteraksi satu sama lain. TIK memungkinkan manusia untuk belajar, mencari informasi, menjalankan bisnis dengan lebih efektif dalam hitungan detik. TIK juga memungkinkan berkomunikasi dengan orang lain di seluruh dunia melalui berbagai aplikasi yang tersedia.

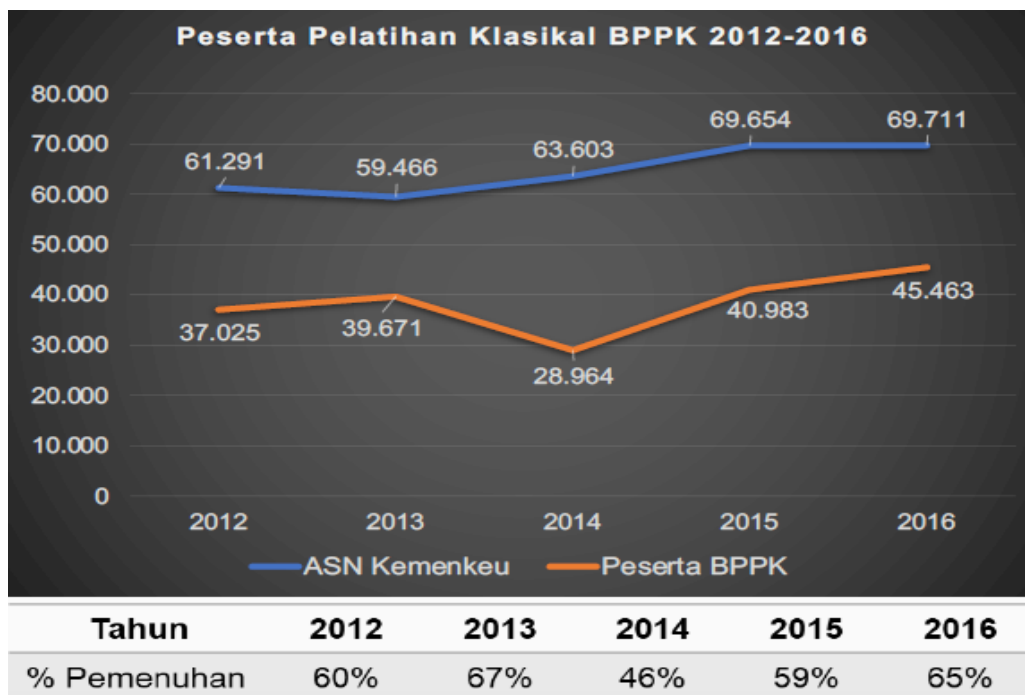
Dalam dunia pemerintahan, pemerintah Indonesia telah menetapkan Peraturan Presiden (PERPRES) Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik [1]. Hal ini menjadi upaya Pemerintah Indonesia untuk memodernisasi sektor pemerintahan dengan memanfaatkan teknologi informasi. Dengan penggunaan teknologi yang tepat, diharapkan pelayanan publik dapat lebih efisien, transparan, dan dapat memenuhi kebutuhan masyarakat secara lebih baik.

Dalam dunia pendidikan, teknologi ini telah mengubah cara belajar. Platform pembelajaran jarak jauh, kursus online, dan sumber daya pendidikan digital telah membuat pendidikan lebih fleksibel dan dapat diakses oleh siapa saja. Pembelajaran digital yang didorong oleh kemajuan TIK dan era digital telah mengubah cara kita belajar, mengajar, dan mendapatkan pengetahuan. Dunia pelatihan SDM juga mengalami pergeseran. Pembelajaran digital telah mengubah dunia pelatihan dan hal ini semakin penting dalam masyarakat yang terhubung dengan internet.

Peserta dapat belajar secara mandiri dengan bantuan perangkat lunak pembelajaran atau kursus online, yang memungkinkan pembelajaran yang lebih personalisasi dan disesuaikan dengan kebutuhan individu. Pembelajaran digital juga menawarkan akses ke pelatihan berkualitas tinggi tanpa batasan geografis. Peserta di wilayah yang beragam di seluruh dunia dapat mengikuti pembelajaran online tanpa harus meninggalkan tempat tinggal mereka. Pembelajaran era digital memberi fleksibilitas waktu karena mereka dapat belajar kapan saja. Hal ini bagus untuk orang yang memiliki jadwal yang padat atau tanggung jawab lainnya.

Kementerian Keuangan memiliki visi “Menjadi Pengelola Keuangan Negara untuk Mewujudkan Perekonomian Indonesia yang Produktif, Kompetitif, Inklusif, dan Berkeadilan untuk Mendukung Visi dan Misi Presiden dan Wakil Presiden: Indonesia Maju yang Berdaulat, Mandiri dan Berkepribadian Berlandaskan Gotong Royong.” Untuk menuju visi tersebut diperlukan SDM yang memiliki nilai-nilai unggul dan berkinerja. Pengembangan SDM dilakukan melalui pelatihan. Pelatihan SDM merupakan suatu keharusan dan menjadi strategi penting di era yang menuntut untuk adaptif dan responsif. Organisasi yang berinvestasi dalam pengembangan dan pelatihan SDM akan cenderung lebih sukses, lebih kompetitif, dan lebih siap menghadapi tantangan masa depan [2].

BPPK sebagai institusi pelatihan di Kementerian Keuangan bertugas menyelenggarakan pelatihan. Selama rentang waktu 2012 - 2016 jumlah peserta pelatihan secara klasikal di BPPK mengalami perkembangan signifikan sebagai berikut (Gambar 1). Jumlah tersebut masih belum sebanding dengan jumlah pegawai Kemenkeu.



Gambar 1 Jumlah Peserta Pelatihan Klasikal BPPK

Sumber: BPPK (2023) [3]

Menyediakan pembelajaran yang relevan dan sesuai dengan kebutuhan Kemenkeu adalah komponen penting dalam mencetak SDM yang handal tersebut. Oleh karena itu BPPK sebagai lembaga pelatihan yang bukan sekadar sebagai training center namun menerapkan strategi pembelajaran melalui Kemenkeu corporate university. Salah satu teknologi pembelajaran yang digunakan Kemenkeu adalah penerapan platform Kemenkeu Learning Center sebagai tools pembelajaran untuk pelatihan di BPPK. Diharapkan bahwa penggunaan KLC ini akan mendorong keberhasilan penyelenggaraan pelatihan untuk mewujudkan

visi dan misi Kemenkeu. Sebagai aplikasi pembelajaran, KLC2 tersedia di google play store dan bisa diunduh oleh siapa saja.

Selama rentang waktu 2020 s.d. 2022, KLC telah memberikan manfaat sebagai berikut

Publikasi Dokumen Pengetahuan	8.071
Jumlah course	2.519
Pengguna ASN Kemenkeu teregistrasi	85.409
Total pengguna teregistrasi	207.273

Sumber: BPPK (update s.d. Nov 2022)

Penelitian ini bertujuan untuk menganalisis ulasan sentimen aplikasi KLC2 ada *google play store*. Rumusan masalah pada penelitian ini adalah “Berapa tingkat akurasi KLC2? Bagaimana sentimen yang ada pada KLC2?”

1. Learning Management System (LMS)

LMS adalah perangkat lunak atau platform yang digunakan untuk mengelola dan menyampaikan pendidikan atau pelatihan secara online. LMS biasanya digunakan oleh institusi pendidikan, perusahaan, atau organisasi untuk memantau dan mengevaluasi kemajuan peserta didik/ pelatihan. LMS adalah aplikasi perangkat lunak yang secara otomatis dapat menangani administrasi, pelaksanaan, dan pelaporan dari sebuah aktifitas pelatihan [4]. LMS pertama kali muncul pada tahun 1990an ketika dikenal sebagai sistem pembelajaran terpadu berbasis komputer. Hal ini kemudian diadopsi dengan nama lain misalnya “learning platform”, “distributed learning system”, “course management system”, “content management system”, “instructional management system”, “portal”, dan “virtual learning environment” [5].

LMS cukup banyak memberikan kontribusi dari segi pemanfaatan. Fleksibilitas LMS sangat memungkinkan pendidik dan siswa mengaksesnya kapan saja dan dimana saja serta melalui device apa saja (PC, tablet, ataupun smartphone) [6].

Beberapa fitur LMS termasuk:

1. Pengelolaan Konten pembelajaran. LMS memungkinkan penyelenggara pelatihan mengunggah, mengatur, dan menyebarkan materi pembelajaran seperti video, teks, gambar, dan file lainnya.
2. Manajemen user. LMS memungkinkan pengguna mendaftar, mengatur data pribadi, dan mengakses kursus atau pelatihan yang sesuai.
3. Pembelajaran online. Pengajar dapat menyampaikan materi, mengirim tugas, mengatur ujian, dan berinteraksi dengan peserta secara online.
4. Kolaborasi. LMS memiliki fitur kolaborasi seperti forum diskusi, obrolan, atau alat berbagi yang membantu peserta pelatihan berkomunikasi satu sama lain.
5. Evaluasi LMS memungkinkan peserta pelatihan melacak kemajuan perkembangan mereka mengikuti pelatihan. Hal ini biasanya ditunjukkan dengan prosentase progress.

2. Knowledge Management System (KMS)

KMS adalah sistem atau platform yang dimaksudkan untuk mengelola, menyimpan, mengakses, dan berbagi pengetahuan dan informasi dalam sebuah organisasi atau entitas. Tujuan utama KMS adalah untuk memfasilitasi pengumpulan, organisasi, distribusi, dan pemanfaatan pengetahuan yang dimiliki oleh individu atau entitas untuk mencapai tujuan bisnis atau tujuan lainnya. KMS juga membantu organisasi dalam meningkatkan produktivitas, inovasi, efisiensi, dan efisiensi. Azzumar (2021) menyebutkan KMS merupakan sebuah sistem yang dimiliki perusahaan maupun organisasi untuk mengidentifikasi, memproses dan membagikan knowledge atau pengetahuan, sesuai dengan berjalannya mencapai tujuan organisasi ataupun Perusahaan [7].

Arribathi (2020) menyebutkan Knowledge management memiliki kemampuan untuk menciptakan, mengkomunikasikan, dan menerapkan pengetahuan untuk menciptakan dan mempertahankan nilai kunci kompetensi bisnis. Istilah ini secara luas didefinisikan sebagai "pengelolaan atau manajemen pengetahuan organisasi untuk menciptakan nilai bisnis dan membangun daya saing [8].

Beberapa komponen penting KMS adalah sebagai berikut.

1. Tempat penyimpanan pengetahuan. KMS menyediakan tempat untuk menyimpan dan mengatur pengetahuan. Jenis pengetahuan ini dapat berupa catatan, presentasi, dokumen, basis data, video, dan lain-lain yang menjadi perwujudan dari pengetahuan tacit yang dimiliki individu atau organisasi

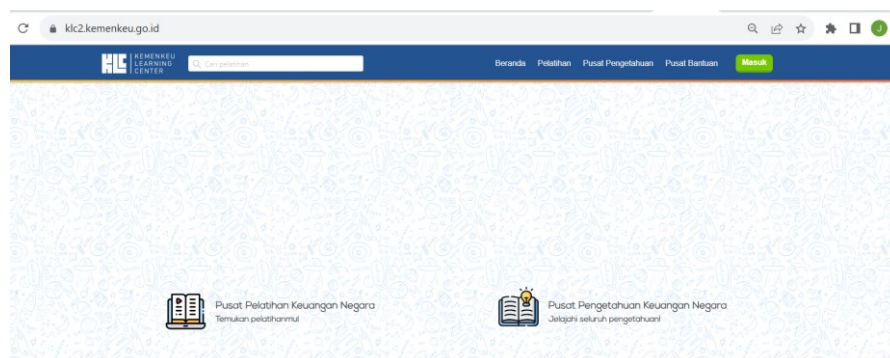
2. Akses knowledge. KMS memungkinkan pengguna menemukan dan mengakses pengetahuan atau informasi dengan cepat dan mudah.

3. Sarana berbagi pengetahuan dan kolaborasi. KMS memungkinkan seseorang untuk berbagai pengetahuan yang dimilikinya dan bekerja sama dengan berbagi ide, pengetahuan, dan pengalaman pihak lain.

3. *Kemenkeu Learning Center*

Kemenkeu Learning Center merupakan media pembelajaran online yang membahas berbagai materi tentang Pengelolaan Keuangan Negara yang dapat diakses oleh seluruh pegawai Kementerian Keuangan dan masyarakat umum. Kemenkeu Learning Center berfungsi untuk mendukung proses pendidikan dan pelatihan yang diselenggarakan di lingkungan Kementerian Keuangan. Di KLC tersedia berbagai pelatihan yang diselenggarakan. Peserta pelatihan dapat mengikuti pelatihan dan melihat beragam microlearning di KLC. Pelatihan dan microlearning tersebut diharapkan dapat menjadi sarana untuk peningkatan kualitas, produktivitas kerja, dan kinerja pegawai. KLC dapat diakses pada alamat www.klc2.kemenkeu.go.id. KLC menjadi media pembelajaran online terkait dengan pengelolaan keuangan negara. KLC dapat diakses oleh pegawai Kemenkeu dan masyarakat umum. KLC berfungsi sebagai Learning Management System (LMS) dan Knowledge Management System (KMS) (BPPK, 2023)

Tampilan KLC2 adalah sebagai berikut.

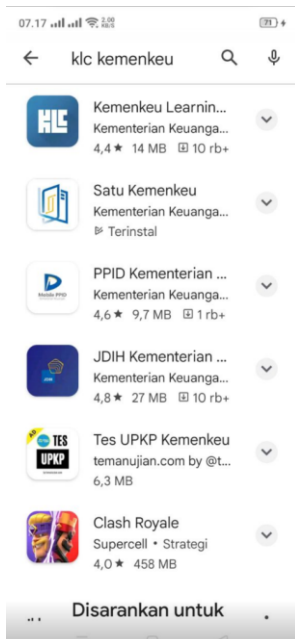


Gambar 2 Tampilan KLC2

Sumber: website KLC2

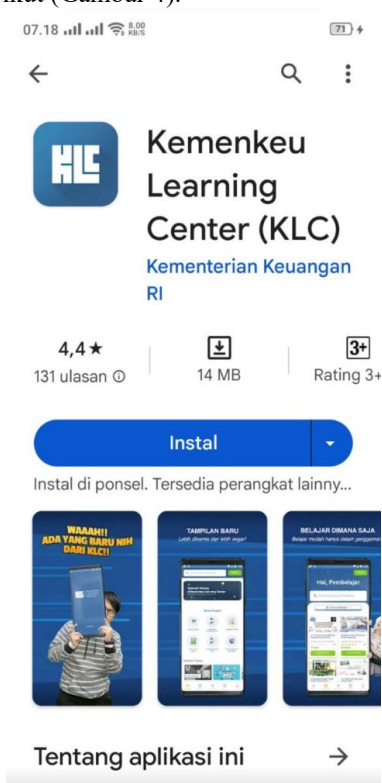
Karena KLC2 menawarkan pembelajaran yang dapat diikuti oleh siapa saja, dimana saja, dan kapan saja, maka untuk dapat membuka KLC2 ini tidak harus melalui laptop atau komputer. KLC2 juga dapat diakses melalui handphone. Untuk dapat mengikuti elearning/ micro learning di KLC, peserta harus mendaftar terlebih dahulu menggunakan Nomor KTP (untuk peserta non-Kemenkeu) atau NIPnya (untuk peserta Kemenkeu).

KLC2 merupakan aplikasi yang bisa diunduh dari google play store. Peserta dapat menginstal aplikasi ini pada gawainya. Hal ini memudahkan peserta untuk tetap dapat mengikuti pembelajaran di sela-sela aktivitas harian mereka. Berikut tampilan google play store dengan kata kunci klc Kemenkeu (Gambar 3).



Gambar 3. Tampilan Google Play Store dengan kata kunci Klc Kemenkeu
Sumber: data diolah (2023)

Setelah menemukan logo klc seperti tampilan di atas, maka klik di klc Kemenkeu. Tampilannya sebagai berikut (Gambar 4).



Gambar 4 Tampilan KLC Kemenkeu di google playstore
Sumber: data diolah (2023).

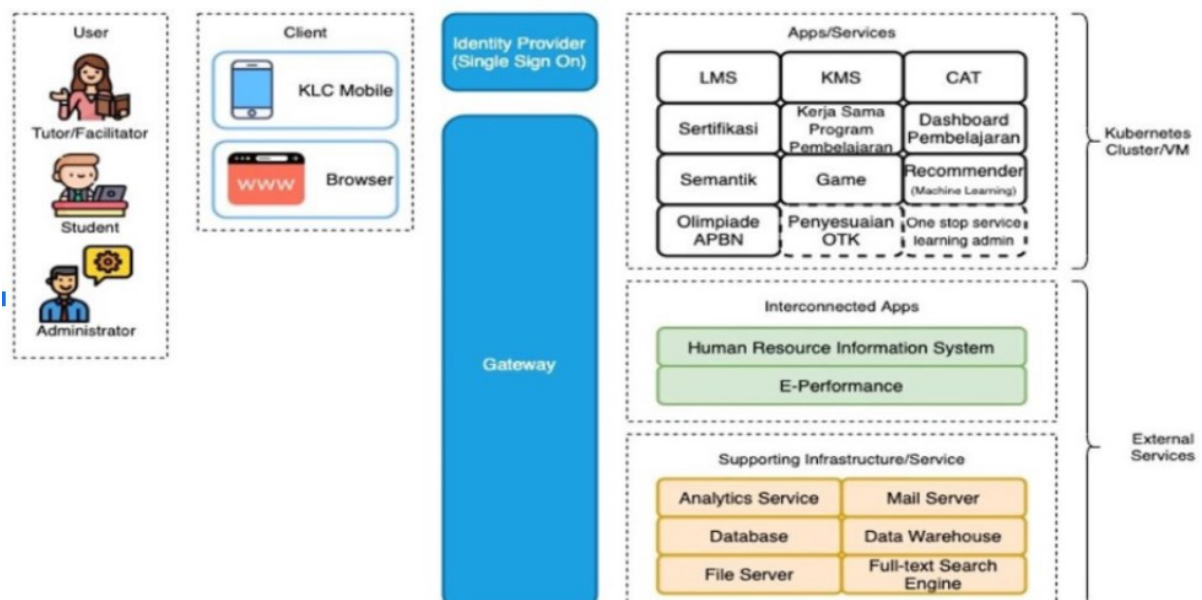
Implementasi pembelajaran menggunakan platform KLC ini berpedoman kepada aturan yang mendasarinya, misalnya PMK 226/PMK.011/2019 tentang manajemen pengetahuan di lingkungan Kementerian Keuangan, KMK 283/KMK.011/2021 tentang implementasi organisasi pembelajar (learning organization) di lingkungan Kementerian Keuangan, PER-2/PP/2019 tentang pedoman e-learning di lingkungan Kemenkeu, PER-8/PP/2018 tentang pedoman microlearning di lingkungan Kemenkeu, KEP

118/PP/2019 tentang standar pendokumentasian pengetahuan (knowledge capture) di lingkungan BPPK, dan Per-6/PP/2021 tentang pedoman pelatihan jarak jauh di lingkungan Kemenkeu.

KLC dibangun dengan beberapa alasan:

1. Selaras dengan strategi kemenkeu Corpu yaitu belajar kapan saja dan dimana saja, menyediakan platform untuk pengelolaan pembelajaran dan manajemen pengetahuan, dan mendukung demokratisasi pembelajaran (self learning)
2. memenuhi target pembelajaran digital sesuai dengan Inisiatif Strategis RBTK Kemenkeu. Target pembelajaran berbasis digital naik dari 30% pada tahun 2019 menjadi 70% pada tahun 2021.
3. Kapasitas BPPK untuk memenuhi pembelajaran klasikan pada tahun 2012 - 2016 masih sekitar 46-65% dari jumlah ASN Kemenkeu.
4. Arahan Menkeu untuk menjadikan Kemenkeu sebagai knowledge institution.

Pengembangan KLC dilakukan sejak tahun 2016 dimana pada saat itu pembelajaran klasikal diterapkan dengan penggunaan aplikasi manajemen diklat yaitu SemantikBPPK. pada tahun 2017 KLC Gen 1 selesai dikembangkan dan mulai operasional per 30 April 2017 dengan mengandalkan framework WPLLMS (Wordpress Learning Management System). Pada tahun 2019-2020 KLC Gen 2 selesai dikembangkan mulai operasional Februari 2020 dengan fitur utama klc2 (LMS) menggantikan KLC1. Pada saat itu mulai dikembangkan KMS, KLC mobile dan ujian online yang rilis 2021. KLC dibangun dengan arsitektur sebagai berikut (Gambar 5).



Gambar 5 Arsitektur KLC

Sumber: BPPK (2023)

Melalui KLC terdapat manfaat yang dirasakan. Aksesibilitas pembelajaran menjadi lebih mudah karena LMS dan KMS terdapat pada satu platform yang dapat diakses secara online. Terdapat penghematan biaya pelatihan sampai dengan 70% karena tidak membutuhkan biaya akomodasi dan transportasi. Layanan pembelajaran lebih variatif baik untuk pegawai Kemenkeu dan K/L lainnya. Sampai dengan tahun 2021, peserta pelatihan via KLC mencapai 672.088 Kemenkeu dan 123.978 non Kemenkeu (BPPK, 2023).

4. Analisis Sentimen

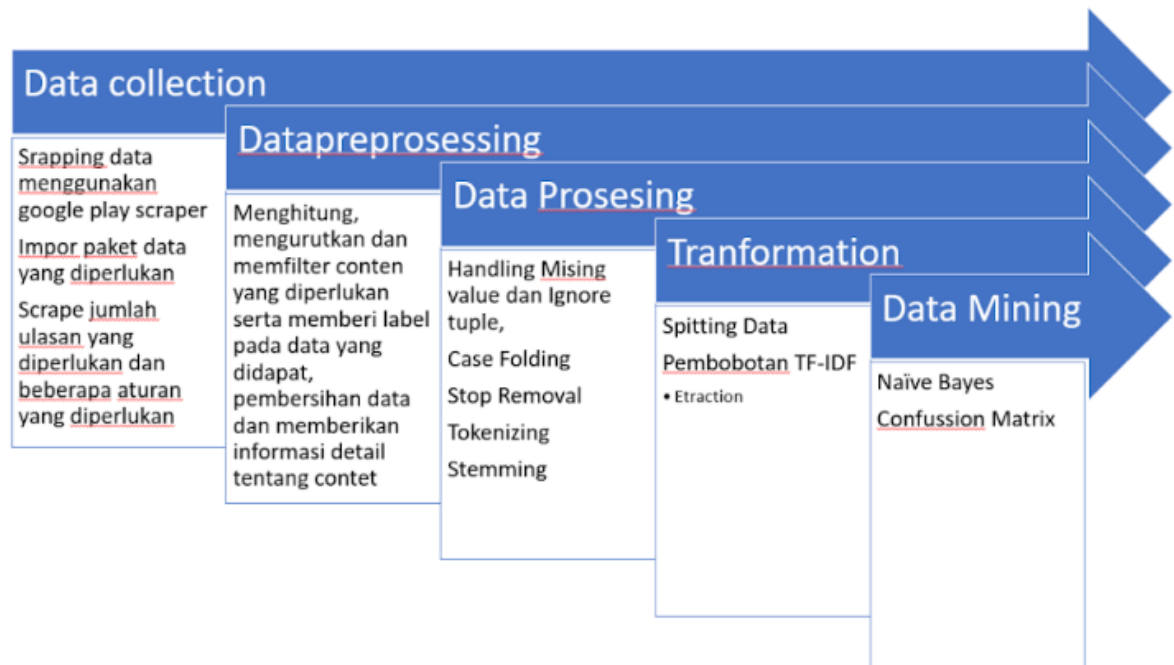
Analisis sentimen, juga dikenal sebagai opinion mining, adalah jenis penelitian yang mempelajari pendapat dan opini seseorang, sentimen, perilaku, dan emosi yang diungkapkan dalam teks [9]. Analisis sentimen dapat digunakan dalam berbagai konteks, seperti produk konsumen, layanan kesehatan, layanan keuangan, dan peristiwa sosial dan politik seperti pemilu. Penelitian tentang analisis sentimen biasanya berkonsentrasi pada pendapat yang menunjukkan atau menunjukkan perasaan yang positif atau negatif. Pendapat terkait dengan hampir semua aktivitas manusia karena mereka dapat mempengaruhi perilaku seseorang. Dalam kenyataannya, perusahaan dan lembaga selalu ingin mengetahui pendapat publik tentang suatu barang atau jasa.

Maulana (2023) menyitir bahwa analisis sentimen adalah metode penggalian informasi yang menganalisis pendapat, opini, sikap, penilaian, aksi, dan emosi seseorang terhadap suatu entitas, seperti masalah, layanan, produk, atau masalah tertentu. Perkembangan yang cepat pada opini di media sosial dan diskusi tentang ulasan produk atau layanan memungkinkan analisis sentimen ini menjadi penting. Analisis Sentimen atau Opinion Mining adalah studi komputasional dari opini – opini orang, sentimen dan emosi melalui entitas atau atribut yang dimiliki yang diekspresikan dalam bentuk teks. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral [10].

Di Google Play Store, penilaian aplikasi diikuti oleh ulasan pengguna. Ulasan ini menggambarkan pendapat pengguna tentang aplikasi dan membantu pengguna memutuskan untuk menggunakannya [11].

II. RESEARCH METHOD

Metode penelitian yang digunakan dalam penelitian ini adalah *Knowledge Discovery from Data* (KDD). KDD sendiri memiliki 5 tahapan, dengan tahapan didalamnya yakni Data Selection, Preprocessing, Transformation, Data Mining, Evaluation [12]. Agar dapat lebih mudah keterangan dari alur penelitian terdapat pada gambar 6.



Gambar 6 Alur penelitian
Sumber: data diolah (2023)

3.1. Data Collection

Data dikumpulkan dari google play store, mengenai ulasan pengguna aplikasi KLC.

3.2. Preprocessing

Tahap text preprocessing adalah tahap awal dari text mining. Text Preprocessing merupakan proses seleksi data yang akan diproses pada setiap dokumen. Tahap Preprocessing diperlukan untuk membersihkan data dari kata yang tidak terlalu perlu. Dalam tahap preprocessing pada penelitian ini terdiri dari cleaning, normalisasi dan pemberian label (labeling) dan informasi detail tentang content.

3.3. Data Prosesing

Text transformation merupakan tahap ke empat dalam text mining yang bertujuan untuk membuat data yang berupa text siap di proses data mining, tahap text transformation pada penelitian ini terdiri dari Handling Missing Value dan ignore tuple, Case folding, Stop Removal, Tokenizing, filtering, & Stemming.

3.4. Text Transformation

Tahap ini dengan data splitting data, memberikan pembobotan pada TF-IDF dan extraction

3.5. Data Mining

Naive Bayes Classifier Pada tahap klasifikasi menggunakan Algoritma Naive Bayes, dibagi menjadi dua proses yaitu proses training dan proses testing. Pertama dilakukan dulu proses training untuk pelatihan, kemudian dilakukan proses testing dengan mengacu pada probabilitas dari dataset training. Metode Naïve Bayes Classifier merupakan salah satu metode yang digunakan untuk melakukan pengelompokan suatu opini dengan baik. Mampu mengelompokkan komentar menjadi positif, netral, ataupun negatif terhadap suatu produk atau isu yang sedang berkembang di khalayak ramai.

3.5. Evaluasi

Tahap evaluasi adalah tahap yang dilakukan untuk melihat seberapa baik performansi algoritma klasifikasi yang digunakan dalam penelitian. Tolok ukur yang digunakan untuk mengukur performansi tersebut adalah accuracy, precision, recall, dan f-measure. Untuk memberikan gambaran kinerja yang lebih jelas, matriks konfusi memungkinkan analisis di empat dimensi:

1. Accuracy : Untuk mengetahui tingkat kemiripan antara nilai prediksi dengan nilai aktual.
2. Precision : Untuk mengetahui seberapa tepat atau akurat model dari yang diprediksi positif. Precision juga merupakan ukuran yang baik untuk menentukan ketika nilai False Positive pada suatu model tinggi.
3. Recall : Untuk menghitung berapa banyak dari nilai Actual Positive yang berhasil diidentifikasi oleh model melalui pelabelan sebagai True Positive. Recall juga akan menjadi metrik model yang digunakan untuk memilih model terbaik ketika ada nilai yang tinggi terkait dengan FalseNegative.
4. F-Measure : Perbandingan rata-rata nilai precision dan recall yang dibobotkan.

Naive Bayes Classifier (NBC)

Secara umum rumus dasar persamaan teorema Bayes adalah:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Gambar 7. Rumus Dasar Persamaan Teorema Bayes

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesis X merupakan suatu kelas spesifik

P(H|X) = Probabilitas H berdasarkan kondisi X

P(H) = Probabilitas pada hipotesis H (prior)

P(X|H) = Probabilitas X berdasarkan hipotesis H

P(X) = Probabilitas X (data sampel yang diamati)

Berdasarkan rumus tersebut, maka peluang masuknya sampel tertentu ke dalam kelas H adalah peluang munculnya kelas H (sebelum masuknya sampel tersebut), dikalikan dengan peluang kemunculan karakteristik sampel pada kelas H, dibagi dengan peluang kemunculan karakteristik sampel keseluruhan [11].

Confusion Matrix

Confusion matrix merupakan suatu matriks pengukuran performa yang berisi nilai perkiraan benar dan nilai perkiraan salah dimana keluarannya dapat berupa dua kelas atau lebih. Nilai recall, precision, accuracy dan f1-score akan menjadi evaluasi dalam pengujian keakuratan hasil pencarian nilai dalam penelitian. Struktur Confusion matrix akan menghasilkan matrix sebagai berikut:

	Actual Positive (0)	Actual Negative (0)
Predicted Positive (1)	TP	FP

Predicted Negative (0)	FN	TN
------------------------	----	----

dari tabel ini bisa dijelaskan 4 sisi nilai dari confusion matrix terdiri dari TP merupakan singkatan dari True Positive, yaitu angka perkiraan positif sesuai dengan nilai aktual benar. FP merupakan singkatan dari False Positive, yaitu angka perkiraan positif tidak sesuai dengan nilai aktual salah. FN merupakan singkatan dari False Negative, yaitu angka perkiraan negatif tidak sesuai dengan nilai aktual benar. TN merupakan singkatan dari True Negative, yaitu angka perkiraan negatif sesuai dengan nilai aktual salah.

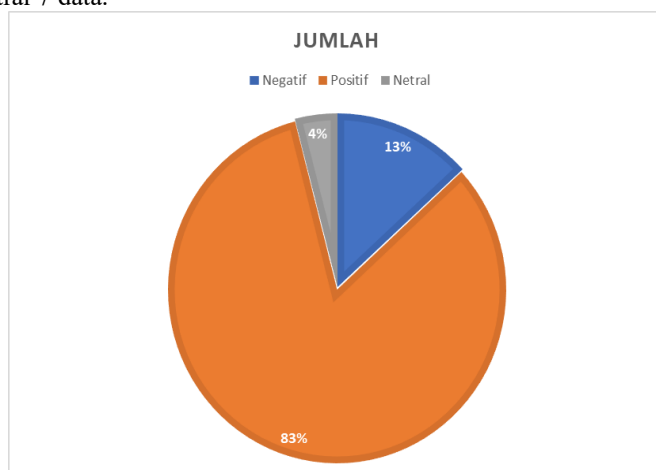
III. RESULTS AND DISCUSSION

1. Pengumpulan data

Pengambilan data pada penelitian kali ini menggunakan hasil scraping data. Data yang ingin diolah adalah rating dan ulasan dari pengguna aplikasi KLC2 di Google Play Store. Rating tersebut akan digunakan untuk memilah sentimen positif dan negatif. Data yang diambil untuk melakukan analisis sentimen yaitu pada rentang tanggal 1 Agustus 2021 sampai tanggal 1 Agustus 2023 dan dataset diambil dengan acak. Pengklasifikasian rating sebagai berikut yaitu rating 4 dan 5 sebagai sentimen positif, rating 1 dan 2 sebagai sentimen negatif dan rating 3 sebagai data netral. Hasil scraping yang diambil pada ulasan setelah melakukan proses pengolahan hasil data yang telah dihimpun dan dilakukan cleansing data dapat dilihat pada Tabel 1 berikut.

Sentimen	Jumlah
Negatif	23
Positif	145
Netral	7
Total	175

Dataset yang terkumpul 175 data. Kemudian data terbagi menjadi negatif 23 data, positif 145 data, netral 7 data.



2. Data preprocessing

Data Pre-processing merupakan tahap pertama dalam pemrosesan data supaya data dapat diolah dan data siap untuk proses penelitian salah satunya analisa sentimen. Proses data pre-processing dapat berupa cleaning, dan sorting

3. Data Processing

a. Casefolding

Casefolding merupakan tahapan yang mengubah suatu kata atau huruf pada ulasan menjadi huruf yang besar kecil menjadi huruf kecil semua. Pada implementasi casefolding untuk penelitian yang sekarang mengubah semua tulisan menjadi huruf kecil semua. Contoh proses casefolding dapat dilihat pada tabel berikut

	content	score	Label	text_clean
0	Ok sekali	5	Positif	ok sekali
1	Semakin mobile	5	Positif	semakin mobile
2	meningkatkan kompetensi asn	5	Positif	meningkatkan kompetensi asn
3	Tiap ganti mode full screen, video ngulang dar...	4	Positif	tiap ganti mode full screen video ngulang dari...
4	Luar biasa mantap	5	Positif	luar biasa mantap
6	top	5	Positif	top
7	Terimakasih sudah sangat membantu menambah pen...	5	Positif	terimakasih sudah sangat membantu menambah pen...
8	Sangat bermanfaat dan efektif untuk pembelajar...	5	Positif	sangat bermanfaat dan efektif untuk pembelajaran
9	Good	4	Positif	good
10	Mantap	5	Positif	mantap

b. Stop word Removal

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar #dan dianggap tidak memiliki makna. #Contoh stopwords dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan stopwords yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.

text_clean	text_StopWord
ok sekali	ok
semakin mobile	mobile
meningkatkan kompetensi asn	meningkatkan kompetensi asn
tiap ganti mode full screen video ngulang dari...	ganti mode full screen video ngulang diperbaiki
luar biasa mantap	mantap
top	top
terimakasih sudah sangat membantu menambah pen...	terimakasih membantu menambah pengetahuan paja...
sangat bermanfaat dan efektif untuk pembelajaran	bermanfaat efektif pembelajaran
good	good
mantap	mantap
perlu penelitian yang serius	penelitian serius

c. Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token diartikan sebagai “kata” meskipun tokenize juga dapat dilakukan pada paragraf maupun kalimat, setiap kata akan di pisah-pisah atau dilakukan separasi. Contoh proses tokenizing dapat dilihat pada Tabel

	content	score	Label	text_clean	text_StopWord	text_tokens
0	Ok sekali	5	Positif	ok sekali	ok	[ok]
1	Semakin mobile	5	Positif	semakin mobile	mobile	[mobile]
2	meningkatkan kompetensi asn	5	Positif	meningkatkan kompetensi asn	meningkatkan kompetensi asn	[meningkatkan, kompetensi, asn]
3	Tiap ganti mode full screen, video ngulang dar...	4	Positif	tiap ganti mode full screen video ngulang dari...	ganti mode full screen video ngulang diperbaiki	[ganti, mode, full, screen, video, ngulang, di...]
4	Luar biasa mantap	5	Positif	luar biasa mantap	mantap	[mantap]

d. Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata dan mengkonversikan kata-kata yang berimbuhan menjadi kata dasar atau menghilangkan imbuhan menjadi bentuk kata dasarnya. Untuk melakukan stemming bahasa Indonesia kita dapat menggunakan library Python Sastrawi yang sudah kita siapkan di awal. Library Sastrawi menerapkan Algoritma Nazief dan Adriani dalam melakukan stemming bahasa Indonesia. sebagai acuan ketentuan bahasa Indonesia, maka keluaran hasil stemming akan sesuai dengan ketentuan library sastrawi. Contoh proses stemming dapat dilihat pada Tabel berikut.

```

1 : ok : ok
2 : mobile : mobile
3 : meningkatkan : tingkat
4 : kompetensi : kompetensi
5 : asn : asn
6 : ganti : ganti
7 : mode : mode
8 : full : full
9 : screen : screen
10 : video : video
11 : ngulang : ngulang
12 : diperbaiki : baik
13 : mantap : mantap
14 : top : top
15 : terimakasih : terimakasih
16 : membantu : bantu
17 : menambah : tambah

```

4. Transfromation

a. Data Spliting

Proses data splitting bertujuan untuk membagi dataset menjadi dua bagian: data latih dan data uji. Data latih ditujukan untuk membangun model sentiment sedangkan data uji untuk mengevaluasi performansi model sentiment yang dihasilkan. Terdapat tiga skenario eksperimen berdasarkan splitting data yang dilakukan oleh penelitian sebelumnya, skenario 1 dengan komposisi 70% data latih, skenario 2 dengan komposisi 75% data latih sedangkan skenario 3 dengan komposisi 80% data latih dalam hal ini penulis mencoba semua skenario yang ada dan disesuaikan dengan data yang ada membagi data menjadi data training dan testing dengan `test_size = 0.30, 0.25 dan 0,20` dan random state nya menggunakan angka nol

b. DF TFF

Pembobotan kata dengan menggunakan vectorizing TF-IDF dilaksanakan dengan cara memberikan bobot pada setiap kata dari masing-masing dokumen/kalimat dengan keseringan kata yang muncul terbanyak akan ditampilkan terlebih dahulu yang sudah melalui tahap pre-processing.

Ada dua proses yang digunakan dalam tahapan pembobotan kata yaitu tahap Term Frequency (TF) dan tahap Inverse Document Frequency (IDF). Term Frequency (TF) merupakan suatu term dalam dokumen terkait yang dihitung frekuensi kemunculannya. Bobot kata akan semakin tinggi apabila banyak jumlah kata yang sering muncul (term) dalam suatu dokumen semakin banyak. Inverse Document Frequency (IDF) merupakan hubungan ketersediaan sebuah term dalam seluruh dokumen yang menunjukkan sebuah perhitungan. Besar nilai IDF akan semakin besar apabila besaran dokumen yang didalamnya terdapat suatu term tertentu semakin sedikit

5. Data mining

a. Naive Bayes

Pengolahan data pada penelitian ini akan menggunakan algoritma Naïve Bayes Classifier. Algoritma NBC adalah klasifikasi probabilitas sederhana dengan menggunakan model statistik untuk menghitung peluang kelas yang memiliki setiap kelompok dengan atribut yang tampil, dimana label kelas diambil dari beberapa himpunan berhingga. Metode Naïve Bayes merupakan metode yang terdiri atas beberapa keunggulan, seperti perhitungan sederhana, presisi yang tinggi, dan memiliki kecepatan apabila melakukan pemrosesan database yang banyak dan besar. Rumus Naïve Bayes sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Accuracy dapat memprediksi nilai dengan benar sehingga menggambarkan seberapa akurat dan tepat (machine learning). Dengan kata lain, merupakan perbandingan rasio kasus yang diidentifikasi benar (negative dan positive)

dengan semua data yang ada. Rumus accuracy adalah :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision mengevaluasi kemampuan tingkat ketelitian data antara data aktual dengan hasil prediksi dalam model untuk menemukan peringkat yang paling relevan. Rumus precision adalah :

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall untuk memaparkan keberhasilan model dalam mendapatkan kembali sebuah informasi. Sehingga recall untuk mengevaluasi kemampuan sistem untuk menemukan informasi yang sesuai dari semua data yang benar positif. Rumus recall adalah :

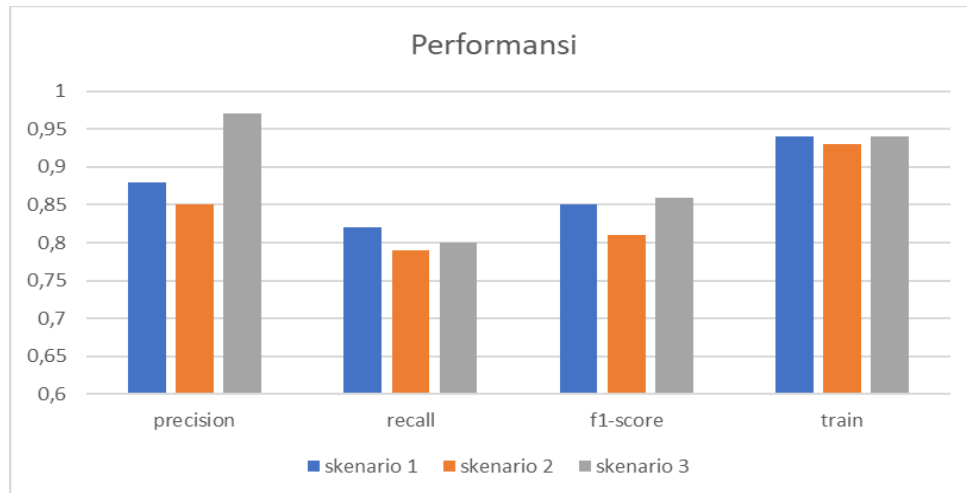
$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score adalah komparasi perhitungan rata-rata nilai dari recall dan precision yang telah di bobotkan.

$$\text{F1 Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Berdasarkan hasil uji yang didapatkan, pada Tabel dengan menggunakan skenario 1 menggunakan 70% data sebagai data train, dengan menggunakan algoritma Naïve Bayes Classifier menghasilkan hasil accuracy sebesar 94%, hasil precision sebesar 88%, hasil recall sebesar 82% dan hasil f1-score sebesar 85%. Dengan menggunakan Skenario 2 yaitu 75% data sebagai data train diperoleh hasil accuracy sebesar 93 %, hasil precision sebesar 85%, hasil recall sebesar 79% dan hasil f1-score sebesar 81%. Dengan menggunakan Skenario 3 yaitu 80 % data sebagai data train diperoleh hasil accuracy sebesar 94 %, hasil precision sebesar 97%, hasil recall sebesar 8% dan hasil f1-score sebesar 86%.

Rata rata performansi	skenario 1	skenario 2	skenario 3
accuracy	0,94	0,93	0,94
precision	0,88	0,85	0,97
recall	0,82	0,79	0,8
f1-score	0,85	0,81	0,86



b. Confusion Matrix

Confusion matrix merupakan suatu matriks pengukuran performa yang berisi nilai perkiraan benar dan nilai perkiraan salah dimana keluarannya dapat berupa dua kelas atau lebih. Nilai recall, precision, accuracy dan f1-score akan menjadi evaluasi dalam pengujian keakuratan hasil pencarian nilai dalam penelitian. Struktur Confusion matrix akan menghasilkan matrix sebagai berikut:

Skenario 1	Skenario 2	Skenario 3
$\begin{bmatrix} 4 & 2 \\ 1 & 44 \end{bmatrix}$	$\begin{bmatrix} 3 & 2 \\ 1 & 36 \end{bmatrix}$	$\begin{bmatrix} 3 & 2 \\ 0 & 29 \end{bmatrix}$

Hasil data tersebut, untuk skenario 1, dengan rincian sebanyak 145 ulasan positif diprediksi benar sebanyak 4 ulasan dan sebanyak 1 ulasan diprediksi salah. Kemudian, sebanyak 23 ulasan negatif diprediksi benar sebanyak 44 ulasan dan sebanyak 2 ulasan diprediksi salah. Untuk skenario 2, dengan rincian sebanyak 145 ulasan positif diprediksi benar sebanyak 3 ulasan dan sebanyak 1 ulasan diprediksi salah. Kemudian, sebanyak 23 ulasan negatif diprediksi benar sebanyak 36 ulasan dan sebanyak 2 ulasan diprediksi salah. Sedangkan untuk skenario 3, dengan rincian sebanyak 145 ulasan positif diprediksi benar sebanyak 3 ulasan dan sebanyak 0 ulasan diprediksi salah. Kemudian, sebanyak 23 ulasan negatif diprediksi benar sebanyak 29 ulasan dan sebanyak 2 ulasan diprediksi salah.

IV. CONCLUSION

Berdasarkan hasil penelitian diatas, dapat ditarik kesimpulan sebagai berikut.

1. Secara keseluruhan dengan menggunakan algoritma *Naïve Bayes* dalam mengklasifikasikan ulasan komentar perihal aplikasi KLC2 pada *Google Play Store*, klasifikasi label value *positive* dan *negative* tingkat akurasi mencapai 94%.
2. Terhadap aplikasi KLC2, ulasan positif sebanyak 83%, negative sebanyak 13%, dan 4% netral.

REFERENCES

- [1] Pemerintah Indonesia. Peraturan Presiden (PERPRES) Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik. (2018)
- [2] Kemenkeu. www.kemenkeu.go.id. (2023)
- [3] BPPK. Paparan BPPK. (2023)

-
- [4] Al Mamary. 2022. Understanding the use of learning management systems by undergraduate university students using the UTAUT model: Credible evidence from Saudi Arabia. *International Journal of Information Management Data Insights* 2 (2022)
- [5] Toring, Harold et al. 2023. Evaluation of students' satisfaction toward an adopted learning management system at Indiana Aerospace University: A structural equation modelling approach. *Asia Pacific Management Review* 28 (2023)
- [6] Alifiyanti, Intan Firda et al. 2018. Pemanfaatan Learning Management System (LMS) Berbasis Edmodo Materi Fluida Dinamis Untuk Peningkatan Minat Dan Prestasi Belajar Fisika Siswa Sekolah Menengah. *Prosiding SNFA (Seminar Nasional Fisika dan Aplikasinya) 2018*
- [7] Azzumar, Muhammad Fathi dan Nurmiati, Evy. 2021. Perancangan Knowledge Management System Pada E-Commerce Xyz Berbasis Mobile. *Just IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*. Vol. 11, No. 3, May 2021, pp. 47 – 54
- [8] Arribathi, Abdul Hamid. 2020. Efektifitas Penerapan Knowledge Management System (Kms) Berbasis Website Terhadap Pembelajaran Agama Islam (Studi Kasus Di Smk Al-Fattah Tangerang). *Jurnal Tahdzibi: Manajemen Pendidikan Islam* Volume 5 No. 1 Mei 2020
- [9] Feldman, R & Sanger, J. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York, 2007.
- [10] Maulana, Rihan et al. 2023. Analisis Sentimen Ulasan Aplikasi MyPertamina Pada Google Play Store Menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu* Vol. 9 No. 1 2023, 42-48
- [11] Nurian, Andriani dan Sari, Betha Nurina. 2023. Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* Vol.11 No.3
- [12] Permana, Muhamad Anton et al. 2023. Analisis Sentimen Penggunaan Aplikasi Video Conference Pada Ulasan Google Play Store Menggunakan Metode Nbc (Naive Bayes Classifier). *JURSISTEKNI (Jurnal Sistem Informasi dan Teknologi Informasi)* Vol 5, No.1, Januari 2023: Hal 178- 191