

Analisis Sentimen Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore menggunakan K-Nearest Neighbor dan Naïve Bayes

Digdaya Tualang Ksatria¹, Yogi Yunefri², Lucky Lhaura Van FC³, Sutejo⁴, Afriansyah⁵, Ilham Chairuddin⁶

^{1,2,3,4}Program Studi Teknik Informatika Universitas Lancang Kuning

⁵Program Studi Sistem Informasi Universitas Lancang Kuning

⁶Program Studi Sistem Informasi Sekolah Tinggi Teknologi Pekanbaru

^{1,2,3,4,5}Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

Jl. Dirgantara No 4 Arengka, Pekanbaru, Riau telp. 0761-61815

e-mail: ¹digdaya@unilak.ac.id, ²yogyunefri@unilak.ac.id, ³Lucky@unilak.ac.id,
⁴sutejo@unilak.ac.id, ⁵Afriansyah@unilak.ac.id, ⁶ilham.chairuddin@gmail.com

Abstrak

MyPertamina adalah program loyalty dan e-payment yang memberikan user experiences dari PT Pertamina (Persero) dengan mudah untuk seluruh pelanggan Pertamina. Dengan menggunakan aplikasi MyPertamina, pembayaran dapat dilakukan secara non-tunai melalui LinkAja!, sehingga lebih ringkas tanpa perlu mengantri menunggu kembalian dan lain sebagainya. Namun, dibalik berbagai fasilitas yang diberikan, masyarakat masih menemui beberapa kendala misalnya aplikasi MyPertamina yang sulit dibuka, banyaknya SPBU yang belum bisa memakai LinkAja!, pendaftaran akun yang sering kali gagal, transaksi yang sukses dengan MyPertamina namun poin nya tidak bertambah yang mengakibatkan kepercayaan pada MyPertamina berkurang dan konsumen kurang mendapatkan keuntungan dari adanya aplikasi tersebut. Dalam hal ini muncul berbagai ulasan di halaman ulasan Google Playstore pada aplikasi MyPertamina, beberapa pengguna kerap memberikan ulasan yang tidak sesuai dengan ekspektasinya, dan ada pula yang memberikan penilaian yang baik terhadap aplikasi tersebut. Melihat pentingnya ulasan pengguna terhadap keberlangsungan suatu aplikasi, maka akan baik apabila analisis sentimen dilakukan pada data ulasan aplikasi di Google Playstore dengan menggunakan metode K-Nearest Neighbor dan Naïve Bayes. Pada penelitian ini menghasilkan sentimen positif dan negatif pada masing-masing kedua metode. Untuk metode K-Nearest Neighbor terdapat sentimen positif dengan total 117 ulasan dan sentimen negatif dengan total 1099 ulasan berdasarkan dari analisis klasifikasi yang telah dilakukan. Sementara untuk metode K-Nearest Neighbor terdapat sentimen positif dengan total 84 ulasan dan sentimen negatif dengan total 1132 ulasan berdasarkan dari analisis klasifikasi yang telah dilakukan. Dari kedua metode didapat hasil dari masing-masing nilai akurasi sebesar 70,73% untuk Naïve Bayes dan sebesar 85,97% untuk K-Nearest Neighbor. Hal ini menunjukkan bahwa tingkat akurasi dari metode K-Nearest Neighbor lebih baik dalam melakukan klasifikasi teks ulasan aplikasi MyPertamina jika dibandingkan dengan tingkat akurasi yang dihasilkan dari metode Naïve Bayes.

Kata Kunci: Data Mining, Klasifikasi, Naïve Bayes, K-Nearest Neighbor, Google Colaboratory

Abstract

MyPertamina is a loyalty and e-payment program that provides user experiences from PT Pertamina (Persero) easily for all Pertamina customers. By using the MyPertamina application, payments can be made non-cash via LinkAja!, so it is more concise without having to queue for change and so on. However, despite the various facilities provided, people still encounter several obstacles, for example the MyPertamina application which is difficult to open, many gas stations that cannot use LinkAja!, account registration which often fails, successful transactions with MyPertamina but the points do not increase which results in trust in MyPertamina is decreasing and consumers are getting less benefit from this application. In this case, various reviews appear on the Google Playstore review page for the MyPertamina application, some users often provide reviews that do not match their expectations, and there are also those who give a good assessment of the application. Seeing the importance of user reviews on the sustainability of an application, it would be good if sentiment analysis was carried out on application review data on Google Playstore using the K-Nearest Neighbor and Naïve Bayes methods. This research produces positive and negative sentiments for each of the two methods. For the K-Nearest Neighbor method, there is positive sentiment with a total of 117 reviews and negative sentiment with a total of 1099 reviews based on the classification analysis that has been carried out. Meanwhile, for the Naïve Bayes method, there is positive sentiment with a total of 84 reviews and negative sentiment with a total of 1132 reviews based on the classification analysis that has been carried out. From the two methods, the results obtained for each accuracy value were 70.73% for Naïve Bayes and 85.97% for K-Nearest Neighbor. This shows that the level of accuracy of the K- Nearest Neighbor method is better in classifying text reviews of the MyPertamina application when compared to the level of accuracy resulting from the Naïve Bayes method.

Keywords: Data Mining, Classification, Naïve Bayes, K-Nearest Neighbor, Google Colaboratory

1. PENDAHULUAN

Kebutuhan akan kecepatan dan kemudahan pemrosesan data dan informasi membuat berbagai perusahaan maupun instansi berlomba-lomba untuk menciptakan inovasi baru. Dengan semakin pesatnya perkembangan teknologi informasi dapat mendukung kemudahan realisasi dari inovasi tersebut. Inovasi-inovasi tersebut tersebar pada seluruh lini kehidupan, baik transportasi, pendidikan, perdagangan, hingga menyentuh sektor pemerintahan[1].

MyPertamina adalah program *loyalty* dan *e-payment* yang memberikan *user experiences* dari PT Pertamina (Persero) dengan mudah untuk seluruh pelanggan Pertamina. Dengan menggunakan aplikasi MyPertamina, pembayaran dapat dilakukan secara non-tunai melalui LinkAja!, sehingga lebih ringkas tanpa perlu mengantri menunggu kembalian dan lain sebagainya. Namun, dibalik berbagai fasilitas yang diberikan, masyarakat masih menemui beberapa kendala misalnya aplikasi MyPertamina yang sulit dibuka, banyaknya SPBU yang belum bisa memakai LinkAja!, pendaftaran akun yang sering kali gagal, transaksi yang sukses dengan MyPertamina namun poin nya tidak bertambah yang mengakibatkan kepercayaan pada MyPertamina berkurang dan konsumen kurang mendapatkan keuntungan dari adanya aplikasi tersebut[2]. Dalam hal ini muncul berbagai ulasan di halaman ulasan Google Playstore pada aplikasi MyPertamina, beberapa pengguna kerap memberikan ulasan yang tidak sesuai dengan ekspektasinya, dan ada pula yang memberikan penilaian yang baik terhadap aplikasi tersebut.

Melihat pentingnya ulasan pengguna terhadap keberlangsungan suatu aplikasi, maka akan baik apabila analisis sentimen dilakukan pada data ulasan aplikasi di Google Playstore. Analisis sentimen dilakukan sebagai suatu proses untuk menghasilkan pengkategorian pendapat melalui ulasan yang diberikan pada suatu topik menjadi berlabel positif, negatif, atau netral. Analisis sentimen merupakan bagian dari penelitian *text mining* yang memiliki potensi nilai komersial[1].

Penelitian akan berfokus pada pengukuran sentimen pengguna aplikasi dari ulasan yang diunggah oleh pengguna pada aplikasi MyPertamina yang kemudian akan diproses dan diolah untuk dapat menghasilkan kelas tingkatan kepuasan pengguna MyPertamina. Kemudian data diolah berdasarkan kelasnya (Positif dan Negatif) untuk melihat faktor-faktor umum yang dibicarakan pelanggan yang mempengaruhi tingkatan kepuasan tersebut. Penelitian ini juga akan melakukan analisis berdasarkan hasil-hasil yang didapatkan untuk dapat menjadi saran kepada pihak instansi Pertamina dalam upaya peningkatan dan perbaikan kualitas. Oleh karena itu, penulis mengangkat penelitian dengan judul “Analisis Sentimen Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan *K-Nearest Neighbor* Dan *Naïve Bayes*” dengan melakukan ekstraksi data ulasan pengguna dari Google Playstore dan melakukan *preprocessing* untuk mengetahui nilai positif dan negatif dengan model *Wordcloud* pada *Tools* GoogleColaboratory. *Google Colaboratory* atau disebut juga *Colab* adalah tools baru yang dikeluarkan oleh yang dibuat oleh *Google Internal Research*. Tools ini membantu para Researcher dalam mengolah data untuk keperluan belajar pada pengolahan data menggunakan *Machine Learning*[3].

Beberapa penelitian terdahulu tentang analisis sentimen antara lain yaitu, Penerapan *Naïve Bayes Classifier*, *K-Nearest Neighbor (KNN)* dan *Decision Tree* untuk

Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah [4], Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma *Naïve Bayes*[1], Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi JNE Dan J&T Express Menggunakan Metode *Naïve Bayes*[5], Analisa Sentimen Terhadap Review Fintech Dengan Metode *Naïve*

Bayes Classifier Dan *K- Nearest Neighbor* [6], Analisis Sentimen Aplikasi Transportasi Online KRL Access Menggunakan Metode *Naive Bayes* [7], Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, *Random Forest* Dan *Support Vector Machine* [8], Analisis Sentimen Aplikasi Pembelajaran Online Di Play Store Pada

Masa Pandemi Covid-19 Menggunakan Algoritma *Support Vector Machine (SVM)* [9], Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode *Naïve Bayes Classifier* pada Situs *Google Play Store* [10], Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access

Di *Google PlayStore* Menggunakan Metode *Multinomial Naive Bayes* [11], dan Analisa Sentimen Pada *Twitter* Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode *Naïve Bayes* [12].

Analisis Sentimen adalah sebuah metode yang digunakan untuk memahami, dan mengolah data teks guna mendapatkan informasi negatif, positif, dan netral yang ada pada kumpulan teks [13]. Analisis sentimen sering juga disebut sebagai *opinion mining*. Saat ini, pengguna sangat senang mengekspresikan perasaannya melalui *platform online*, seperti

media sosial, *e-commerce*, dan website [14]. Tugas dasar dalam melakukan analisis sentimen yaitu mengelompokkan teks yang terdapat dalam sebuah kalimat dan menentukan pendapat tersebut apakah positif, negatif, atau netral[5] Adapun Tahapannya adalah sebagai berikut [15]:

- a. Definisikan domain dataset yaitu mengumpulkan dataset seperti *review* agen travel, *review* restoran, *review* produk dan lain-lain.
- b. *Pre-processing*, pada tahapan ini biasanya dilakukan *Tokenization*, *Stopwords Removal*, dan *Stemming* [16].
- c. *Transformation* yaitu pembobotan dari data tekstual, proses yang sering digunakan adalah *TF-IDF*. Berikut masing-masing perhitungan *TF* dan *IDF* [17]:

$$TF = \begin{cases} 1 + \log_{10}(tf_{t,d}), & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \quad (1)$$

$$IDF = \log\left(\frac{N}{df_t}\right)$$

Keterangan :

$Tf_{t,d}$: jumlah kemunculan term (t) pada dokumen (d), jika tidak ada term atau $t=0$, maka *TF* menjadi 0.

N : jumlah dokumen teks.

df_t : jumlah dokumen yang mengandung term (t).

- d. Feature Selection yaitu membuat pengklasifikasi lebih efisien dengan mengurangi jumlah data yang dianalisa.
- e. Classification yaitu pengklasifikasi teks biasanya menggunakan metode Naïve Bayes, K-Nearest Neighbor, SVM dan lain-lain.
- f. Interpretation/Evaluation yaitu biasanya evaluasi untuk menghitung nilai akurasi dan nilai AUC.

NBC adalah salah satu metode untuk mengklasifikasi data dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan karakter independen yang tinggi. Metode ini sesuai untuk banyak dataset dengan performa yang cepat dalam mengklasifikasi data dan memiliki akurasi tinggi [18]. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes [19]. Berikut persamaan dari teorema bayes: [12]:

$$P(c|X) = \frac{P(X|c) \cdot P(c)}{P(X)} \quad (2)$$

Keterangan :

$P(c|X)$: Probabilitas hipotesis berdasar kondisi (*posteriori probability*)

$P(c)$: Probabilitas hipotesis (*prior probability*)

$P(X|c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$: Probabilitas c

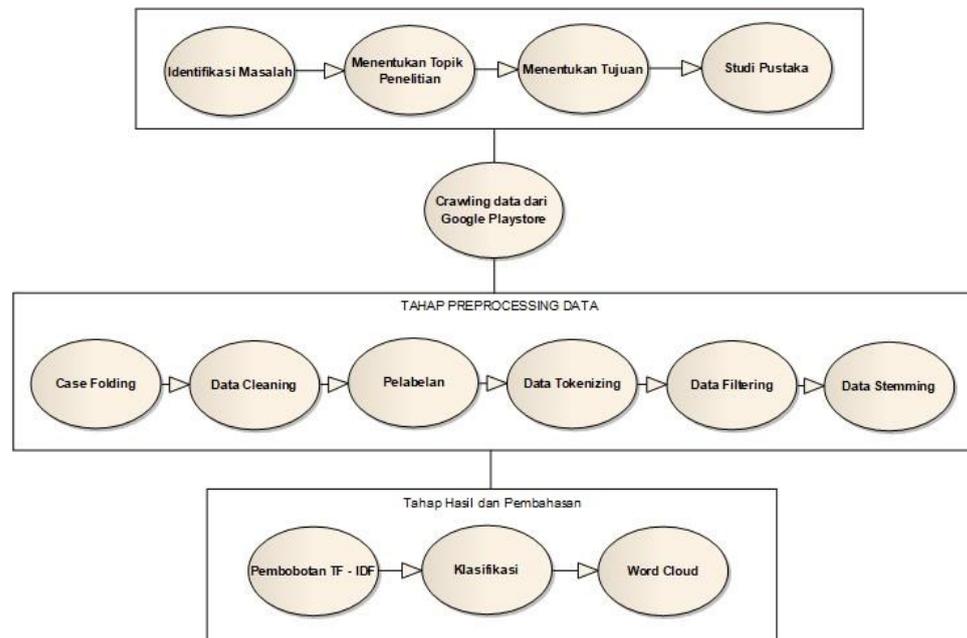
Algoritma *K-Nearest Neighbor* adalah suatu algoritma supervised learning dimana hasil dari instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sample-sample dari training data [14]. Tidak ada pemodelan apapun dalam klasifikasi ini dan berdasar pada memori yang menggunakan jumlah terbanyak diantara klasifikasi dari nilai K sebagai prediksi dari titik yang terbaru dekatnya jarak titik dengan tetangganya bisa dihitung dengan menggunakan eucliden distance [20]. Untuk mendefinisikan jarak antara dua titik pada data training (x) dan titik pada data testing (y) maka digunakan rumus euclidean, seperti yang ditunjukkan pada persamaan berikut [21]:

$$d(x, y) = \sqrt{\sum_{i=1}^r (x_{ik} - y_{ik})^2} \quad (3)$$

Dengan d adalah jarak euclidean antara titik pada data training (x) dan data testing (y) yang akan diklasifikasi, dimana adalah nilai ke-i variabel ke-k dari x, adalah nilai ke-i variabel ke-k dari y, dan r adalah jumlah variabel, diaman konteks klasifikasi data, "d" (jarak Euclidean) dihitung antara dua titik, yaitu titik pada data training (x) dan titik pada data testing (y) yang akan diklasifikasi. Dalam rumus jarak Euclidean tersebut, "i" merujuk pada variabel ke-i, "k" merujuk pada dimensi atau variabel ke-k dari data, "x" adalah nilai variabel ke-i variabel ke-k dari data training, "y" adalah nilai variabel ke-i variabel ke-k dari data testing, dan "r" adalah jumlah variabel atau dimensi total dari data. Dengan menggunakan rumus ini, kita dapat mengukur seberapa dekat atau jauh kedua titik tersebut dalam ruang variabel yang dimiliki.

2. METODE PENELITIAN

Adapun beberapa tahapan pada penelitian ini adalah sebagai berikut :



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar diatas, berikut adalah penjabaran urutan-urutan tahapan penelitiannya :

1. Tahap Perencanaan

Pada tahap ini peneliti menjelaskan permasalahan sebagai latar belakang penelitian dan tujuan yang akan dicapai. Permasalahan dalam penelitian ini berasal dari studi literatur yang telah dipelajari oleh peneliti sehingga mendapatkan informasi dan data yang dibutuhkan.

a. Identifikasi Masalah

Sebelum melakukan penelitian, penulis terlebih dahulu mengidentifikasi permasalahan yang terjadi mengenai seputar topik penelitian yang akan dilakukan.

b. Menentukan Topik Penelitian

Untuk topik penelitian penulis mengamati dan mengidentifikasi permasalahan yang terdapat pada ulasan-ulasan yang diberikan oleh pengguna Playstore pada aplikasi MyPertamina. Subjek penelitian ini adalah ulasan dari individu sebagai pengguna akun dalam menyatakan opini dan pendapatnya terhadap bagaimana aplikasi MyPertamina. Dengan demikian penulis mengangkat topik penelitian tentang Analisis Sentimen Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan *K-Nearest Neighbor* Dan *Naïve Bayes*.

c. Menentukan Tujuan Penelitian

Untuk mendukung tercapainya sasaran penelitian ini, tahapan selanjutnya adalah penentuan tujuan dari penelitian ini dengan maksud agar tujuan dalam penulisan lebih fokus. Tujuan dari penelitian ini adalah untuk menganalisis ulasan pengguna aplikasi MyPertamina dan juga membuat visualisasi kata menggunakan *WordCloud*.

d. Studi Pustaka

Studi pustaka adalah melakukan pengumpulan data sebagai landasan teori yang kokoh untuk melakukan penyusunan laporan dan menganalisa permasalahan.

Studi pustaka yaitu dengan cara mengumpulkan informasi dari beberapa jurnal nasional maupun internasional, buku-buku dan website resmi yang berhubungan dengan topik yang diteliti untuk mendukung penelitian ini.

2. *Crawling* Data Dari Google Playstore

Teknik pengumpulan data ulasan diambil dengan cara *crawling* data melalui Google Playstore pada ulasan aplikasi MyPertamina. Waktu penelitian yang dilakukan penulis adalah dari 4 November hingga 12 November 2022. Proses penarikan data dilakukan dengan menggunakan aplikasi tambahan dari *browser Google Chrome* yaitu *Instant Data Scraper*. Dalam tahap ini dilakukan analisa terhadap data yang akan digunakan pada penelitian ini seperti kategori data yang akan dipakai.

a. Kategori Data

Kategori data yang dipakai pada penelitian ini menjadi 2 kategori kelas ulasan yaitu kategori ulasan positif dan negatif. Berikut ini merupakan penjelasan kategori data yang dibutuhkan pada penelitian ini yaitu sebagai berikut :

1) Positif

Merupakan sebuah kalimat apabila kalimat tersebut mengandung arti positif terhadap pembaca atau akan memungkinkan menggiring pembaca membentuk pendapat positif terhadap kalimat tertentu seperti, pujian, informasi yang berguna, kalimat dukungan, kabar yang menyenangkan, keberhasilan, prestasi, terobosan baru, kejadian penting dan keindahan.

2) Negatif

Merupakan sebuah kalimat apabila kalimat tersebut mengandung arti negatif terhadap pembaca atau akan memungkinkan menggiring pembaca membentuk pendapat negatif terhadap kalimat tertentu seperti, ditandai adanya sindiran, perkataan kotor, bantahan, celaan, atau kalimat kekecewaan masyarakat terhadap informasi.

b. Pengumpulan Data

Proses pengambilan data dari Google Playstore dilakukan dengan cara menggunakan teknik *crawling*, yaitu proses penyalinan data *review* dari *website* terkait. Pada tahapan pengumpulan data hasil *crawling* merupakan data yang akan diolah dengan kebutuhan penelitian yang akan digunakan dan data disimpan dengan format *.xls*.

3. Tahap Preprocessing Data

Pada tahap *preprocessing* data terdapat *text mining* yaitu untuk menganalisa teks yang berguna untuk mencari informasi bermanfaat dengan tujuan tertentu. Penambahan teks sebagai proses dua tahap yang diawali dengan penerapan terstruktur terhadap sumber data teks dan yang kedua dengan ekstraksi informasi dan pengetahuan yang relevan. *Preprocessing* yang digunakan dalam penelitian ini adalah *Tokenizing*, *Filtering* dan *Stemming*.

a. *Data Tokenizing*

Pada tahapan *tokenizing* dilakukan dengan memisahkan setiap teks menjadi setiap potongan kalimat dan kata berdasarkan spasi yang ditemukan. Berfungsi untuk menghilangkan tanda baca atau simbol dan membuat token pada dokumen.

b. *Data Filtering*

Pada tahapan *filtering* dilakukan dengan membuang kata-kata yang tidak penting atau tidak bermakna secara signifikan dari hasil token. Dan kemudian dilakukan penghapusan tanda baca dan *stopword*. *Stopword* adalah pembagian

bahasa yang alami. Ada beberapa kata yang harus dihapus dari sebuah teks karena membuat teks menjadi lebih berat dan kurang penting untuk dianalisis sebab tidak bisa diukur sebagai kata kunci.

c. *Data Stemming*

Pada tahapan stemming dilakukan dengan mengubah kata berimbuhan yang terdapat dalam dokumen menjadi kata dasar seperti mengkritik yang kata dasarnya yaitu kritik dan banyak contoh lainnya.

4. Tahap Hasil dan Pembahasan

Setelah data dikumpulkan tahapan selanjutnya adalah tahap hasil dan pembahasan. Adapun tahapan yang dilakukan adalah :

a. Pembobotan TF-IDF

Pembobotan ini yaitu melakukan pemecahan kalimat menjadi beberapa kata serta memberikan bobot pada setiap kata dengan menggunakan TF-IDF. Tahap ini merupakan *statistic numeric* yang digunakan untuk mengungkapkan tingkat kepentingan sebuah kata untuk dokumen dalam suatu koleksi.

b. Klasifikasi

Pada tahap ini dilakukan pengklasifikasian teks yang sudah melalui tahap-tahap *preprocessing* sebelumnya. Proses pemodelan klasifikasi diawali dengan membagi data sesuai dengan kategori dari teks tersebut menjadi bagian untuk data latih (*training*) dan data uji (*testing*). Metode yang digunakan adalah *Naïve Bayes* dan *K-Nearest Neighbor* (KNN).

c. Hasil Analisis

Dalam tahapan ini hasil dari klasifikasi berupa pengelompokan ulasan sentimen pengguna aplikasi *berdasarkan* frekuensi kata yang muncul sesuai dengan kelas masing-masing yaitu positif dan negatif.

d. Visualisasi Data

Tahapan ini yaitu akan membuat visualisasi kata-kata dataset berdasarkan frekuensi kata yang muncul sesuai dengan kelas masing-masing yaitu positif dan negatif dengan menggunakan *word cloud*. Hasil dari *word cloud* ini akan dianalisis dengan cara membandingkan frekuensi kata dan frekuensi kelas yang telah didapatkan dengan realita yang terjadi dilapangan.

3. HASIL DAN PEMBAHASAN

3.1 Analisis dan Pengumpulan Data

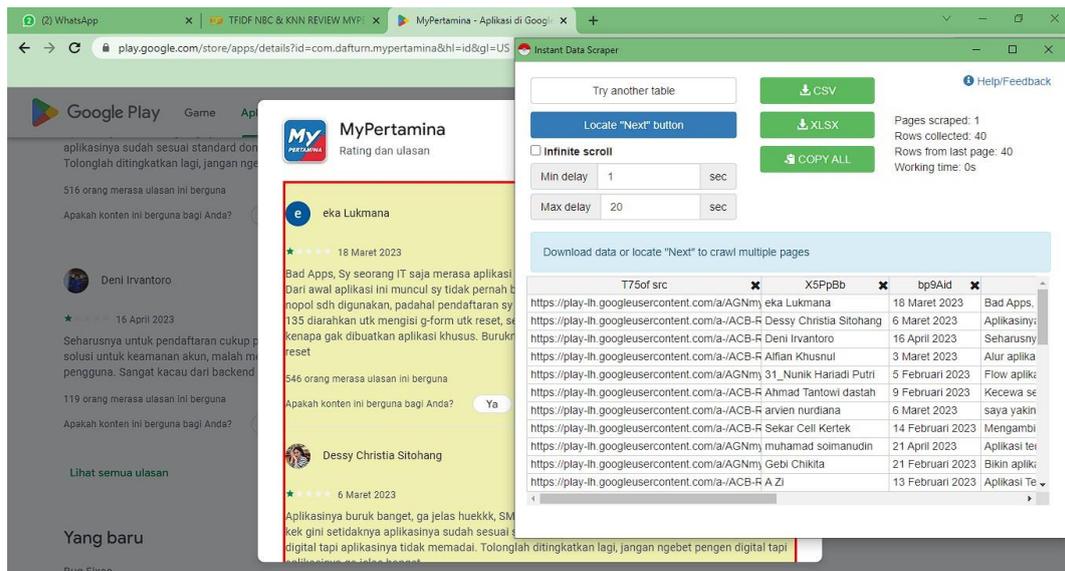
Analisis sentimen terhadap opini atau pendapat dari pengguna aplikasi MyPertamina yang diunduh melalui Google Play Store mengenai fitur-fitur dan mekanisme penggunaan dan pembayaran yang disediakan oleh aplikasi dengan metode *Naïve Bayes* dan *K-Nearest Neighbor*. Analisis sentimen berdasarkan data ulasan dari para pengguna aplikasi MyPertamina yang dilakukan dengan cara menarik data sesuai kebutuhan dari sebuah website Google Play Store dengan cara menyalin data tersebut, adapun *tools* yang penulis gunakan adalah *Extensions* pada *Browser Google Chrome* yaitu *Instant Data Scrapper*.



Gambar 2. *Instant Data Scrapper*

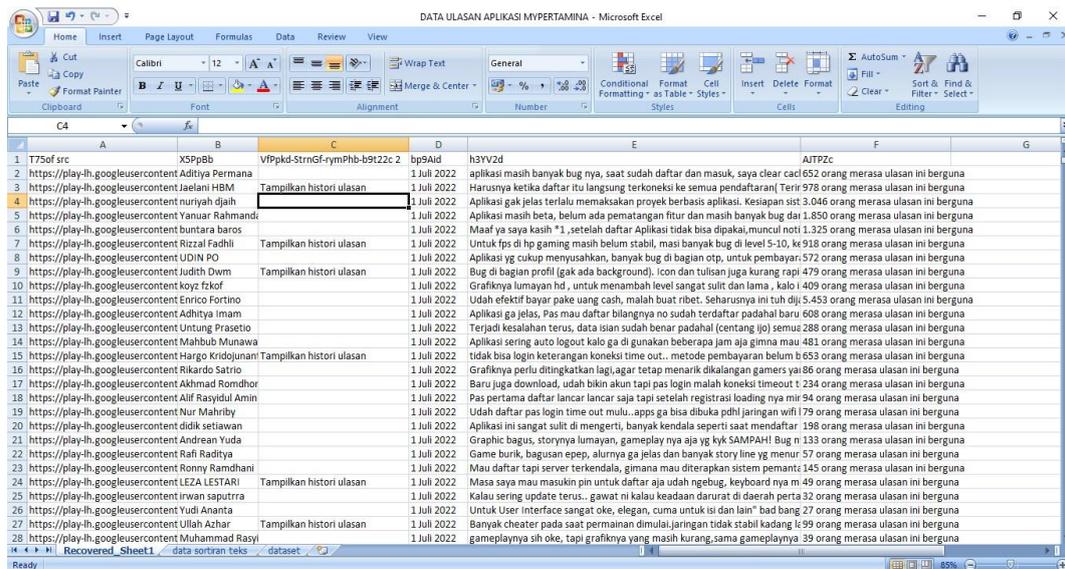
Gambar diatas merupakan aplikasi *Instant Data Scrapper* yang digunakan untuk menarik data dari Google Play Store. Aplikasi ini dapat menarik data dari sebuah website dan menyimpannya ke dalam file Ms.Excel.

220 ■ Digdaya Tualang Ksatria, Analisis Sentimen Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan K- Nearest Neighbor Dan Naive Bayes



Gambar 3. Data Ulasan Yang Akan Ditarik

Pada gambar diatas adalah proses penarikan data dengan menggunakan *InstantData Scrapper*. Pada gambar diatas contoh data ulasan pada yang akan ditarik data teksnya.



Gambar 4. Hasil Data Teks Yang Sudah Ditarik Dalam Format Ms.Excel

Total 1400 data teks diatas didapat dengan menggunakan aplikasi *Instant Data Scrapper* (Gambar 2) dan menarik data pada Google Play Store (Gambar 3).

3.2 Praproses Data

Praproses yang dilakukan pada tahap awal ini merupakan pembersihan data secara manual. Data yang telah ditarik sebelumnya selanjutnya dilakukan pemberian kode ulasan, penyortiran data serta pemberian kelas pada data training yaitu data urutan U0001 sampai dengan U1200, sementara data dari urutan U1201 sampai dengan U1400 dijadikan sebagai data testing.

1. Sortir Data

Penyortiran data dilakukan dengan tujuan untuk membuang data teks yang tidak digunakan atau tidak berhubungan dengan pembahasan dari aplikasi MyPertamina. Terdapat beberapa data teks ulasan yang tidak sesuai dengan pembahasan pada data yang di tarik, sebagai contoh data yang tidak digunakan dapat dilihat pada tabel berikut:

Tabel 1. Sortir Data

Ulasan	Text
U001	aplikasi masih banyak bug nya, saat sudah daftar dan masuk, saya clear cache malah di suruh daftar / login lagi. aplikasi juga sangat berat, padahal jaringan internet lancar, terlalu lama dalam merespon data nya, apa spek server nya kurang mumpuni? keliatan sisi developer nya masih pemula, atau mungkin teknologi yang digunakan masih kuno tertinggal jauh dengan teknologi yang sedang trend / sedang banyak digunakan saat ini
U002	Harusnya ketika daftar itu langsung terkoneksi ke semua pendaftaran(Terintegrasi), user tinggal pilih opsi. Ini malah tiap menu harus isi nomor KTP, Tgl Lahir Dan lain lain berulang ulang. Bukan memudahkan dalam penggunaan aplikasi, malah jadi menyulitkan dan membebani pengguna. Tolong Disederhanakan cara input registrasinya, jangan berulang ulang Jaelani /Bandung
U003	Aplikasi gak jelas terlalu memaksakan proyek berbasis aplikasi. Kesiapan sistem sangat buruk, time connection, tidak terdaftar, aplikasi lemot dan keluar sendiri. Sistem pembayaran juga tidak fleksibel, bukan mempermudah tetapi semakin mempersulit masyarakat. Bagaimana nasib yang tidak mempunyai HP sedangkan membutuhkan bensin buat mencari nafkah? Semoga bisa refleksi dan evaluasi agar tidak terkesan maksa, dipikir ulang kalo mau buat suatu sistem dikira gampang kali
U004	Aplikasi masih beta, belum ada pematangan fitur dan masih banyak bug dan eror di dalamnya. Untuk grafis dan alur storynya cukup bagus sayang untuk meningkatkan achivement sangat susah dan untuk naikin level juga susah. Thanks Netmarbel saya jadi dapat kartu S+
U005	Maaf ya saya kasih *1 ,setelah daftar Aplikasi tidak bisa dipakai,muncul notif sesi habis sesi habis terus. Setiap masuk selalu muncul notif tersebut,sudah dicoba pakai wifi - paket data tetep saja gak bisa. Tolong diperbaiki dulu aplikasinya .
U006	Untuk fps di hp gaming masih belum stabil, masi banyak bug di level 5-10, kestabilan koneksi juga masih jelek padahal mb/s stabil aman di game lain, Loading masih suka macet trus force close. Semoga cepat diperbaiki ya terimakasih
U007	Aplikasi yg cukup menyusahkan, banyak bug di bagian otp, untuk pembayaran msh blm flexible cuma bisa pke link aja, cukup meribetkan pengendara karena tidak semua punya handphone dan semua bisa pake aplikasi ini, blm lagi kalau kuota habis/hpnya lemot jadi kurang efektif ketika terjadi antre panjang, kurangnya sosialisasi sehingga masyarakat blm siap untuk menerima kebijakan ini.
U008	terjadi kesalahan, tapi pas back, malah ngelooop. Ini kerjaan developernya masih banyak. Kayak aplikasi beta. UI UX masih perlu banyak improve, padahal ini sekelas BUMN. Kedepannya perlu banyak inovasi, termasuk metode pembayaran yang jangan cuma pake linkAja yang penggunaanya jarang.. dan yang paling ganggu adalah, loginnya yang pake timeout. Jarang dipake jadi logout sendiri
U009	terjadi kesalahan, tapi pas back, malah ngelooop. Ini kerjaan developernya masih banyak. Kayak aplikasi beta. UI UX masih perlu banyak improve, padahal ini sekelas BUMN. Kedepannya perlu banyak inovasi, termasuk metode pembayaran yang jangan cuma pake linkAja yang penggunaanya jarang.. dan yang paling ganggu adalah, loginnya yang pake timeout. Jarang dipake jadi logout sendiri
U010	Udah efektif bayar pake uang cash, malah buat ribet. Seharusnya ini tuh dijadikan sebuah opsi bukan kewajiban . Hadeh, belum lagi masalah bug dalam aplikasi kadang force close dan pas dibuka aplikasi nya putih tampilan menu munculnya lama. Emang dah kalo ada yang susah kenapa pilih yg mudah wkwkwkwk

Pada tabel diatas dapat dilihat pada data dengan urutan U0004, U0006 dan U0009 merupakan data teks yang tidak sesuai dengan pembahasan mengenai ulasan dari aplikasi MyPertamina sehingga data-data teks yang berwarna kuning perlu dihapuskan dan tidak perlu digunakan.

2. Pemberian Kelas Pada Dataset (Data Training)

Pemberian kelas yang dimaksud adalah melakukan klasifikasi secara manual kepada data *training* (data teks U0001 – U1200). Dalam hal ini peneliti melihat dan membaca satu per satu teks dari data training dan mengambil keputusan bahwasanya data-data teks tersebut bernilai positif ataupun negatif. Adapun tampilan datanya adalah sebagai berikut:

Tabel 2. Dataset

Ulasan	Kelas	Text
U001	Negatif	aplikasi masih banyak bug nya, saat sudah daftar dan masuk, saya clear cache malah di suruh daftar / login lagi. aplikasi juga sangat berat, padahal jaringan internet lancar, terlalu lama dalam
U002	Negatif	Harusnya ketika daftar itu langsung terkoneksi ke semua pendaftaran(Terintegrasi), user tinggal pilih opsi. Ini malah tiap menu harus isi nomor KTP, Tgl Lahir Dan lain lain berulang ulang. Bukan memudahkan dalam penggunaan aplikasi, malah jadi menyulitkan dan membebani pengguna. Tolong Disederhanakan cara input registrasinya, jangan berulang ulang Jaelani /Bandung
U003	Negatif	Aplikasi gak jelas terlalu memaksakan proyek berbasis aplikasi. Kesiapan sistem sangat buruk, time connection, tidak terdaftar, aplikasi lemot dan keluar sendiri. Sistem pembayaran juga tidak fleksibel, bukan mempermudah tetapi semakin mempersulit masyarakat. Bagaimana nasib yang tidak mempunyai HP sedangkan membutuhkan bensin buat mencari nafkah? Semoga bisa refleksi dan evaluasi agar tidak terkesan maksa, dipikir ulang kalo mau buat suatu sistem dikira gampang kali
U004	Negatif	Aplikasi yg cukup menyusahkan, banyak bug di bagian otp, untuk pembayaran msh blm flexible cuma bisa pke link aja, cukup meribetkan pengendara karena tidak semua punya handphone dan semua bisa pake aplikasi ini, blm lagi kalau kuota habis/hpnya lemot jadi kurang efektif ketika terjadi antre panjang, kurangnya sosialisasi sehingga masyarakat blm siap untuk menerima kebijakan ini.
U005	Negatif	Bug di bagian profil (gak ada background). Icon dan tulisan juga kurang rapih. Waktu klik debit instant ada tulisan terjadi kesalahan, tapi pas back, malah ngeloop. Ini kerjaan developernya masih banyak. Kayak aplikasi beta. UI UX masih perlu banyak improve, padahal ini sekelas BUMN.
U006	Negatif	Kedepannya perlu banyak inovasi, termasuk metode pembayaran yang jangan cuma pake linkAja yang penggunaanya jarang.. dan yang paling ganggu adalah, loginnya yang pake timeout. Jarang dipake jadi logout sendiri
U007	Negatif	Udah efektif bayar pake uang cash, malah buat ribet. Seharusnya ini tuh dijadikan sebuah opsi bukan kewajiban . Hadeh, belum lagi masalah bug dalam aplikasi kadang force close dan pas dibuka aplikasi nya putih tampilan menu munculnya lama. Emang dah kalo ada yang susah kenapa pilih yg mudah...wkwkwkw

Berdasarkan data yang telah melalui proses sortir data sebelumnya, data pada tabel 5.2 diatas menampilkan kolom kelas yang mana data kolom kelas ini ditentukan dan dibuat oleh peneliti. Data yang sudah diberikan kelas ini sudah dapat diproses dengan *tools* dan dapat diuploadkan pada Google Drive.

3.3 Pengujian Dengan Google Colaboratory

Adapun tahapan pengujian pada *Data Mining* dengan menggunakan Google Colaboratory adalah sebagai berikut :

1. Unggah Data Ke Google Drive

Upload data dilakukan untuk memasukkan data yang akan diuji dalam bentuk format Excel. Selanjutnya memasukkan data ke Google Drive agar data dapat dipanggil dengan *coding* pada Google Colaboratory.

2. Mengimport *Library*

Tabel 3. Deskripsi *Import Library*

Kode	deskripsi
Import pandas as pd	Untuk menganalisis, membersihkan, mengeksplorasi.
From google.colab import drive	Untuk memanggil data dari google drive

3. Memanggil Data Dari *Google Drive*

Pada tahap ini melakukan pemanggilan data yang sudah diunggah ke *Google Drive* sebelumnya.

4. Membuat Sub Program Dengan Nama "*Cleansing*"

Membuat sub program dengan nama *cleansing* yang mempunyai parameter penampung yang berada didalam kurung yaitu data. Sub program ini berisi *function* yang dapat mengubah data menjadi huruf kecil, menghapus tanda baca, *emoticon*, dan enter (*new line*). Sub program ini dapat membalikkan data setelah data di olah pada *function* tadi.

5. Menginstall *Library* Sastrawi

Python Sastrawi merupakan *library* sederhana yang dapat mengubah kata berimbuhan Bahasa Indonesia menjadi bentuk dasarnya. Sastrawi juga dapat diinstal melalui "pip".

6. *Import Library Stopwordremover* Dari *Library* Sastrawi

Mengimport *library stopwordsremover* dari *library* sastrawi. Memanggil *function* *StopWordRemoverFactory()* yang disimpan kedalam variabel *factory*, selanjutnya membuat *stop word remover* tadi pada *function* nya yang disimpan kedalam variabel *stopword*. Selanjutnya mengimplementasikannya pada dataset.

7. *Import Library Stemmer* Dari *Library* Sastrawi

library Stemmer dari *library* sastrawi. Memanggil *function* *StemmerFactory()* yang disimpan kedalam variabel *factory*, selanjutnya membuat *stemmer* tadi pada *function* nya yang disimpan kedalam variabel *stemmer*. Selanjutnya mengimplementasikan nya pada data olah kita dengan cara membuat variabel *review* yang berisi *array* kosong, selanjutnya melakukan perulangan pada dataset yang dimana didefinisikan sebagai variabel *row*, kemudian menggabungkan *array* variabel *review* dengan hasil dari *function* *stemmer.stem* dimana kita mengirimkan parameter perulangan dataset kolom 'teks' secara satu persatu untuk diolah. Kemudian kita replace dataset kolom 'teks' dengan variabel *review*.

8. Hapus Kata Yang Kurang Dari 4 Dan Lebih Dari 20 Karakter

Selanjutnya kita menghapus kata yang berisi kurang dar 4 huruf dan lebih dari 20 huruf. Membuat variabel *filterlength* dan *text* yang berisi *array* kosong, membuat variabel *i* yang berisi int 0. Melakukan perulangan pada dataset yang dimana didefinisikan sebagai variabel *row*, kemudian menggabungkan *array* variabel *text*

2. Visualisasi Wordcloud

Selanjutnya membuat visualisasi *Wordcloud* untuk dataset (gabungan data training dan data testing) yang sudah terklasifikasi dengan metode *Naïve Bayes* dan *K-Nearest Neighbor*.



Gambar 7. Visualisasi *Wordcloud* *Naïve Bayes* (Kiri Positif, Kanan Negatif)



Gambar 8. Visualisasi *Wordcloud* *K-Nearest Neighbor* (Kiri Positif, Kanan Negatif)

4. KESIMPULAN

Pada penelitian ini menghasilkan sentimen positif dan negatif pada masing-masing kedua metode. Untuk metode *K-Nearest Neighbor* terdapat sentimen positif dengan total 117 ulasan dan sentimen negatif dengan total 1099 ulasan berdasarkan dari analisis klasifikasi yang telah dilakukan. Sementara untuk metode *K-Nearest Neighbor* terdapat sentimen positif dengan total 84 ulasan dan sentimen negatif dengan total 1132 ulasan berdasarkan dari analisis klasifikasi yang telah dilakukan.

Pada penelitian ini menghasilkan output visualisasi dari dari kedua metode klasifikasi *K-Nearest Neighbor* dan *Naïve Bayes* dalam analisis sentimen terhadap review mengenai aplikasi *MyPertamina* yang terdapat pada *Google Playstore* dan penerapan langsung dengan bahasa pemrograman *Python* pada *Tools Google Colaboratory*. Visualisasi yang ditampilkan berupa *Wordcloud* dari klasifikasi sentimen positif dan juga negatif. Dari kedua metode didapat hasil dari masing-masing nilai akurasi sebesar 70,73% untuk *Naïve Bayes* dan sebesar 85,97% untuk *K-Nearest Neighbor*. Hal ini menunjukkan bahwa tingkat akurasi dari metode *K-Nearest Neighbor* lebih baik dalam melakukan klasifikasi teks ulasan aplikasi *MyPertamina* jika dibandingkan dengan tingkat akurasi yang dihasilkan dari metode *Naïve*

DAFTAR PUSTAKA

- [1] Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 785–795.
 - [2] Hikmawati, N. K. (2022). Analisis Kualitas Layanan My Pertamina Menggunakan Pendekatan e-GovQual pada Beberapa Kota Percobaan. *Jurnal Manajemen Informatika (JAMIKA)*, 12, 100–111.
 - [3] Rahma, L., Syaputra, H., Mirza, A. H., & Purnamasari, S. D. (2021). Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once). *Jurnal Nasional Ilmu Komputer*.
 - [4] Anam, M. K., Pikir, B. N., Firdaus, M. B., Erlinda, S., & Agustin. (2021). Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*.
 - [5] Kalingara, N. I. P., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes. *E-Proceeding of Engineering Telkom University*.
 - [6] Surohman, S., Aji, S., Rousyati, R., & Wati, F. F. (2020). Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor. *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 93–105.
 - [7] Nurwahyuni, S. (2019). Analisis Sentimen Aplikasi Transportasi Online KRL access menggunakan metode Naive Bayes. *Swabumi*, 7(1), 31–38.
 - [8] Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71.
 - [9] Erfina, A., Basryah, E. S., Saepulrohman, A., & Lestari, D. (2020). Analisis Sentimen Aplikasi Pembelajaran Online Di Play Store Pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine (Svm). *Seminar Nasional Informatika (SEMNASIF)*, 1(1), 145–152.
 - [10] Santoso, D. P., & Wibowo, W. (2022). Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store. *Jurnal Sains Dan Seni ITS*, 11(2).
 - [11] Izunnahdi, M., Aburrahman, G., & Wardoyo, A. E. (2021). Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naïve Bayes. *Jurnal Smart Teknologi*, 1(1), 1–7.
 - [12] Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes. *Jurnal Ilmiah Teknologi Informasi Asia*.
 - [13] Setiawan, R. A., Estethika, A. K. A., Nurhayanto, E. M. O., Asmara, Y., & Wahyudi, A. (2021). Analisis Sentimen Hotel di Nusa Tenggara Barat Menggunakan Algoritma SVM. *Seminar Multimedia Dan Artificial Intelligence: Optimalisasi Artificial Intelligence Di Era Revolusi Industri 4.0 Dan Society 5.0*.
 - [14] Permana, K., Putra, Z. P., & Nugroho, A. (2021). Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop ! Surabaya. *Jurnal Ilmu Komputer Dan Bisnis (JIKB)*.
-

- [15] Amrullah, A. Z., Sofyan Anas, A., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal BITE*.
- [16] Suryadi, Ridho, A., & Murhaban. (2021). Analisis Sentimen Review Hotel Menggunakan Algoritma Naive Bayes Classifier dan Pendekatan Lexicon based. *Techsi: Jurnal Penelitian Teknik Informatika*.
- [17] Negara, A. B. P., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi Dan Ilmu Komputer*.
- [18] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- [19] Pati, G. K., & Umar, E. (2022). Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K- Nearest Neighbor. *Jurnal Media Informatika Budidarma*, 6, 2309–2315.
- [20] Darmawan, R., & Amini, S. (2022). Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naive Bayes dan K-Nearest Neighbor pada Twitter. *Seminar Nasional Mahasiswa, September*, 78–85.
- [21] Novalia, V., Goejantoro, R., & Sifriyani. (2020). Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor (Studi Kasus : Status Kerja Penduduk Di Kabupaten Kutai Kartanegara Tahun 2018). *Jurnal EKSPONENSIAL*.

