



# ANALISIS SENTIMEN PADA ULASAN APLIKASI MAXIM DI *GOOGLE PLAY STORE DENGAN K-NEAREST NEIGHBOR*

## TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat  
untuk Memperoleh Gelar Sarjana Komputer pada  
Program Studi Sistem Informasi

Oleh:

**RESTU RAMADHAN**

**11653101468**



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2023

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR PERSETUJUAN

### ANALISIS SENTIMEN PADA ULASAN APLIKASI MAXIM DI GOOGLE PLAY STORE DENGAN K-NEAREST NEIGHBOR

#### TUGAS AKHIR

Oleh:

**RESTU RAMADHAN**

**11653101468**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir  
di Pekanbaru, pada tanggal 17 Juli 2023

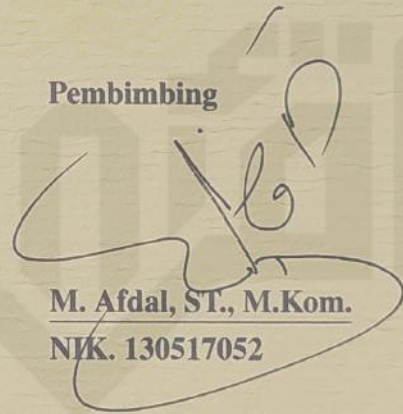
**Ketua Program Studi**



**Eki Saputra, S.Kom., M.Kom.**

**NIP. 198307162011011008**

**Pembimbing**



**M. Afdal, ST., M.Kom.**

**NIK. 130517052**

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengummumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**LEMBAR PENGESAHAN**

**ANALISIS SENTIMEN PADA ULASAN APLIKASI MAXIM DI  
GOOGLE PLAY STORE DENGAN K-NEAREST NEIGHBOR**

**TUGAS AKHIR**

Oleh:

**RESTU RAMADHAN**

**11653101468**

Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer  
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau  
di Pekanbaru, pada tanggal 7 Juli 2023

Pekanbaru, 7 Juli 2023

Mengesahkan,

**Ketua Program Studi**

**Eki Saputra, S.Kom., M.Kom.**

**NIP. 198307162011011008**



**Dr. Hartono, M.Pd.**

**NIP. 196403011992031003**

**DEWAN PENGUJI:**

**Ketua : Syaifullah, SE., M.Sc.**

**Sekretaris : M. Afdal, ST., M.Kom.**

**Anggota 1 : Inggih Permana, ST., M.Kom.**

**Anggota 2 : Muhammad Jazman, S.Kom., M.InfoSys.**

© Hak Cipta milik UIN Suska Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 7 Juli 2023

Yang membuat pernyataan,



**RESTU RAMADHAN**

**NIM. 11653101468**

UIN SUSKA RIAU

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*Alhamdulillah rabbil 'aalamiin*, Puji syukur atas karunia dan kuasa Allah SWT sehingga pada akhirnya saya bisa menyelesaikan Tugas Akhir ini. Tidak lupa pula sholawat dan salam hadiahkan kepada Nabi besar Muhammad SAW yang telah menuntun kita dari zaman jahiliyah ke zaman yang terang benderang ini.

Untuk bapak dan ibu tersayang, inilah yang dapat ku persembahkan meski tak akan pernah tergantikan atas pengorbanan yang kalian berikan. Namun, saat ucapan syukur itu menggema, hilanglah sudah dahagaku. Ampunilah dosa mereka Ya Allah.

Buat Orang tua kedua saya, Penasehat Akademik Bapak Bapak Arif Marsal, Lc., MA yang ramah serta baik hati membimbing saya. Terimakasih telah mendengarkan keluh kesah saya, dan banyak memberi arahan serta motivasi sehingga saya mampu bangkit dari keterpurukan selama kuliah ketika masa-masa sulit, semoga bapak dan keluarga selalu diberi perlindungan oleh Allah SWT.

Untuk pembimbing Tugas Akhir saya Bapak M. Afdal, ST., M.Kom terimakasih telah membimbing saya, sehingga saya mampu melanjutkannya hingga menjadi Tugas Akhir seperti saat ini. Saya sangat berterimakasih sebesar-besarnya dan saya mohon maaf karena selama bimbingan sudah banyak merepotkan bapak.

Buat kakak terimakasih buat semangat yang telah kalian berikan dan do'a yang kalian berikan. Untuk teman-teman seperjuanganku FAM'S dan lainnya, terimakasih sudah memberi dorongan dan masukan sehingga Tugas Akhir ini terlaksana.

Semoga Allah membalas segala kebaikan bapak, ibu, abang, kakak, dan teman yang telah berkontribusi dalam penyelesaian Tugas Akhir ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## KATA PENGANTAR

Puji beserta syukur yang amat luar biasa atas kehadiran Allah SWT berkat karunia-Nya Laporan Tugas Akhir dengan judul “Analisis Sentimen pada Ulasan Aplikasi Maxim di *Google Play Store* dengan *K-Nearest Neighbor*” ini bisa terselesaikan oleh penulis dengan sebaik mungkin. Adapun sholawat beserta salam tidak lupa pula penulis ucapkan kepada Nabi Muhammad SAW dengan melafadzkan “*Al-lahumma Sholli ‘ala Sayyidina Muhammad, Wa’ala Aalihi Sayyidina Muhammad*”.

Dalam penyelesaian laporan ini, penulis tidak terlepas dari bantuan-bantuan dan masukan-masukan yang telah diberikan oleh berbagai pihak sehingga menjadi seperti saat ini. Oleh karena itu, penulis mengucapkan terimakasih sebanyak-banyaknya kepada:


1. Bapak Prof. Dr. Hairunas, M.Ag sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd sebagai Dekan Fakultas Sains dan Teknologi.
3. Bapak Eki Saputra, S.Kom., M.Kom sebagai Ketua Program Studi Sistem Informasi.
4. Bapak Arif Marsal, Lc., MA sebagai Dosen Penasehat Akademik yang sudah memberikan informasi, arahan, maupun bimbingan sehingga saya berhasil menyelesaikan Tugas Akhir ini.
5. Bapak M. Afdal, ST., M.Kom yang telah senantiasa membantu dan membimbing saya sehingga saya bisa berhasil menyelesaikan penulisan Tugas Akhir ini.
6. Seluruh dosen Program Studi Sistem Informasi yang senantiasa membantu penulis dalam memberikan informasi yang berkaitan dengan Tugas Akhir ini.
7. Bapak tercinta saya, Yuswardi yang selalu mendukung dan mendo’akan anaknya agar cepat selesai.
8. Ibu tercinta saya, Witneti Hasri yang selalu memberikan do’a, motivasi, mendukung anaknya agar cepat selesai.
9. Kakak saya Yossy Fraseptya Utami dan Abang Yayat terimakasih atas dukungan dan masukkan.
10. Helma, Andre, Dede, Fadhil, Fikri, Reynaldi, Rezky, Wahyu, Yonda yang selama Tugas Akhir ini memberikan tempat untuk berdiskusi.
11. Pacar Dea Rahmadayanti yang selama ini selalu mendorong agar semangat mengerjakan Tugas Akhir ini, Sahabat Nur Oktavia dan teman-teman yang tidak disebutkan masih senantiasa memberikan dukungan dan moti-

vasi kepada penulis hingga saat ini.

Selama penulisan Tugas Akhir ini, atas kesadaran penulis masih banyak hal yang harus di perbaiki. Maka penulis mengharapkan kritik beserta saran dari segala pihak demi terciptanya Tugas Akhir ini dengan sempurna. Semoga dengan ini, dapat memberikan suatu manfaat yang baik bagi kita semua *Aamiin yaa robbal 'aalamiin*.

Pekanbaru, 17 Juli 2023

Penulis,



**KESU RAMADHAN**

**NIM. 11653101468**

UIN SUSKA RIAU

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





Medan, 30 Juni 2023

No : 496/LOA-JURIKOM/VI/2023

Lamp :

Hal : Surat Penerimaan Naskah Publikasi

Kepada Yth,

Bapak/Ibu **Restu Ramadhan**

Di Tempat

Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada JURIKOM (Jurnal Riset Komputer) (pISSN 2407-389X / eISSN 2715-7393), dengan judul:

Judul : **Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor**

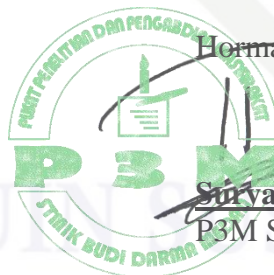
Penulis: **Restu Ramadhan(\*) , M Afdal, Inggih Permana, Muhammad Jazman**

Berdasarkan hasil review dari reviewer, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Volume 10, Nomor 3, Juni 2023**

QR-Code untuk melihat link LOA Jurnal JURIKOM (Jurnal Riset Komputer), **Volume 10, Nomor 3, Juni 2023** yang telah dikeluarkan.

Sebagai informasi tambahan, saat ini JURIKOM telah **TERAKREDITASI** dengan Peringkat **SINTA 4** berdasarkan SK Kepmendikbudristek No. [158/E/KPT/2021](#) tertanggal **09 Desember 2021** dimulai dari Volume 5 No 4, tahun 2018, hingga Volume 10 No 3 Tahun 2023.

Demikian informasi yang kami sampaikan, atas perhatiannya kami ucapkan terimakasih.



Hormat Kami,

**Surya Darma Nasution, M.Kom**

P3M STMIK Budi Darma

Tembusan:

1. Ketua STMIK Budi Darma
2. Author
3. Files

ISSN 2407-389X (media cetak)  
ISSN 2715-7393 (media online)

© Hak Cipta milik UIN Suska Riau

# JURIKOM

## Jurnal Riset Komputer

Volume 7, Nomor 1, Februari 2020

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

State Islamic University of Sultan Syarif Kasim Riau



Diterbitkan Oleh :  
**P3M STMIC Budi Darma**  
Telp : (061) 787 5998 - 821 5652  
Fax : (061) 787 648  
Website :  
[ejurnal.stmik-budidarma.ac.id](http://ejurnal.stmik-budidarma.ac.id)



DAFTAR ISI

<b>Analisis Metode SAW dan WP dalam Pemilihan Costumer Service Berdasarkan Pembobotan ROC</b>	686-696
Agus Iskandar	
<b>Penerapan Algoritma FP-Growth Dalam Pencarian Hubungan Antara Waktu Pembelian Dan Barang yang Dibeli Untuk Strategi Promosi Penjualan</b>	697-703
Tasya Marzuqah, Inggih Permana, M Afdal	
<b>Perancangan UI/UX Fitur Asrama Mahasiswa Berbasis Website dengan Pendekatan User Centered Design</b>	704-714
Fika Febrika, Pacu Putra Suarli, Nabila Rizky Oktadini, Allsela Meiriza, Putri Eka Sevtiyuni, Endang Lestari Ruskan, Dedy Kurniawan	
<b>Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor</b>	715-724
Restu Ramadhan, M Afdal, Inggih Permana, Muhammad Jazman	
<b>Rancang Bangun Sistem Pemantauan dan Kendali Budidaya Anggur Dengan Penerapan Internet Of Things (IoT) Berbasis Android</b>	725-734
Nur Fitriana Putri, Rahmi Hidayati, Irma Nirmala	
<b>Development of Village Information System using Waterfall and TAM Method in XYZ Ward</b>	735-741
Suwarno, Suwarno, Yefta Christian, Jonathan Jonathan, Marvin Christian, Vendryan Vendryan, Chris Tan	
<b>Sistem Pemantauan dan Kontrol Smart Cat Room Berbasis Internet Of Things (IoT)</b>	742-751
Ariansyah, Sudarsono, Rahmi Hidayati, Uray Ristian	
<b>Perancangan Aplikasi Augmented Reality Sebagai Media Pembelajaran Tata Surya Pada Sekolah Dasar Kelas 6</b>	752-761
Aslam Hadil Matin, Hari Widi Utomo	

Hak Cipta © dan Hak Pengelolaan Unesa-Ui-Ui  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mengemukakan sumber.  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor

Restu Ramadhan<sup>\*</sup>, M Afdal, Inggih Permana, Muhammad Jazman

Fakultas Sains dan Teknologi, Sistem Infomasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia  
Email: <sup>1</sup>restu301297@gmail.com, <sup>2</sup>m.afdal@uin-suska.ac.id, <sup>3</sup>inggihpermana@uin-suska.ac.id, <sup>4</sup>jazman@uin-suska.ac.id  
Email Penulis Korespondensi: restu301297@gmail.com  
Submitted 18-06-2023; Accepted 30-06-2023; Published 30-06-2023

## Abstrak

Transportasi online adalah Inovasi dalam teknologi yang muncul untuk memecahkan berbagai masalah yang timbul pada transportasi umum konvensional seperti halnya dalam kemudahan pemesanan, ketersediaan dan digitalisasi pembayaran. Maxim adalah perusahaan transportasi online yang sudah beroperasi sejak tahun 2018 di Indonesia. Seiring meningkatnya jumlah pengguna aplikasi maxim, membuat tuntutan terhadap mutu pelayanan aplikasi juga meningkat. Dalam Google Play Store, ulasan dan informasi suatu aplikasi disimpan dalam bentuk teks. Salah satu proses pengambilan informasi text mining dalam kategori teks adalah Analisis Sentimen untuk melihat kecenderungan suatu sentimen atau pendapat apakah bersifat positif, netral, atau negatif pada ulasan pengguna aplikasi Maxim. Proses klasifikasi sentimen menggunakan algoritma K-NN menghasilkan akurasi, presisi, dan recall berturut-turut sebesar 90,23%; 90,23%; dan nilai recall 72,38% dengan percobaan menggunakan 90% data latihan dan 10% data uji serta nilai  $k = 5$ .

**Kata Kunci:** Analisis Sentimen; Google Play Store; K-Nearest Neighbor; Maxim; Text Mining

## Abstract

Online transportation is an innovation in emerging technology to solve various problems that arise in conventional public transportation such as in the ease of ordering, availability, and digitization of payments. Maxim is an online transportation company that has been operating since 2018 in Indonesia. As the number of users of the maxim application increases, demands for the quality of application service also increase. In the Google Play Store, reviews and information about an app are stored in text form. One of the processes of extracting text mining information in the text category is Sentiment Analysis to see the tendency of a sentiment or opinion whether it is positive, neutral, or negative at the Maxim application user reviews. The sentiment classification process using the K-NN algorithm produces accuracy, precision, and recall of 90.23%; 90.23%; and a recall value of 72.38% with an experiment using 90% training data, 10% test data, and a value of  $k = 5$ .

**Keywords:** Google Play Store; K-Nearest Neighbor; Maxim; Sentiment Analysis; Text Mining

## 1. PENDAHULUAN

Transportasi online adalah suatu penyelenggaraan Lalu Lintas dan Angkutan Jalan yang berjalan dengan mengikuti serta memanfaatkan perkembangan ilmu pengetahuan (teknologi) berbasis aplikasi dan online baik untuk pemesanan maupun pembayaran [1]. Inovasi ini muncul untuk memecahkan berbagai masalah yang timbul pada transportasi umum konvensional seperti kemudahan pemesanan, ketersediaan dan digitalisasi pembayaran. Perkembangan teknologi yang sangat pesat membawa pengaruh yang besar bagi manusia. Salah satu pengaruh dari perkembangan teknologi adalah semakin maraknya penggunaan transportasi online melalui media internet. Menurut data dari riset *Google*, Temasek dan Bain & Company [2]. Pertumbuhan angka pasar transportasi online di Indonesia mengalami pertumbuhan sangat besar, pada tahun 2015 bernilai US\$ 900 juta meningkat di tahun 2018 sebesar US\$ 12,7 miliar dan di prediksi mencapai US\$ 40 miliar pada 2025. Dari riset tersebut juga menunjukkan bahwa Indonesia merupakan negara dengan pertumbuhan pangsa pasar transportasi online terbesar di Asia Tenggara. Sementara itu, berdasarkan survei yang dilakukan Asosiasi Penyelenggara Jasa Internet Indonesia memaparkan pemesanan transportasi online menempati urutan ke 16 dari 22 alasan mengapa masyarakat menggunakan internet dan 93,9% masyarakat mengakses internet menggunakan smartphone [3].

Terdapat beberapa transportasi online di Indonesia yaitu, Gojek, Grab, Maxim, FastGo, Bitcar, Bonceng, Anterion.id [4]. Maxim adalah perusahaan internasional yang bergerak dibidang teknologi informasi dimana perusahaan ini menyediakan platform yang mempertemukan pengemudi dan pelanggan secara mudah. Maxim adalah perusahaan transportasi online yang bisa dibilang baru di Indonesia. Maxim sendiri sudah ada sejak tahun 2003 lalu. Namun dulunya hanya sebuah layanan aplikasi taksi di kota Chardinsk yang terletak di pegunungan Ural, Rusia. Di Indonesia maxim sendiri pertama kali beroperasi pada tahun 2018. Perusahaannya semakin melebarkan sayap dengan tidak hanya menjadi perusahaan transportasi online yang berfokus ke taksi saja, melainkan juga jenis layanan angkutan lain seperti ojek atau mobil pada umumnya. Maxim mengembangkan sendiri aplikasinya dan menciptakan sistem perangkat keras dan perangkat lunak yang memungkinkan mitra-mitranya terhubung kelayanan perusahaan, memperoses jutaan setiap hari, memantau kualitas kerja dan pelayanan, serta menganalisis dan mengoptimalkan bisnis mereka. Maxim memiliki kebijakan yang menguntungkan bagi pengemudi maupun pelanggan, seperti jadwal yang fleksibel untuk pengemudi, harga terjangkau serta sistem reservasi order untuk pelanggan [5].

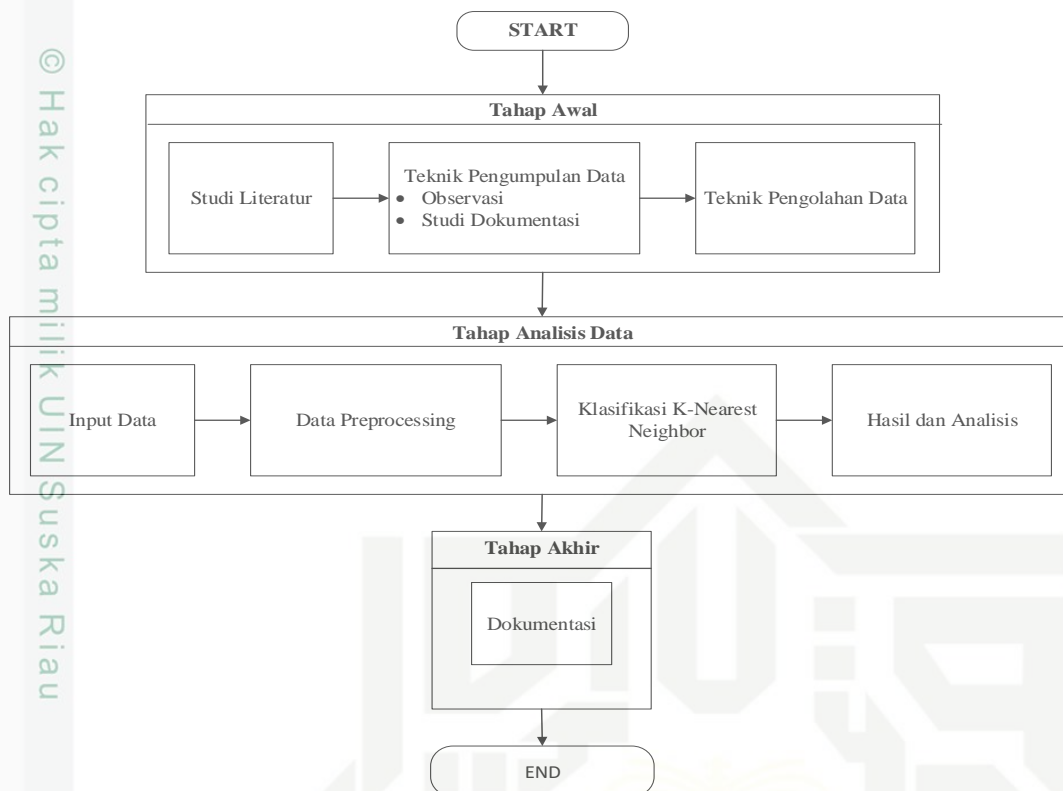
Seiring meningkatnya jumlah pengguna *smartphone* dan kendaraan berakibat pada semakin banyaknya para pengguna aplikasi maxim, hal ini membuat tuntutan terhadap mutu pelayanan aplikasi juga meningkat. Dalam Google Playstore setiap pengguna dapat memberikan rating dan ulasan terhadap suatu aplikasi. Menurut [6], ulasan dan informasi suatu produk disimpan dalam bentuk teks, maka text mining merupakan solusi dalam pengambilan informasi yang berbentuk teks. Salah satu proses pengambilan informasi text mining dalam kategori teks adalah *Sentiment Analysis* atau

Analisis Sentimen. Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks [7]. Analisis sentimen diperlukan sebagai alat untuk mengklasifikasikan sebuah informasi yang berbentuk teks bahasa ke dalam kategori positif dan negatif [8]. Analisis sentimen digunakan untuk melihat kecenderungan suatu sentimen atau pendapat, apakah pendapat tersebut cenderung beropini positif atau negatif. Menurut [9], analisis sentimen pada review pengguna adalah proses menyelidiki review produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan. Analisis sentimen adalah teknik komputasi pendapat, perasaan dan subjektivitas teks [10]. Tugas dasar dalam analisis sentimen adalah untuk mengklasifikasikan polaritas teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas berarti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif. Oleh karena itu, dibutuhkan suatu analisis terhadap opini-opini tersebut dalam penelitian ini agar bisa dijadikan tolak ukur baik atau tidaknya jasa transportasi online menurut pelanggan. Tujuan dari analisis sentimen selain untuk mendapatkan informasi sentimen yang terkandung dalam sebuah teks, juga untuk melihat pendapat teks yang berkaitan dengan masalah atau objek, apakah itu cenderung memiliki pandangan positif atau negatif. Proses pengklasifikasian opini melalui pendekatan text mining memerlukan metode yang mampu mengklasifikasikan opini secara akurat. Penelitian sebelumnya terkait analisis sentimen pernah dilakukan untuk menganalisis opini mahasiswa terhadap penilaian kinerja dosen berdasarkan kuesioner dengan menerapkan algoritma *K-Nearest Neighbor* (KNN) yang menghasilkan akurasi sebesar 81%, *precision* sebesar 81%, *recall* sebesar 81%, dan *F1-Score* sebesar 81% [11]. Penelitian lainnya pada tahun 2020 dilakukan oleh Angelina, *dkk* untuk menganalisis sentimen pengguna Twitter terhadap Aplikasi Ruang Guru yang menghasilkan nilai optimal pada model optimasi algoritma PSO berbasis SVM dengan nilai akurasi sebesar 78,55% dan AUC sebesar 0,853 [12]. Sehingga dalam pengklasifikasian ini diterapkan algoritma KNN dan pembobotan untuk setiap kata menggunakan TF-IDF. Ada beberapa kelebihan dari metode K-NN yaitu algoritma klasifikasi *K-Nearest Neighbor* yang terbukti mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan dalam sebuah aplikasi [13]. Hal tersebut diketahui karena algoritmanya yang mudah dan efisien untuk klasifikasi teks. Bukan hanya mudah dan efisien, sifat dari algoritma *K-Nearest Neighbor* sendiri bersifat *self-learning*, dimana algoritma ini dapat mempelajari struktur data yang ada dan mengkategorikan dirinya sendiri [14]. Penelitian ini hanya berfokus menganalisa kepuasan pengguna atau pelanggan Maxim, dimana kepuasan adalah suatu sikap yang diputuskan berdasarkan pengalaman yang didapat [15]. Penelitian ini menggunakan data ulasan pengguna Maxim untuk kemudian diproses dan dianalisis dengan algoritma KNN sehingga menghasilkan klasifikasi sentimen terhadap tiap ulasan pengguna yang dibagi menjadi tiga kelas sentimen yaitu sentimen positif, negatif, dan netral.

## 2. METODOLOGI PENELITIAN

Tahapan penelitian ini digambarkan melalui diagram alir metodologi penelitian pada Gambar 1. Tahap awal dalam penelitian adalah melakukan studi literatur untuk menemukan landasan teori dan referensi tertulis yang mendasari penelitian guna menyelesaikan permasalahan serta dijadikan referensi yang kuat dalam melakukan analisa. Pemrosesan data dimulai dengan (a) Tahap pra-proses data; (b) Pembagian data latih dan data uji; (c) Pemodelan klasifikasi data dengan K-NN; dan (d) Analisis sentimen hasil klasifikasi. Tahapan pra-proses data ulasan terdiri dari tahap *case folding*, *tokenizing*, *cleansing*, *stopword removal*, *stemming*, pelebelan data ulasan, dan diakhiri dengan proses pembobotan TF-IDF. Setelah data selesai diproses kemudian dilakukan proses distribusi data latih dan data uji dengan kuantitas yang beragam, kemudian dilakukan pemodelan data dengan KNN menggunakan beberapa nilai parameter *k*. Data yang digunakan merupakan data yang diperoleh dari ulasan pengguna aplikasi Maxim selama satu bulan terhitung dari tanggal 7 Agustus 2022 sampai dengan 7 September 2022 dengan jumlah 10291 ulasan. Data ulasan pengguna aplikasi Maxim diperoleh di *Google Play Store* yang dapat diakses melalui tautan <https://play.google.com/store/apps/details?id=com.taxsee.taxsee&hl=id&gl=US&pli=1> dengan menggunakan teknik *scraping* melalui bahasa pemrograman *Python*. Proses *scraping* data ulasan menggunakan *Package google\_play\_scraper* yang tersedia di *python*.

UIN SUSKA RIAU



**Gambar 1.** Alir Diagram Penelitian

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber atau dengan cara lain.  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.1 Analisis Sentimen

Analisis sentimen adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial berdasarkan fakta dan opini [16]. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut, dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien. Sentiment analysis merupakan salah satu bidang dari *Natural Language Processing (NLP)* yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks.

## 2.2 Text Mining

*Text mining* adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data. Ini juga dikenal sebagai analisis teks, meskipun beberapa orang menarik perbedaan antara dua istilah; dalam pandangan itu, analitik teks adalah aplikasi yang diaktifkan oleh penggunaan teknik *text mining* untuk memilah-milah set data [17] [18].

## 2.3 Text Preprocessing

Berdasarkan ketidakteraturan struktur data teks, maka proses sistem temu kembali informasi ataupun *text mining* memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Salah satu implementasi dari *text mining* adalah tahap *text preprocessing* [19]. Tahap *Preprocessing* diantaranya:

### a. Case Folding

Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau lowercase). *Case folding* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf „a“ sampai dengan „z“ yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter*.

### b. Tokenizing

Tahap *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

### c. Filtering

Tahap *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya.

#### d. Stemming

Teknik Stemming diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan katakata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

#### 2.4 K-Nearest Neighbour (KNN)

Algoritma k-Nearest Neighbor adalah 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. Algoritma k-Nearest Neighbor menggunakan Neighborhood Classification sebagai nilai prediksi dari nilai instance yang baru [20].

### 3. HASIL DAN PEMBAHASAN

Data ulasan pada Tabel 1 yang diperoleh dari hasil *scraping* tidak dapat langsung digunakan pada proses mining dengan algoritma Data Mining KNN. Data ulasan perlu dipraproses untuk menghilangkan data noise, *missing values*, atau data ulasan yang tidak memiliki informasi yang berguna, tidak memiliki makna, dan data yang sekiranya dapat mengganggu proses mining.

**Tabel 1.** Ulasan Pengguna Aplikasi Maxim

No	Ulasan	Tanggal
1	Lokasinya tambahin dong	07 September 2022
2	drvr nya ramah	07 September 2022
3	sangan bagus dan baik	07 September 2022
4	makasih ya driver	07 September 2022
5	Ok trmks	07 September 2022
6	sangat membantu Dan sopir nya sang at ramah	07 September 2022
7	Tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah	07 September 2022
8	sangat memuaskan	07 September 2022
...	...	...
10290	terimakasih banyak..supirnya sangat santun dan baik banget..	07 August 2022
10291	Bagusss	07 August 2022

#### 3.1 Praproses Data

Tahapan praproses data ulasan terdiri dari tahap *case folding*, *tokenizing*, *cleansing*, *stopword removal*, *stemming*, pebelan data ulasan, dan diakhiri dengan proses pembobotan TF-IDF.

##### 3.1.1 Tahap Case Folding

*Case folding* merupakan tahap untuk mengubah ulasan yang berbentuk teks atau string berhuruf kapital atau berhuruf kapital dan huruf kecil, menjadi keseluruhan teks berhuruf kecil. Hal ini dilakukan agar setiap string yang memiliki susunan kata yang sama hanya akan membentuk satu fitur dari susunan kata tersebut. Berikut merupakan hasil dari proses *Case Folding* pada Tabel 2.

**Tabel 2.** Data Hasil Case Folding

No	Ulasan	Case Folding
1	Lokasinya tambahin dong	lokasinya tambahin dong
2	drvr nya ramah	drvr nya ramah
3	sangan bagus dan baik	sangan bagus dan baik
4	makasih ya driver	makasih ya driver
5	Ok trmks	ok trmks
6	sangat membantu Dan sopir nya sang at ramah	sangat membantu dan sopir nya sang at ramah
7	Tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah	tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah
8	sangat memuaskan	sangat memuaskan
...	...	...
10290	terimakasih banyak..supirnya sangat santun dan baik banget..	terimakasih banyak..supirnya sangat santun dan baik banget..
10291	Bagusss	bagusss

##### 3.1.2 Tahap Tokenizing

Tahap *tokenizing* merupakan proses memecahkan kalimat atau string menjadi token berupa potongan karakter atau kata. Contohnya seperti kalimat "titik jemput akurat" akan dipecah menjadi potongan-potongan kata "titik", "jemput", dan

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

“akurat”. Proses *tokenizing* pada ulasan menggunakan *library nltk* yang tersedia di *python*. Berikut merupakan hasil pada data yang dilakukan proses *tokenizing* pada Tabel 3.

### 3.1.3 Tahap *Cleansing*

Tahap *Cleansing* dibagi menjadi dua tahap, tahap pertama adalah proses perubahan kata slang atau kata tidak baku pada ulasan, proses normalisasi mengacu pada kamus kata normalisasi yang digunakan untuk mengubah data tidak baku menjadi kata baku atau normal. Seperti kata “7an” dinormalisasi menjadi kata “tujuan”; kata “abis” menjadi “habis”; kata “adalah” menjadi “adalah”, dan lain sebagainya. Tahap selanjutnya adalah tahap *filtering* yang merupakan proses untuk menghapus karakter dalam string yang bukan bagian dari alfabet dan tidak memiliki makna tertentu sehingga mengurangi karakter yang tidak dibutuhkan, seperti angka; *hashtag*; emoji; *link*; tanda baca, dan lain sebagainya. Berikut merupakan hasil akhir dari proses *cleansing* data ulasan pada Tabel 3.

**Tabel 3.** Data Hasil *Tokenizing & Cleansing*

No	Tokenizing	Cleansing
1	['lokasinya', 'tambahin', 'dong']	lokasinya tambah dong
2	['drvr', 'nya', 'ramah']	drvr nya ramah
3	['sangan', 'bagus', 'dan', 'baik']	sangan bagus dan baik
4	['makasih', 'ya', 'driver']	terima kasih iya driver
5	['ok', 'trmks']	oke terima kasih
6	['sangat', 'membantu', 'dan', 'sopir', 'nya', 'sang', 'at', 'ramah']	sangat membantu dan sopir nya sang di ramah
7	['tolong', 'perbaiki', 'masa', 'sy', 'nungguin', 'sampe', 'menit', 'y', 'udah', 'sy', 'batalin', 'aja', 'cari', 'yg', 'sebelah']	tolong perbaiki masa saya menunggu sampai menit ya sudah saya batal saja cari yang sebelah
8	['sangat', 'memuaskan']	sangat memuaskan
...	...	...
10290	['terimakasih', 'banyak', 'supirnya', 'sangat', 'santun', 'dan', 'baik', 'banget']	terima kasih banyak supirnya sangat santun dan baik banget
10291	['bagusss']	bagus

### 3.1.4 Tahap *Stopword Remove*

Tahap *stopword remove* adalah tahap untuk menghilangkan kata yang umum dan tidak memiliki arti tertentu, seperti kata konjungsi atau kata hubung dan kata adverbial atau kata keterangan yang tidak memiliki dampak pada tahap analisis mining. Seperti kata “saya”, “dia”, “bawah”, “dengan”, “nya”, “itu”, “ini”, dan kata lainnya. Tahap *stopword remove* menggunakan kamus *stopword* yang mengacu dengan *stoplist* pada penelitian [21] dan kata-kata lainnya merupakan hasil dari pengumpulan yang dilakukan oleh peneliti. Berikut merupakan hasil dari tahap *stopword remove* menyisihkan data akhir sebanyak 4269 *raw data* sebagaimana pada Tabel 4.

### 3.1.5 Tahap *Stemming*

Proses *stemming* pada penelitian ini menggunakan *library Sastrawi* yang tersedia di Python. Proses ini bertujuan untuk mengubah kata ke bentuk dasarnya sebagaimana aturan dalam Kamus Besar Bahasa Indonesia (KBBI), misalnya seperti kata “menunggu” menjadi kata dasar “tunggu”; kata membantu menjadi kata “bantu”; dan lain sebagainya. Adapun hasil dari tahap *stemming* tertera pada Tabel 4.

**Tabel 4.** Hasil *Stopword Remove* dan *Stemming*

No.	Stopword Remove	Stemming
1	lokasinya	lokasi
2	drvr ramah	drvr ramah
3	bagus baik	bagus baik
4	terima kasih driver	terima kasih driver
5	oke terima kasih	oke terima kasih
6	membantu sopir sang ramah	bantu sopir sang ramah
7	tolong perbaiki menunggu batal cari sebelah	tolong baik tunggu batal cari belah
8	memuaskan	muas
...	...	...
10290	terima kasih supirnya santun baik	terima kasih supirnya santun baik
10291	bagus	bagus

Berdasarkan hasil dari proses casefolding hingga proses stemming diperoleh 269 record data yang *missing values* atau *null*/kosong, hal ini disebabkan data tersebut tidak memiliki makna atau informasi berguna berdasarkan hasil praproses, sehingga data akhir yang dapat digunakan untuk proses selanjutnya berjumlah 10022 *record* data ulasan.



### 3.1.6 Tahap Pelebelan

Data ulasan merupakan data teks yang bersifat tidak terstruktur dan tidak memiliki kelas/label, agar dapat digunakan dalam proses klasifikasi dengan algoritma K-NN, maka perlu dilakukan proses pelebelan data. Proses pelebelan data memerlukan kamus sentimen kata positif dan kata negatif. Kamus kata negatif dan positif yang digunakan pada penelitian ini adalah kamus berdasarkan rujukan penelitian [22] dan dilengkapi dengan kamus yang terdapat pada *repository* milik Ramaprakoso (2017) [23] serta berdasarkan tinjauan Pustaka oleh peneliti. Adapun tahapan dalam melakukan pelebelan data teks adalah sebagai berikut [24]:

- Menentukan term yang mengacu dan mampu merepresentasikan kata positif dan negatif, dalam hal ini adalah adanya kamus sentimen kata positif dan kata negatif. Kata positif adalah kata yang menggambarkan sanjungan, pujian, atau yang bermakna baik dan terpuji, sedangkan kata negatif merupakan kata yang merepresentasikan kata yang memiliki arti tidak atau kurang baik yang biasanya mengarah pada antonim atau negasi dari kata-kata pujian atau positif.
- Melakukan penghitungan jumlah kata positif dan negatif pada setiap kalimat, dimana ketika jumlah kata positif lebih besar dari jumlah kata negatif pada kalimat, maka kalimat tersebut masuk ke kelas sentimen positif, sebaliknya jika jumlah kata positif lebih kecil dari jumlah kata negatif pada kalimat, maka Kelas kalimat tersebut masuk ke kelas sentimen negatif. Adapun jika jumlah kata positif sama dengan jumlah kata negatif dalam sebuah kalimat, maka kalimat tersebut masuk ke dalam kelas sentimen netral.

Tabel 7. Data Hasil Pelebelan

No	Stemming	Count	Sentimen
1	lokasi	0	netral
2	drvr ramah	1	positif
3	bagus baik	2	positif
4	terima kasih driver	2	positif
5	oke terima kasih	3	positif
6	bantu sopir sang ramah	2	positif
7	tolong baik tunggu batal cari belah	-1	negatif
8	muas	0	netral
...	...	...	...
10021	terima kasih supirnya santun baik	4	positif
10022	bagus	1	positif

### 3.1.7 Tahap Pembobotan TF-IDF

Data hasil *stemming* akan digunakan sebagai acuan untuk membuat *fitur* data. Untuk menghasilkan *fitur* perlu menghitung bobot pada setiap term dengan metode perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF akan menghasilkan nilai dari setiap kata yang sudah diekstrak sebelumnya. Untuk menghitung bobot pada tiap *fitur* data dengan *library scikit-learn* pada bahasa pemrograman python menggunakan modul *TfidfVectorizer* dan *CountVectorizer*. Hasil proses pembobotan menghasilkan potongan term atau kata yang berfungsi sebagai atribut sebanyak 3368 term sebagaimana yang tertera pada Tabel 8.

Tabel 8. Hasil Pembobotan TF-IDF

No	Term/Kata														
	abai	abal	abang	abg	abu	ac	aceh	acnya	acs	acuh	...	zendo	zone	zonk	zos
	1	2	3	4	5	6	7	8	9	10	...	3365	3366	3367	3368
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6714	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0
6715	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10021	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10022	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

### 3.2 Pembagian Data Latih dan Data Uji

Untuk membuat pemodelan klasifikasi sentimen pada Data Ulasan, data yang telah melalui tahap pra-proses kemudian dibagi menjadi data latih dan data uji. Data latih adalah data yang digunakan untuk membuat model klasifikasi sedangkan data uji adalah data yang digunakan untuk menguji model yang telah dibangun dengan data latih. Adapun proporsi pembagian data pada penelitian ini menggunakan teknik pembagian data *Hold-out*. Proses pembagian data menggunakan modul *selection* yang terdapat pada *library scikit learn*, adapun untuk menentukan proporsi pembagian data yang paling optimal, dilakukan beberapa percobaan pembagian data sebagaimana tertera pada Tabel 9.

Tabel 9. Percobaan Pembagian Data

No.	Percobaan	Data Latih		Data Uji	
		Persentase	Jumlah Data	Persentase	Jumlah Data
1	P-1	50%	5011	50%	5011
2	P-2	60%	6013	40%	4009
3	P-3	70%	7015	30%	3007
4	P-4	75%	7516	25%	2506
5	P-5	80%	8017	20%	2005
6	P-6	85%	8518	15%	1504
7	P-7	90%	9019	10%	1003
8	P-8	95%	9620	5%	502

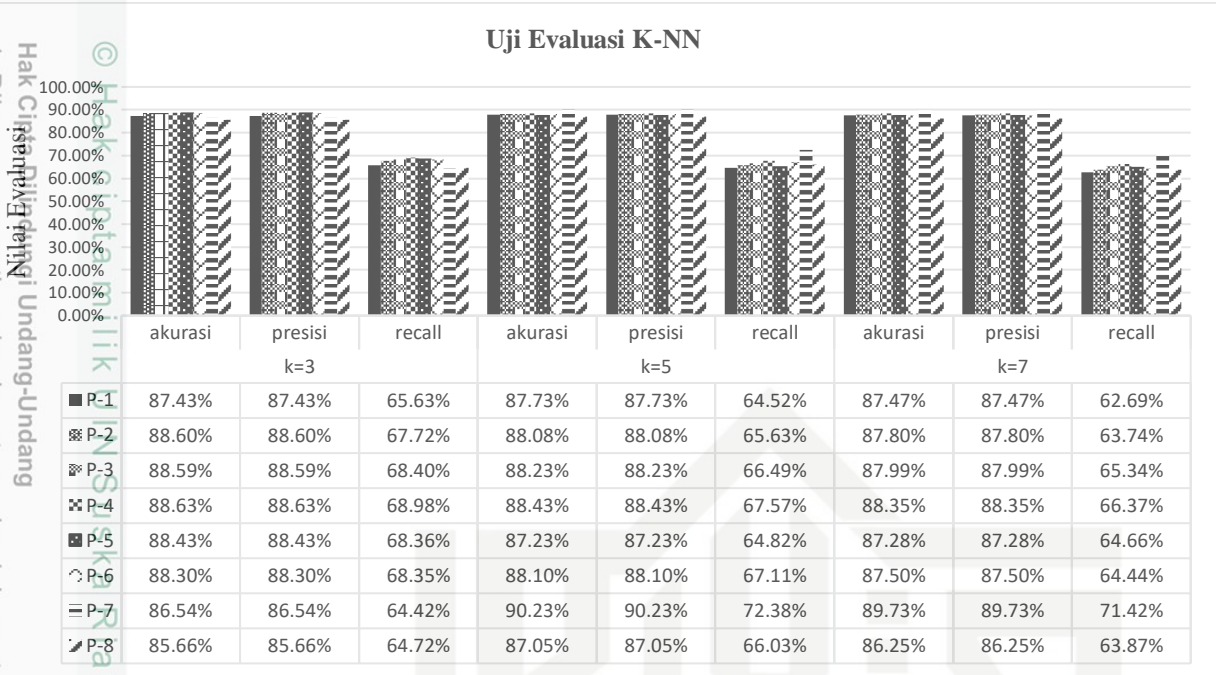
Karena sebaran sentimen ulasan yang tidak seimbang (*inbalanced*) dan kemungkinan memiliki pengaruh terhadap model klasifikasi, maka pemodelan akan dilakukan dengan menerapkan teknik rekonstruksi data SMOTE yang bertujuan untuk menyeimbangkan jumlah sentimen dengan meningkatkan jumlah sentimen minoritas melalui rekonstruksi ulang sentimen minoritas secara acak sehingga memiliki jumlah yang sama dengan sentimen mayoritas. Dalam hal ini kelas negatif dan netral akan memiliki jumlah yang sama dengan sentiment positif sebanyak 7978 ulasan. Sehingga secara keseluruhan penerapan SMOTE menghasilkan jumlah dataset sebanyak 23934 ulasan. Tabel 10 merupakan pembagian data latih dan data uji pada dataset hasil resampling SMOTE.

Tabel 10. Percobaan Pembagian Dataset Resampling SMOTE

No.	Percobaan	Data Latih		Data Uji	
		Persentase	Jumlah Data	Persentase	Jumlah Data
1	P-1	50%	11967	50%	11967
2	P-2	60%	14360	40%	9574
3	P-3	70%	16753	30%	7181
4	P-4	75%	17950	25%	5984
5	P-5	80%	19147	20%	4787
6	P-6	85%	20343	15%	3591
7	P-7	90%	21540	10%	2394
8	P-8	95%	22737	5%	1197

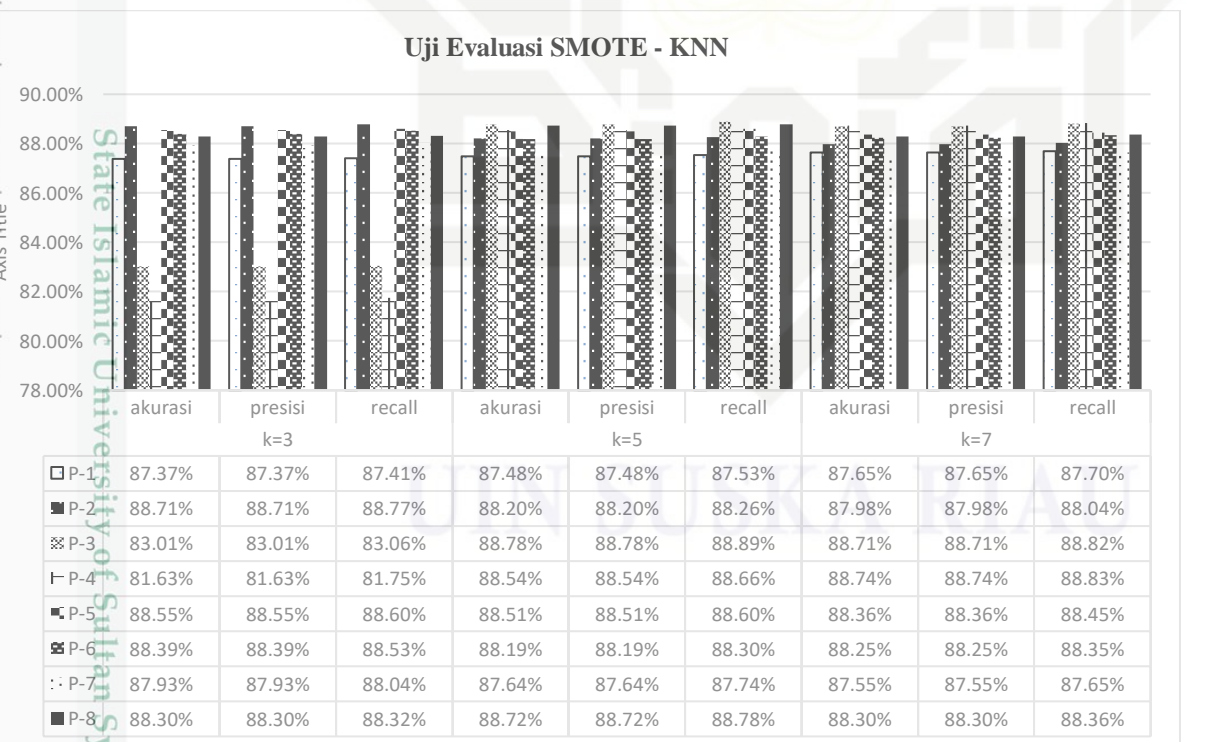
### 3.3 Pemodelan Sentimen dengan K-NN

Pemodelan klasifikasi sentimen dengan algoritma KNN menggunakan *library scikit learn* dengan modul *KNeighborsClassifier*. Adapun pemodelan pada penelitian ini menggunakan nilai parameter  $k$  atau jumlah banyaknya tetangga terdekat yang beragam yaitu  $k=3$ ,  $k=5$ , dan  $k=7$ . Untuk mengevaluasi model yang dibuat dilakukan penghitungan performa model menggunakan modul *metrics* pada *python*.



Gambar 2. Hasil Uji Evaluasi K-NN dengan  $k$  beragam

Berdasarkan hasil evaluasi pada Gambar 2 terhadap tiap parameter  $k$ , dapat disimpulkan bahwa percobaan dengan performa terbaik diperoleh pada percobaan dengan ID Percobaan P-7 yang menggunakan skenario 90% data latih atau 9019 record data dan 10% data uji atau 1003 record data dan menggunakan nilai  $k = 5$  dengan nilai akurasi sebesar 90,23%; nilai precision sebesar 90,23%, dan nilai recall sebesar 72,38%. Adapun model yang menghasilkan performa terendah diperoleh pada percobaan dengan ID P-8 dengan proporsi data latih sebesar 95% atau 9620 data ulasan dan data uji sebesar 5% atau 502 data ulasan dan menggunakan nilai  $k = 3$  dengan nilai akurasi dan precision sebesar 85,66% dan nilai recall sebesar 64,72%. Adapun hasil pemodelan K-NN pada dataset *resampling* SMOTE tertera pada Gambar 3.



Gambar 3. Hasil Uji Evaluasi K-NN dengan  $k$  beragam pada Data *Resampling*

Berdasarkan nilai evaluasi pada setiap skenario percobaan dengan data *resampling* SMOTE, model dengan performa terbaik menghasilkan nilai akurasi 88,78%; presisi 88,78% ; dan recall 88,89% pada percobaan dengan nilai  $k = 5$  dan 70% data latih dan 30% data uji (P-3). Sebaliknya, model dengan performa terendah diperoleh pada percobaan dengan nilai  $k = 3$  serta data uji dan data latih berturut-turut 75% dan 25% (P-4) dengan nilai akurasi; precision; dan nilai recall

1. Dilarang mengutip sebagian atau seluruh karya tulis ini untuk dipublikasikan di media massa atau jurnal ilmiah, kecuali dengan izin dan menyebutkan sumber.  
 2. Dilarang mengutip sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sebesar 81,63%; 81,63%; dan 81,75%. Berdasarkan 2 pemodelan algoritma KNN tanpa dan dengan menggunakan teknik resampling SMOTE dapat disimpulkan bahwa pemodelan dengan nilai evaluasi terbaik diperoleh pada pemodelan tanpa menerapkan SMOTE dalam proses rekonstruksi data latih dan data uji, dimana penggunaan SMOTE tidak berpengaruh dalam meningkatkan nilai evaluasi pada klasifikasi data ulasan pengguna Maxim.

### 3.4 Evaluasi Model Terbaik

Evaluasi model dilakukan pada model yang menghasilkan nilai performa terbaik, dalam hal ini adalah model dengan parameter  $k=3$  serta proporsi data latih sebesar 90% dan data uji sebesar 10%. Berikut merupakan selisih perbedaan kelas berdasarkan kelas pada data uji secara aktual dan kelas pada data uji berdasarkan hasil prediksi dengan model KNN.

Tabel 11. Hasil Uji Model Terbaik

Kelas Data	Sentimen			Jumlah Data Uji
	Positif	Netral	Negatif	
Kelas Aktual	776	145	82	1003
Kelas Prediksi	795	173	35	1003

Jika dilihat dari Tabel 11 dapat disimpulkan bahwa model cenderung menghasilkan kelas positif dan kelas netral. Hal ini dapat dilihat dari hasil prediksi yang menghasilkan sentimen netral lebih banyak dari kelas aktual, begitu pula dengan sentimen positif pada hasil prediksi. Sebaliknya sentimen negatif justru mengalami penurunan dari data aktual berdasarkan hasil prediksi dari yang semula berjumlah 82 menjadi 35 ulasan negatif ketika diprediksi menggunakan model KNN. Sehingga, dapat disimpulkan bahwa pada percobaan dengan parameter  $k=3$  terdapat 79,26% sentimen positif, 17,25% sentimen netral, dan 3,49% sentimen negatif.

### 3.5 Analisis Sentimen Ulasan

Sentimen dapat dianalisis dengan melakukan visualisasi menggunakan *library wordcloud* dan *matplotlib* yang tersedia di *python*. Berdasarkan proses visualisasi tersebut, diperoleh wordcloud dari sentimen netral ulasan pengguna aplikasi Maxim tertera pada Gambar 4.



Gambar 4. Wordcloud Sentimen Netral

Kata yang mendominasi pada sentimen netral pengguna aplikasi Maxim diantaranya kata aplikasi; driver; maxim; muas/puas; map; daftar; baik; order; drivernya; dan harga. Adapun jumlah kemunculan kata pada sentimen netral, positif, dan negatif secara lebih rinci tertera pada Tabel 12.

Tabel 12. Total Kemunculan Kata pada Ulasan Netral, Positif, dan Negatif

Netral			Positif			Negatif		
No	Word	Jumlah	No	Word	Jumlah	No	Word	Jumlah
1	aplikasi	135	1	terima kasih	1402	1	driver	275
2	driver	134	2	bagus	969	2	aplikasi	220
3	maxim	107	3	mantap	765	3	maxim	157
4	muas	100	4	baik	716	4	susah	118
5	map	62	5	cepat	509	5	order	103
6	daftar	50	6	driver	349	6	map	76
7	baik	49	7	oke	334	7	lama	75
8	order	49	8	maxim	311	8	tolong	67
9	drivernya	48	9	murah	278	9	error	63
10	harga	46	10	baik ramah	263	10	gagal	58

Berbeda halnya dengan sentimen netral, berikut merupakan visualisasi kata pada sentimen positif dari ulasan pengguna aplikasi Maxim pada Gambar 5.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

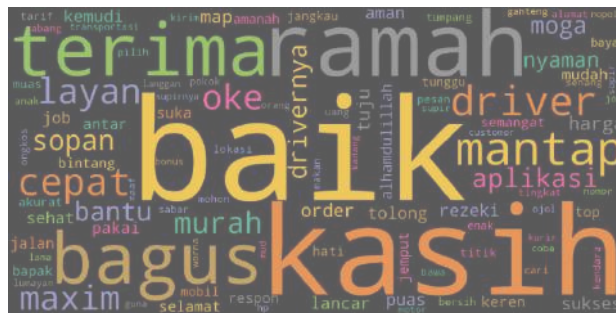
yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

yang merupakan bagian atau seluruhnya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.



**Gambar 5.** Wordcloud Sentimen Positif

Dapat disimpulkan bahwa 10 kata yang paling mendominasi pada sentimen positif ulasan pengguna aplikasi Maxim diantaranya kata terima kasih; bagus; mantap; baik; cepat; driver; oke; maxim; murah; dan baik ramah. Adapun visualisasi kata pada data ulasan yang bersentimen negatif tertera pada Gambar 6.



**Gambar 6.** Wordcloud Sentimen Negatif

Mengacu pada Gambar 5 dapat disimpulkan bahwa 10 kata yang mendominasi pada sentimen negatif yaitu kata driver; aplikasi; maxim; susah; order; map; lama; tolong; error; dan gagal. Jika dianalisis lebih lanjut, terdapat kemunculan kata yang sama antara satu kelas dengan kelas lainnya pada 10 kata dengan kemunculan tertinggi sebagaimana yang tertera pada Tabel 13.

**Tabel 13.** Kemunculan Kata yang Sama

No.	Kata	Sentimen		
		Netral	Positif	Negatif
1	aplikasi	√	-	√
2	driver	√	√	√
3	maxim	√	√	√
4	map	√	-	√
5	baik	√	√	-
6	order	√	-	√

Berdasarkan Tabel 13 dapat ditarik beberapa kesimpulan, yaitu:

- Terdapat 2 kata yang sering muncul dalam semua kelas sentimen yaitu kata “driver” dan “maxim”
- Kata “aplikasi”, “map”, dan “order” sering muncul pada kelas sentimen netral dan negatif
- Kata “baik” sering muncul pada kelas sentimen netral dan positif

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa performa terbaik pada percobaan dengan pembagian data latih dan data uji, serta nilai  $k$  yang beragam diperoleh pada percobaan dengan pembagian data latih 90%, data uji 10% dan menggunakan nilai  $k = 5$  dengan nilai akurasi, presisi, dan recall berturut-turut sebesar 90,23%; 90,23%; dan nilai recall 72,38%. Klasifikasi sentimen dengan model terbaik menggunakan parameter  $k=3$  menghasilkan 79,26% sentimen positif, 17,25% sentimen netral, dan 3,49% sentimen negatif, selain itu terdapat 2 kata dengan frekuensi kemunculan paling tinggi di ketiga kelas sentimen yaitu kata “driver” dan “maxim”.

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sunan Kalijaga Kasim Riau

## REFERENCES

- [1] A. Popy, “RESOLUSI KONFLIK ANTARA PENGEMUDI TRANSPORTASI ONLINE DENGAN PENGEMUDI TRANSPORTASI KONVENSIIONAL (Studi Hambatan Struktural Penyelesaian Konflik antara Pengemudi Gojek, Pengemudi Angkutan Kota (Angkot), Pengemudi Ojek Pangkalan di Kota Padang),” Universitas Andalas, 2018.
- [2] Google and TEMASEK, “e-Conomy SEA 2018 Southeast Asia’s internet economy hits an inflection point.” pp. 1–32, 2018.
- [3] J. P. Kompetitif, “Laporan Survei Penetrasi & Profil Perilaku Pengguna Internet Indonesia 2018,” *J. Pemasar. Kompetitif*, vol. 1, no. 3, pp. 71–80.
- [4] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [5] M. T. Diwandanu and L. M. Wisudawati, “Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K Nearest Neighbors,” *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 1–16, 2023, doi: 10.35760/ik.2023.v28i1.7909.
- [6] M. S. Alrajak, I. Ernawati, and I. Nurlaili, “Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN),” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 1, no. 2, pp. 110–122, 2020.
- [7] B. Liu, *Sentiment analysis and opinion mining*. Springer Nature, 2022.
- [8] O. Somantri and D. Apriliani, “Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 537, 2018, doi: 10.25126/jtiik.201855867.
- [9] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [10] N. Tri Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [11] N. S. N. Salam, A. A. Supianto, and A. R. Perdanakusuma, “Analisis Sentimen Opini Mahasiswa Terhadap Saran Kuesioner Penilaian Kinerja Dosen dengan Menggunakan TF-IDF dan K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 6148–6156, 2019.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, pp. 115–123, 2020.
- [13] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [14] A. N. S. Rahayu, T. I. Hermanto, and I. M. Nugroho, “Sentiment Analysis Using K-Nearest Neighbor Based on Particle Swarm Optimization According To Sunscreen’S Reviews,” *J. Tek. Inform.*, vol. 3, no. 6, pp. 1639–1646, 2022, doi: 10.20884/1.jutif.2022.3.6.425.
- [15] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, “Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.
- [16] A. Deviyanto and M. D. R. Wahyudi, “Penerapan analisis sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, pp. 1–13, 2018.
- [17] S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, “Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [18] R. Siringoringo and J. Jamaludin, “Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online,” *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [19] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [20] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [21] D. Era, S. Andryana, and A. Rubhasy, “Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 7, no. 1, pp. 263–272, 2023.
- [22] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, “Tweets responding to the Indonesian Government’s handling of COVID-19: Sentiment analysis using SVM with normalized poly kernel,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, p. 112, 2020.
- [23] R. Prakoso, “kbba,” 2017. <https://github.com/ramaprakoso/analisis-sentimen/blob/master/kamus/kbba.txt>.
- [24] M. ARIFAH, “Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan Lexicon-Based Dan Naive Bayes,” 2019.

1. Dilarang mengutip, sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber atau dengan cara lain.  
 2. Dilarang mengumumkkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

© Restu Ramadhan, 2023. This Journal is licensed under a Creative Commons Attribution 4.0 International License

Prof. Dr. Sultan Syarif Kasim Riau



# LAMPIRAN A

## SOURCE CODE PENELITIAN

### Akuisis Data

```

1. Akuisis Data

In [20]: import pandas as pd

In [21]: #data tanggal > bulan & - fgl > bulan &
MaxIMData = pd.read_excel("data.xlsx")
ReviewData

Out[21]:

```

	reviewid	username	userstage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent
0	16256259a-437b-419b-71570d75184	Tamara Desmora	https://p...@googleusercontent.com/kdUu.../	Lokasinya tambah dang	1.0	0.0	None	2022-09-07 17:12:02	Kami mohon maaf masalah ini akan segera
1	4d70235-8f58-448b-b7a5-8083a858e84	Arisman Men	https://p...@googleusercontent.com/tp.../	diurinya aneh	5.0	0.0	3.134	2022-09-07 17:11:59	Kami senang ada perhatian Anda terima kasih
2	759e25cb-082-466a-88d5-ct1a3b9e4d2	Hesty Yuliana	https://p...@googleusercontent.com/da.../	siangan bagus dan baik	5.0	0.0	3.141	2022-09-07 17:07:08	Kami senang ada perhatian Anda terima kasih
3	624953c-ba95-454c-3f43-84d56566a3e	Dani Daman	https://p...@googleusercontent.com/ia.../	rekaah ya dmer	5.0	0.0	3.141	2022-09-07 17:03:09	Terimakasih dan berant urtuk kami terima kasih
4	041ec58a-478-4a47-8444-d2bc5864681	Terran Supriatman	https://p...@googleusercontent.com/ia.../	Ok thank	5.0	0.0	3.141	2022-09-07 17:02:51	Terimakasih dan berant urtuk kami terima kasih

### 2. Cleaning Data

```

2. Cleaning Data

In [22]: import re
import nltk

In [23]: MaxIMDatacf = pd.DataFrame(MaxIMData['content']).astype(str)
@MaxIMDatacf

In [62]: MaxIMDatacf.content.isnull().sum()
Out[62]: 0

In [70]: with open('kbbi.txt') as f:
    kbbi_slngpwr = dict(line.strip().split(None, 1) for line in f)

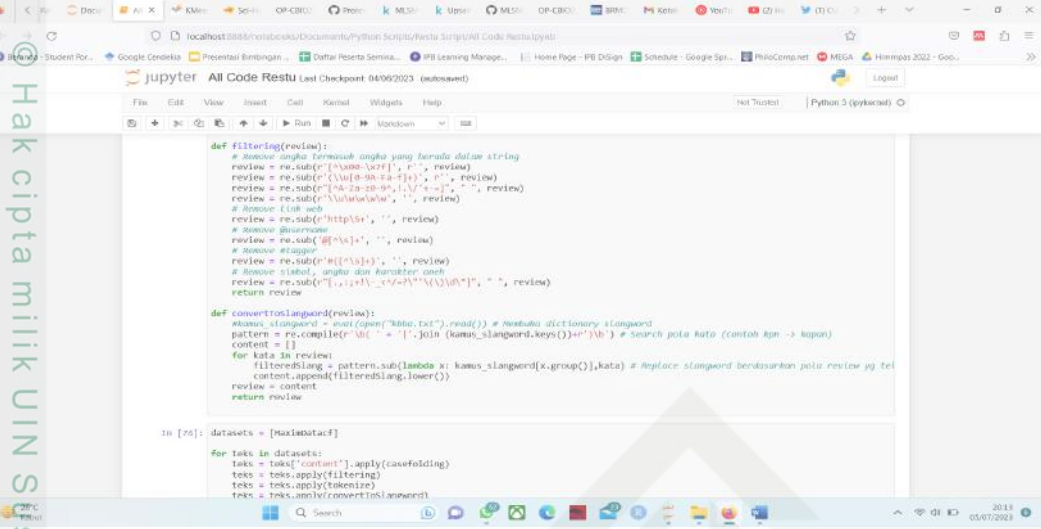
In [73]: #buat function untuk cleaning data
def casefolding(review):
    review = review.lower()
    return review

def tokenize(review):
    token = nltk.word_tokenize(review)
    return token

def filtering(review):
    # Remove angka termasuk angka yang berada diluar string
    review = re.sub(r'[\d+|\s+]', '', review)
    review = re.sub(r'([A-Z])', lambda m: ' '+m.group(1).lower(), review)

```

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

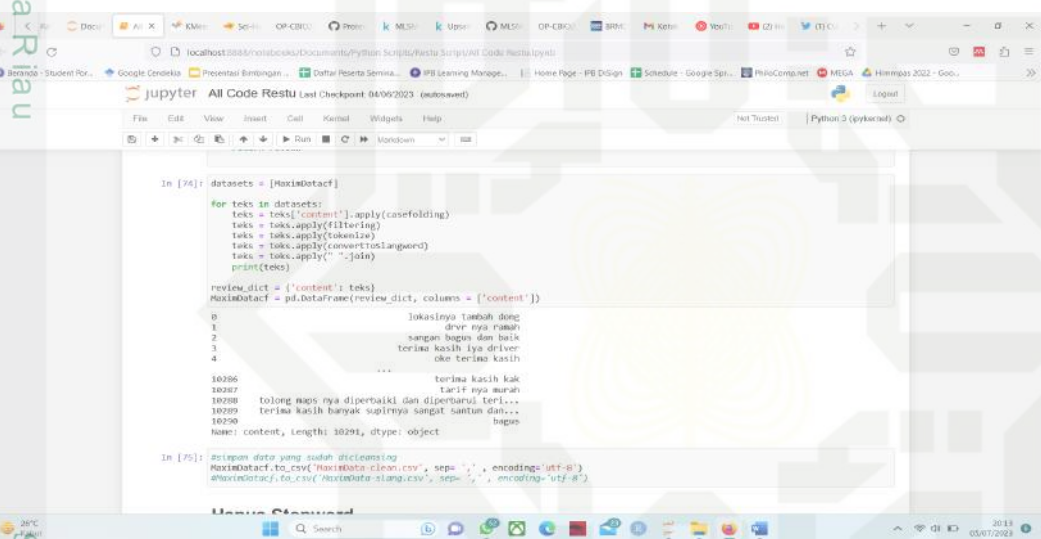


```
def filtering(review):
    # Remove angka termasuk angka yang berada dalam string
    review = re.sub(r"[0-9]+", "", review)
    review = re.sub(r"[a-zA-Z0-9]+", "", review)
    review = re.sub(r"[^a-zA-Z0-9]+", "", review)
    review = re.sub(r"[^a-zA-Z0-9]+", "", review)
    # Remove URL web
    review = re.sub(r"http://", "", review)
    # Remove @username
    review = re.sub(r"@[a-zA-Z0-9]+", "", review)
    # Remove hashtag
    review = re.sub(r"#[a-zA-Z0-9]+", "", review)
    # Remove tanda baca dan karakter spesial
    review = re.sub(r"[^\w\s]", "", review)
    return review

def converttolongword(review):
    #kamus_slangword = json.load(open('kamus.txt')) # Membuat dictionary slangword
    pattern = re.compile(r"([a-zA-Z0-9]+)") # search pola kata (contoh apa -> kapan)
    content = []
    for kata in review:
        filteredslang = pattern.sub(lambda x: kamus_slangword[x.group()], kata) # Replace slangword berdasarkan pola review yg ter
        content.append(filteredslang.lower())
    return content

In [76]: datasets = [MaxisDataCF]

for teks in datasets:
    teks = teks['content']
    teks = teks.apply(casefolding)
    teks = teks.apply(filtering)
    teks = teks.apply(tokenize)
    teks = teks.apply(converttolongword)
    print(teks)
```



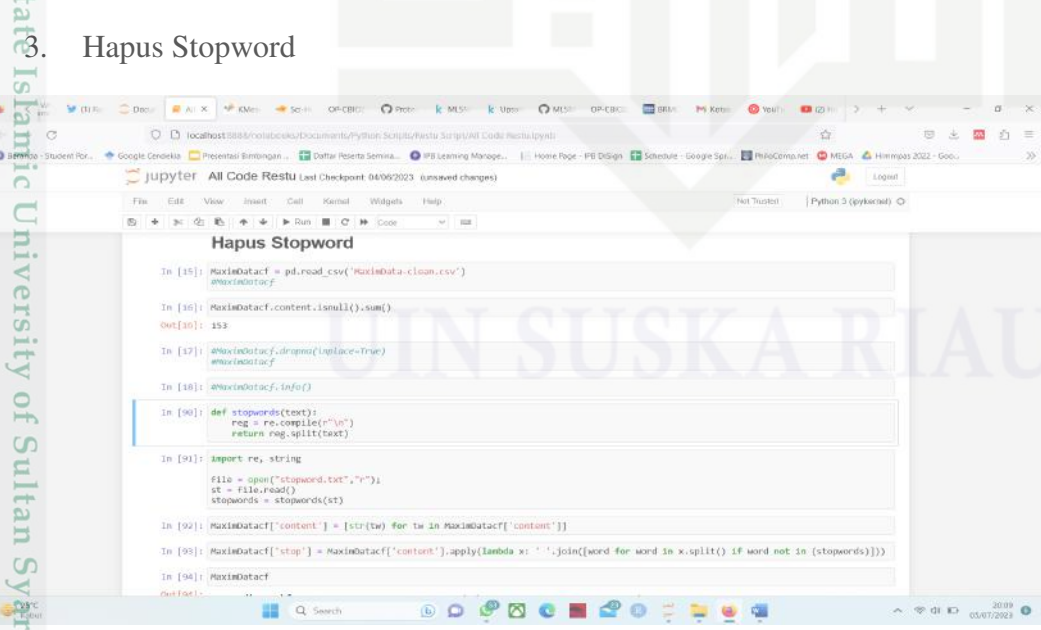
```
In [76]: datasets = [MaxisDataCF]

for teks in datasets:
    teks = teks['content']
    teks = teks.apply(casefolding)
    teks = teks.apply(filtering)
    teks = teks.apply(tokenize)
    teks = teks.apply(converttolongword)
    teks = teks.apply(" ".join)
    print(teks)

review_dict = {'content': teks}
MaxisDataCF = pd.DataFrame(review_dict, columns = ['content'])

0      lokasinya tambah dong
1      driv nya murah
2      sangat bagus dan baik
3      terima kasih iya driver
4      oke terima kasih
...
10286      terima kasih kak
10287      tacif nya murah
10288      tolong maps nya diperbaiki dan dipersarui teri...
10289      terima kasih banyak supirnya sangat santun dan...
10290      bagus
Name: content, Length: 10291, dtype: object

In [75]: siapkan data yang sudah di cleaning
MaxisDataCF.to_csv('MaxisData-clean.csv', sep=';', encoding='utf-8')
MaxisDataCF.to_csv('MaxisData-slang.csv', sep=';', encoding='utf-8')
```



```
Hapus Stopword

In [15]: MaxisDataCF = pd.read_csv('MaxisData-clean.csv')
MaxisDataCF

In [16]: MaxisDataCF.content.isnull().sum()
Out[16]: 153

In [17]: MaxisDataCF.dropna(inplace=True)
MaxisDataCF

In [18]: MaxisDataCF.info()

In [90]: def stopwords(text):
    reg = re.compile(r"(\w+)", re.UNICODE)
    return reg.split(text)

In [91]: import re, string
file = open("stopword.txt", "r")
st = file.read()
stopwords = stopwords(st)

In [92]: MaxisDataCF['content'] = [str(tw) for tw in MaxisDataCF['content']]

In [93]: MaxisDataCF['stop'] = MaxisDataCF['content'].apply(lambda x: " ".join([word for word in x.split() if word not in stopwords]))

In [94]: MaxisDataCF
```

### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau





#### 4. Data Stemming



Hak cipta milik UIN Suska Riau

##### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

stemming

In [10]: from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
In [11]: MaxInDatacf = pd.read_csv('MaxInData-stopword.csv')
          maxInDatacf
In [12]: maxInDatacf.index()
In [13]: maxInDatacf.info()
In [14]: #hapus row dan jika ada yang kosong
          maxInDatacf.dropna(inplace=True)
          maxInDatacf

In [10]: factory = StemmerFactory()
          stemmer = factory.create_stemmer()

In [10]: MaxInDatacf['stop'] = [str(tw) for tw in maxInDatacf['stop']]

In [10]: MaxInDatacf
Out[10]:
   Unnamed: 0  content  stop
0           0  lokasi tembak dng  lokasi
1           1  dir nya rumah  dir rumah
2           2  sangat bagus dan baik  bagus baik
3           3  terima kasih sa diwar  terima kasih diwar
    
```

```

10291 rows x 3 columns

In [105]: maxInDatacf = pd.DataFrame({'stemming': [stemmer.stem(t) for t in maxInDatacf.stop[1:10290]]})

In [105]: #hapus baris stemming
          maxInDatacf.to_csv('MaxInData-stem.csv', encoding='utf-8', index=False)

In [105]: MaxInDatacf
Out[105]:
   stemming
0         lokasi
1         dir rumah
2         bagus baik
3         terima kasih diwar
4         oke terima kasih
...
10286        terima kasih
10287         leaf rumah
10288  isdng mpaq baik baau terima kasih
10289        terima kasih supernya saantun baik
10290         bagus
10291 rows x 1 columns
    
```

State Islamic University of Sultan Syarif Kasim Riau



## 5. Pelabelan

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

jupyter All Code Restu Last Checkpoint: 04/06/2023 (unsaved changes)

pelabelan

In [7]: MaxIDatacf = pd.read_csv('MaxiData-stem.csv')
        @MaxIDatacf

In [8]: MaxIDatacf.isna().sum()
Out[8]:
stemming    269
dtype: int64

In [9]: #hapus row nan jika ada yang kosong
MaxIDatacf.dropna(inplace=True)
@MaxIDatacf

In [10]: pos_sent = open('kamus_sentimen/positive.txt').read()
         positive_words=pos_sent.split('\n')

In [11]: neg_sent = open('kamus_sentimen/negative.txt').read()
         negative_words=neg_sent.split('\n')

In [12]: MaxIDatacf['stemming'] = [str(tw) for tw in MaxIDatacf['stemming']]

In [13]: MaxIDatacf['count'] = MaxIDatacf['stemming'].apply(lambda stemming: sum(0 + (word in positive_words) and 1) + ((word in negati

In [14]: MaxIDatacf['sentimen'] = 'netral'
         MaxIDatacf['sentimen'][(MaxIDatacf['count']>0) = 'positif'
         MaxIDatacf['sentimen'][(MaxIDatacf['count']<0) = 'negatif'
         MaxIDatacf['sentimen'][(MaxIDatacf['count']==0) = 'netral'
    
```

## 6. TF-IDF

```

jupyter All Code Restu Last Checkpoint: 04/06/2023 (unsaved changes)

TF-IDF

In [11]: from __future__ import print_function
         from __future__ import division
         from sklearn.feature_extraction.text import TfidfVectorizer,CountVectorizer

In [14]: MaxIDatacf = pd.read_csv('MaxiData-stem.csv')
         @MaxIDatacf

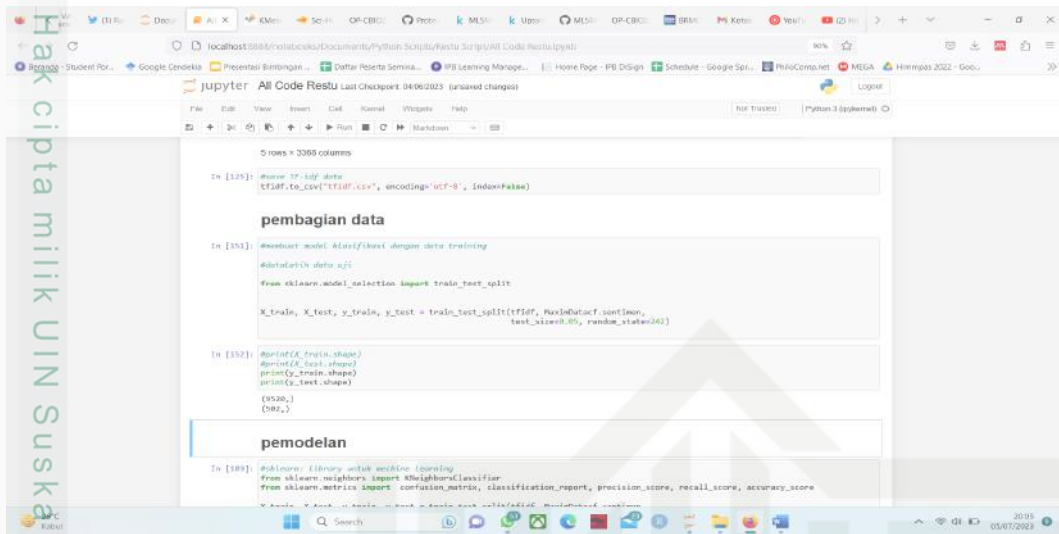
In [5]: @MaxIDatacf.sentimen

In [123]: vector = CountVectorizer()
          vector_tf = TfidfVectorizer()
          matrix_kata = vector_tf.transform(MaxIDatacf.stemming)
          matrix_kata_tf = vector_tf.fit_transform(MaxIDatacf.stemming)
          tfidf = pd.DataFrame(matrix_kata.toarray(), columns=vector_tf.get_feature_names())

In [124]: tfidf.head()

Out[124]:
          abai  abai  atang  abg  abu  ao  aseh  acnya  aos  asuh  ..  yongki  your  youtube  yth  yugi  zaman  zendo  zone  zoni  zok  zok
0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    
```

## 7. Pembagian Data Latih dan Data Uji



```

In [129]: #mereset data
tfidf.to_csv("tfidf.csv", encoding='utf-8', index=False)

pembagian data

In [131]: #membuat model KNeighborsClassifier dengan data training
#data latih data uji
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(tfidf,
                                                  y,
                                                  test_size=0.5,
                                                  random_state=24)

In [132]: #print(X_train.shape)
#print(X_test.shape)
print(y_train.shape)
print(y_test.shape)

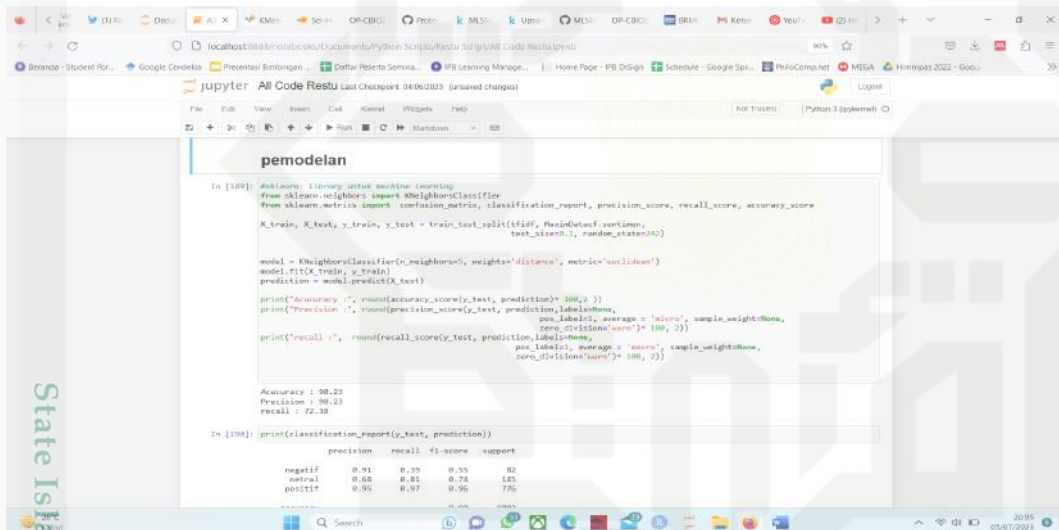
(9526,)
(4763,)

pemodelan

In [189]: #alasan: library untuk machine learning
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, classification_report, precision_score, recall_score, accuracy_score
#anda: X_train, X_test, y_train, y_test = train_test_split(tfidf, y, test_size=0.5, random_state=24)

```

## 8. Pemodelan Data



```

pemodelan

In [189]: #alasan: library untuk machine learning
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, classification_report, precision_score, recall_score, accuracy_score
X_train, X_test, y_train, y_test = train_test_split(tfidf,
                                                  y,
                                                  test_size=0.5,
                                                  random_state=24)

model = KNeighborsClassifier(n_neighbors=5, weights='distance', metric='euclidean')
model.fit(X_train, y_train)
prediction = model.predict(X_test)

print("Accuracy :", round(accuracy_score(y_test, prediction)* 100, 2))
print("Precision :", round(precision_score(y_test, prediction, labels=None,
                                         pos_labels='negatif', sample_weight=None,
                                         zero_division='warn')* 100, 2))
print("recall :", round(recall_score(y_test, prediction, labels=None,
                                    pos_labels='negatif', sample_weight=None,
                                    zero_division='warn')* 100, 2))

Accuracy : 90.23
Precision : 90.23
recall : 72.38

In [198]: print(classification_report(y_test, prediction))

precision    recall  f1-score   support

negatif    0.91     0.39     0.59     82
netral     0.68     0.81     0.74     125
positif    0.95     0.97     0.96     776

```

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



## 9. Visualisasi Data

```

In [206]: from wordcloud import WordCloud

In [207]: negatif_df = MaxDiffDatacf.loc[MaxDiffDatacf['sentimen'] == 'negatif']
          positif_df = MaxDiffDatacf.loc[MaxDiffDatacf['sentimen'] == 'positif']
          netral_df = MaxDiffDatacf.loc[MaxDiffDatacf['sentimen'] == 'netral']

In [219]: text1 = ".join(title for title in negatif_df.stemming)
          text2 = ".join(title for title in positif_df.stemming)
          text3 = ".join(title for title in netral_df.stemming)

          # Creating word_cloud with text as argument in .generate() method
          word_cloud1 = WordCloud(background_color = "black", mode = "RGB", width = 1000, height = 800, colormap = "dark2", collections =
          word_cloud2 = WordCloud(background_color = "black", mode = "RGB", width = 1000, height = 800, colormap = "dark2", collections =
          word_cloud3 = WordCloud(background_color = "black", mode = "RGB", width = 1000, height = 800, colormap = "dark2", collections =

          #Display the generated word cloud
          plt.imshow(word_cloud3, interpolation='nearest')
          plt.axis("off")
          plt.show()
  
```

## 10. Balancing Data dengan SMOTE

### balancing data using SMOTE

```

In [11]: import numpy as np
         from imblearn.over_sampling import SMOTE
         from collections import Counter

         sm = SMOTE(random_state=42)

         X_res, y_res = sm.fit_resample(tfidf, MaxDiffDatacf.sentimen)

         print('Resampled dataset shape %s' % Counter(y_res))

C:\Users\MSI-GF63\anaconda3\lib\site-packages\sklearn\utils\extmath.py:152: RuntimeWarning: invalid value encountered in matmul
  ret = a @ b

Resampled dataset shape Counter({'netral': 7978, 'positif': 7978, 'negatif': 7978})
  
```

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



```
#sklearn: Library untuk mechine Learning
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, classification_report, precision_score, recall_score, accuracy_score

X_train, X_test, y_train, y_test = train_test_split(X_res, y_res,
                                                    test_size=0.05, random_state=242)

model = KNeighborsClassifier(n_neighbors=7, weights='distance', metric='euclidean')
model.fit(X_train, y_train)
prediction = model.predict(X_test)

print("Acucuracy :", round(accuracy_score(y_test, prediction)* 100,2 ))
print("Precision :", round(precision_score(y_test, prediction,labels=None,
                                         pos_label=1, average = 'micro', sample_weight=None,
                                         zero_division='warn')* 100, 2))
print("recall :", round(recall_score(y_test, prediction,labels=None,
                                     pos_label=1, average = 'macro', sample_weight=None,
                                     zero_division='warn')* 100, 2))
```

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## DAFTAR RIWAYAT HIDUP



Penulis bernama Restu Ramadhan lahir di Kota Pekanbaru pada tanggal 30 Desember 1997. Penulis merupakan anak kedua dari dua bersaudara dari pasangan Bapak Yuswardi dan Ibu Witneti Hasri. Penulis beralamat di Jalan Kereta Api Ujung no. 12 B, kelurahan Tangkerang Tengah, kecamatan Marpoyan Damai, kota Pekanbaru, Riau. Kontak Penulis, No. Hp 082284807364 dan alamat email: 11653101468@students.uin-suska.ac.id.

Penulis menyelesaikan pendidikan dasar di SD Negeri 025 Pekanbaru pada tahun 2010, kemudian menyelesaikan sekolah menengah di MTs Negeri Andalan Pekanbaru pada tahun 2013, kemudian penulis melanjutkan serta menyelesaikan sekolah menengah atas di SMK Negeri 4 Pekanbaru pada tahun 2016, lalu penulis meneruskan pendidikan ke jenjang S1 pada tahun 2016 di Universitas Islam Sultan Syarif Kasim Riau pada Fakultas Sains dan Teknologi dengan Program Studi Sistem Informasi.

Pada Tahun 2018, penulis telah menyelesaikan Kerja Praktek, pada tahun 2019 penulis juga telah menyelesaikan Kuliah Kerja Nyata (KKN) di Desa Tanjung Pauh Kecamatan Singingi Hilir Kabupaten Kuantan Singingi Riau. Dan pada tahun 2023 Penulis telah menyelesaikan masa studinya dengan judul skripsi: "Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor".

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.