

# KLASIFIKASI ULASAN PENGGUNA APLIKASI: STUDI KASUS APLIKASI IPUSNAS PERPUSTAKAAN NASIONAL REPUBLIK INDONESIA (PNRI)

Andina Septiani<sup>1)</sup>, Indra Budi<sup>2)</sup>

<sup>1,2</sup>Program Studi Magister Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Indonesia  
Jl. Kenari 2 No.4, RW.5, Kenari, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10430  
e-mail: [andina89septiani@gmail.com](mailto:andina89septiani@gmail.com)<sup>1)</sup>, [indra@cs.ui.ac.id](mailto:indra@cs.ui.ac.id)<sup>2)</sup>

## ABSTRAK

Menurunnya jumlah tren pengguna baru aplikasi iPusnas berpengaruh terhadap penurunan pencapaian nilai target laporan LKIP Pujasintara PNRI 2020 – 2024. Hal tersebut berkaitan dengan nilai peringkat ulasan pengguna aplikasi di Google Playstore yang dinilai masih lebih rendah dibandingkan aplikasi sejenis lainnya. Electronic Word of Mouth (EWOM) yang sangat berpengaruh terhadap keputusan calon pengguna baru aplikasi dalam mempertimbangkan aplikasi terbaik yang sejenis, karena melibatkan tinjauan nilai peringkat dan ulasan pengguna. Beberapa penelitian terdahulu membuktikan bahwa kesulitan selalu dihadapi ketika melakukan analisis atau penggalian informasi penting dalam ulasan pengguna aplikasi secara manual. Analisis ulasan sangat berguna untuk mengembangkan fitur layanan aplikasi agar dapat meningkatkan kepuasan pengguna dan peringkat nilai aplikasi, sehingga diperlukan alat bantu klasifikasi ulasan pengguna secara otomatis dengan mencari model terbaik yang sesuai. Penelitian ini menerapkan metodologi CRISP-DM, tetapi hanya sampai tahap evaluasi. Algoritma klasifikasi yang digunakan adalah Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), serta kombinasi fitur tf-idf unigram, bigram, dan trigram. Hasil penelitiannya adalah kombinasi fitur tf-idf unigram (F1) dengan algoritma SVM mencapai nilai terbaik untuk setiap nilai evaluasi precision, recall, dan f1-score masing-masing sebesar 87%. Nilai evaluasi terendah precision 55% dari hasil kombinasi fitur F2 dengan SVM, recall 42% dan f1-score 32% dari kombinasi fitur F3 dengan logistic regression.

**Kata Kunci:** Klasifikasi Ulasan, Algoritma Classifier, Machine Learning, Text Mining

## ABSTRACT

The decline in the number of new user trends for the iPusnas application affects the achievement of the target value of the LKIP Pujasintara PNRI report 2020 – 2024. This is related to the rating value of application user reviews on Google Playstore which is still considered lower than other similar applications. Electronic Word of Mouth (EWOM) which is very influential on the decision of prospective new users of the application in considering the best similar application, because it involves reviewing the rating value and user reviews. Several previous studies have proven that difficulties are always encountered when analyzing or extracting important information in user reviews of applications manually. Review analysis is very useful for developing application service features in order to increase user satisfaction and application value ratings, so it requires an automatic user review classification tool by finding the best suitable model. This study applies the CRISP-DM methodology, but only until the evaluation stage. The classification algorithm used is Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), and a combination of tf-idf unigram, bigram, and trigram features. The result of this research is that the combination of the tf-idf unigram (F1) feature with the SVM algorithm achieves the best value for each evaluation value of precision, recall, and f1-score of 87% each. The lowest evaluation value is precision 55% from the combination of F2 features with SVM, 42% recall and 32% f1-score from the combination of F3 features with logistic regression.

**Keywords:** Review Classification, Classifier Algorithm, Machine Learning, Text Mining

## I. PENDAHULUAN

DUNIA teknologi informasi yang semakin berkembang meliputi aplikasi ponsel pintar berbasis sistem operasi seperti android, windows, dan ios. Salah satu manfaat dari perkembangan teknologi informasi adalah data yang dapat dimanfaatkan untuk sumber penggalian informasi bagi suatu organisasi [1].

Aplikasi berbasis android yang diunduh melalui Google Play mendapatkan fasilitas layanan berupa fitur pemberian peringkat jumlah bintang dari satu sampai lima, serta kolom komentar untuk mengulas lebih banyak spesifikasi penilaian dari pengguna kepada pengembang aplikasi yang diunduh [2].

Perpustakaan Nasional Republik Indonesia atau disingkat PNRI adalah suatu lembaga pemerintahan yang menyediakan layanan perpustakaan berstandar nasional, serta berperan menjadi perpustakaan pembina, sumber referensi penelitian, dan pusat jaringan semua perpustakaan di Indonesia [3].

Tahun 2016 PNRI meluncurkan aplikasi perpustakaan digital berbasis *mobile* android di Google Playstore, yaitu

iPusnas. Aplikasi iPusnas adalah layanan perpustakaan digital yang memberikan fasilitas baca *online* atau dalam jaringan(daring) berupa buku elektronik [3]. Jumlah tren pengguna baru aplikasi *mobile* iPusnas menjadi nilai salah satu sub indikator yang berpengaruh terhadap target pencapaian nilai Laporan Kinerja Instansi Pemerintah (LKIP) Pujasintara PNRI 2020 – 2024 [4].

LKIP Pujasintara PNRI 2020 – 2024 adalah laporan hasil kinerja PNRI berfungsi sebagai bahan acuan evaluasi kinerja Pusat Jasa Perpustakaan dan Informasi (Pujasintara) PNRI yang dilakukan secara berkala [5]. Setiap aplikasi *mobile* berbasis android diunduh melalui Google Playstore yang menyediakan layanan berupa fitur jumlah nilai bintang dan kolom ulasan. Kedua fitur tersebut dapat dimanfaatkan sebagai salah satu sarana pemberian umpan balik pengguna yang telah memiliki pengalaman dalam menggunakan sebuah aplikasi android kepada pengembang aplikasi tersebut.

Tahun 2021 jumlah tren pengguna baru aplikasi iPusnas menurun sebesar 3,55% jika dibandingkan dengan tahun 2020 [6]. Penurunan jumlah tren pengguna baru tersebut berpengaruh terhadap kualitas pelayanan Perpustakaan Nasional Republik Indonesia(PNRI) yang diberikan kepada para pemustaka, melalui inovasi pemanfaatan teknologi informasi yang tertulis dalam dokumen LKIP Pujasintara PNRI 2020 – 2024 [4].

Google Play Store memberikan fitur jumlah nilai bintang yang merepresentasikan penilaian pengguna aplikasi, berdasarkan pengalaman menggunakan aplikasi dalam bentuk angka [7], serta fitur kolom ulasan sebagai media penyampaian informasi penilaian pengguna setelah menggunakan aplikasi dalam bentuk teks [8]. Namun, nilai bintang aplikasi *mobile* android iPusnas saat ini hanya mencapai 3.9, sedangkan beberapa aplikasi sejenis lainnya mendapatkan nilai bintang di atas 4.1 [8-9].

*Electronic Word of Mouth* (EWOM) memiliki pengaruh terhadap keputusan dari calon pengguna aplikasi dalam memilih aplikasi yang akan diunduh berdasarkan jumlah peringkat bintang dan isi ulasannya [11], karena calon pengguna baru biasanya mengunduh aplikasi berperingkat paling tinggi diantara aplikasi sejenisnya, serta informasi ulasan terbaik [12]. *Electronic Word of Mouth* (EWOM) adalah suatu wadah penting bagi publik untuk membagikan pendapat tentang pengalaman menggunakan suatu jasa atau aplikasi [13], berupa *like*, peringkat, dan ulasan [14].

Ulasan dan peringkat yang diberikan pengguna aplikasi sebagian mengandung informasi yang tidak relevan atau konsisten [15], [16]. Contohnya, pengguna aplikasi memberi peringkat nilai bintang dua, tetapi isi ulasan aplikasi adalah “aplikasi yang bagus” [17]. Hasil penelitian Pagano dan Maalej menemukan fakta bahwa hanya ada 30% ulasan pengguna aplikasi berisi informasi berkaitan dengan topik pengembangan aplikasi, yaitu ide perbaikan, kekurangan aplikasi, *bug*, permintaan fitur dan konten [18]. Penelitian lain juga membuktikan bahwa hanya ada 43% ulasan mengandung topik tentang sistem aplikasi yang berguna sebagai bahan pengembangan aplikasi selanjutnya [19].

Analisis terhadap suatu ulasan secara manual bukan hal yang mudah, salah satu contohnya adalah hasil penelitian Carreno dan Winbladh, yaitu sebanyak 327 data ulasan yang dianalisis secara manual membutuhkan waktu 13 jam [20]. Oleh karena itu, pengembang aplikasi membutuhkan aplikasi perangkat lunak klasifikasi ulasan pengguna secara otomatis untuk mempelajari isi ulasan secara mendalam.

Penelitian ini menerapkan metode *Natural Language Processing* (NLP) yang merupakan cabang dari teknologi *machine learning* atau *artificial intelligence* [21], berfungsi untuk mempelajari pengelolaan bahasa komunikasi alami antara manusia satu dengan manusia lainnya [21]. Penelitian terdahulu tentang klasifikasi ulasan pengguna salah satu aplikasi *mobile* di Google Playstore terhadap 438 kalimat ulasan. Penelitian tersebut menggunakan algoritma klasifikasi *Decision Tree*, *Naïve Bayes*, *Support Vector Machine*, dan *Logistic Regression*. Keempat algoritma tersebut dikombinasikan dengan fitur *unigram*, *bigram*, panjang teks, nilai peringkat, dan rasio jumlah kata bersentimen positif dan negatif. Prapemrosesan data menggunakan teknik tokenisasi, *negation handling*, pembersihan teks, koreksi ejaan, *stemming*, dan *stopword removal* dengan evaluasi menggunakan *f-measure*, *recall*, dan *precision* dari *k-fold* = 10 [22].

Penelitian kedua merupakan analisis sentimen ulasan aplikasi hotel menggunakan kombinasi algoritma *Naïve Bayes* dan fitur *n-gram*, yaitu *unigram*, *bigram*, *trigram*, dan nilai peringkat. 100 data ulasan berlabel positif dan 100 data ulasan berlabel negatif telah melalui tahap prapemrosesan data *cleaning* dan *case folding*. Evaluasi dengan *Accuracy*, *Error Rate*, *Recall*, dan *Precision* [23]. Penelitian ketiga adalah ekstraksi informasi 1.412 ulasan pengguna aplikasi *mobile* dengan algoritma *Logistic Regression*, *Naïve Bayes*, LDA, *Support Vector Machine*, dan NMF, serta kombinasi fitur nilai peringkat, perangkat, panjang teks, *unigram*, dan *bigram*. Dataset ulasan terbagi menjadi kategori *problem*, *other*, *improvement*, dan *request*. Tahap evaluasinya adalah *recall*, *f1-score*, *precision*, dan *accuracy*.

Penelitian ini bertujuan untuk mendapatkan hasil pemodelan tertinggi klasifikasi ulasan pengguna aplikasi *mobile* iPusnas di Google Playstore. Algoritma klasifikasi yang digunakan adalah *Naïve Bayes* (NB), *Support*

*Vector Machine* (SVM), dan *Logistic Regression* (LR) berdasarkan referensi penelitian [22][24], serta penggunaan fitur *unigram*, *bigram*, dan *trigram* seperti yang dilakukan oleh [23]. Klasifikasi ulasan pengguna aplikasi *mobile* iPusnas mengelompokkan ulasan ke dalam tiga topik, yaitu *problem*, *request*, dan *other* yang dikembangkan dari referensi penelitian [25], karena analisis klasifikasi ulasan aplikasi iPusnas belum pernah dilakukan oleh penelitian lain sebelumnya.

## II. METODE PENELITIAN

### A. Klasifikasi Ulasan Aplikasi

Ulasan aplikasi adalah sebuah tempat penyampaian opini pengguna aplikasi berbentuk dokumen teks berdasarkan pengalamannya selama menggunakan suatu aplikasi yang direpresentasikan dengan peringkat nilai dari satu (nilai terendah) sampai lima (nilai tertinggi) [26]. Ulasan biasanya mengandung beragam informasi yang sangat penting bagi developer suatu aplikasi untuk pengembangan fitur dan layanan aplikasi kepada pengguna pada versi berikutnya, tetapi proses penggalian informasi tersebut membutuhkan analisis yang mendalam dan waktu yang lama jika dilakukan secara manual, sehingga perlu menggunakan alat bantu klasifikasi teks otomatis [22].

Klasifikasi teks merupakan suatu metode pemisahan dan pengelompokkan teks berdasarkan kategori-kategori yang telah didefinisikan sebelumnya [27]. Klasifikasi teks telah diterapkan dan dikembangkan pada jenis penelitian data mining atau text mining, basis data, machine learning, dan aplikasi yang meliputi diagnosis medis, gambar, organisasi dokumen, dan lainnya [28]. Ada dua teknik klasifikasi dokumen atau teks, yaitu pendekatan statistik dan pemelajaran mesin atau machine learning [28].

Pendekatan statistik dilakukan dengan cara menghasilkan pemenuhan hipotesis-hipotesis secara manual menggunakan algoritma klasifikasi, sedangkan pemelajaran mesin (machine learning) melakukan klasifikasi dokumen teks dengan penerapan sistem algoritma klasifikasi dalam teknologi sistem klasifikasi otomatis [28]. Beberapa algoritma machine learning meliputi algoritma *Logistic Regression* (LR), *Naïve Bayes* (NB), dan *Support Vector Machine* (SVM) [28]. Penelitian ini menggunakan teknik machine learning.

### B. Text Mining

*Text mining* atau penambangan teks merupakan suatu metode dan alat berguna sebagai media analisis dokumen atau data yang mengumpulkan informasi secara intensif, baik berupa jenis dokumen data semi terstruktur (teks kalimat dan dokumen teks dalam berkas sebuah berkas) maupun tidak terstruktur, dan termasuk dalam bagian penambangan data atau *data mining* [29]. Penambangan data bertujuan untuk mendapatkan dan memahami informasi penting dari sumber dokumen data dengan melakukan eksplorasi dan identifikasi terhadap pola-pola unik linguistik [29]. *Text mining* menggunakan sumber data berupa koleksi atau koleksi yang tidak terstruktur, sehingga memerlukan proses kategorisasi penemuan pola informasi yang sejenis [29]. *Text mining* diterapkan dalam analisis sentiment, sintesis dokumen, *spamming*, dan lainnya [29].

### C. Algoritma Naïve Bayes

Naïve Bayes (NB) adalah algoritma klasifikasi yang dapat melakukan penentuan nilai kemungkinan untuk prediksi peluang berdasarkan data dari pemelajaran pola sebelumnya dan dapat mengelompokkan suatu dokumen data pada sebuah sistem [21]. Metode NB dipilih untuk penelitian ini karena memiliki nilai akurasi tinggi, sederhana, dan perhitungan yang cepat [30]. Rumus algoritma NB dapat dilihat pada persamaan (1) berikut ini [31]:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Keterangan:

A : Hipotesis dokumen data yang merupakan kelas spesifik.

B : Dokumen data kelas yang belum diketahui.

P(A|B) : Peluang kemunculan hipotesis dalam suatu kondisi.

P(A) : Peluang kemunculan hipotesis.

P(B|A) : Peluang kemunculan kondisi dalam hipotesis.

P(B) : Peluang kemunculan B.

Persamaan (1) menampilkan rumus dari algoritma *Naïve Bayes* (NB) yang mana algoritma tersebut dapat berguna sebagai metode pencarian nilai peluang kemunculan tertinggi dalam proses klasifikasi data uji pada kategori yang paling sesuai dengan cepat dan akurat, terutama jika diimplementasikan pada dataset dalam jumlah

yang besar [31].

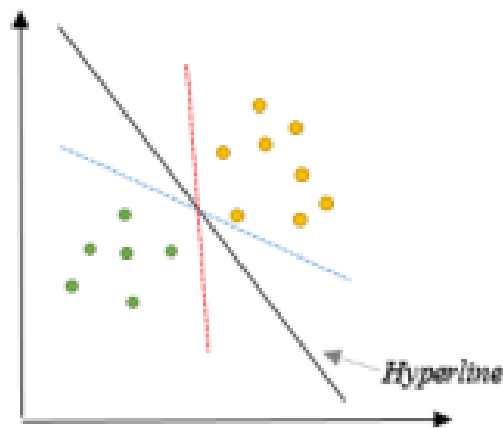
#### D. Algoritma Support Vector Machine

Algoritma *Support Vector Machine* (SVM) merupakan algoritma pembagian data yang telah diketahui sebelumnya berdasarkan klasifikasinya untuk menguji tingkat akurasi data dalam sistem, serta membagi data bersifat non linier dan memisahkan titik vektor menggunakan hyperline [21]. Algoritma SVM juga menghasilkan nilai akurasi terbaik pada penelitian terdahulu [21]. Oleh karena itu, penelitian ini juga menerapkan algoritma SVM.

Pada Gambar 1 dan 2 menunjukkan cara kerja algoritma SVM, dimana pada Gambar 1 ada satu garis hyperline dan dua garis yang membagi dua kelompok titik vektor berwarna hijau dan kuning, kemudian SVM membagi dan mencari kedua kelompok titik vektor tersebut yang paling mendekati kelompoknya dan terdekat dari garis hyperline berwarna hitam untuk mendapatkan hasil tingkat akurasinya.

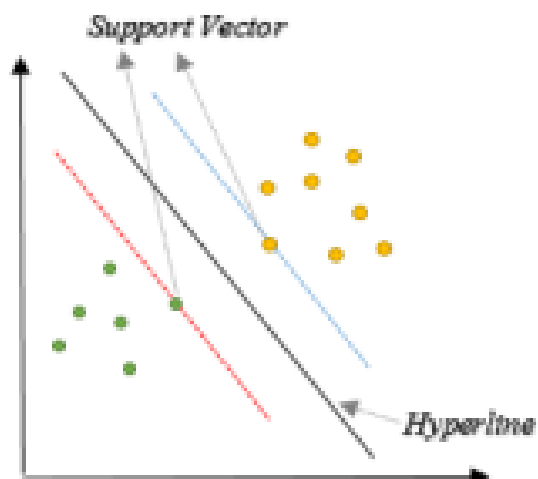
Gambar 2 menampilkan ada dua titik vektor masing-masing berjumlah satu dari kelompok hijau dan kuning yang paling dekat dengan garis hyperline dengan posisi searah lurus disebut sebagai *Support Vector* [21].

#### E. Algoritma Logistic Regression



Gambar. 1. Diagram Pertama Cara Kerja *Support Vector Machine* (SVM)

*Logistic Regression* (LR) adalah modifikasi regresi linier agar dapat melakukan proses klasifikasi ke dalam beberapa kategori berbeda terhadap suatu data. Rumus algoritma LR dapat dilihat pada persamaan (2) berikut ini [32]:



Gambar. 2. Diagram Kedua Cara Kerja *Support Vector Machine* (SVM)

$$p(C = y_i | \bar{X}_i) = \frac{\exp(\bar{A} \cdot \bar{X}_i + b)}{1 + \exp(\bar{A} \cdot \bar{X}_i + b)} \quad (2)$$

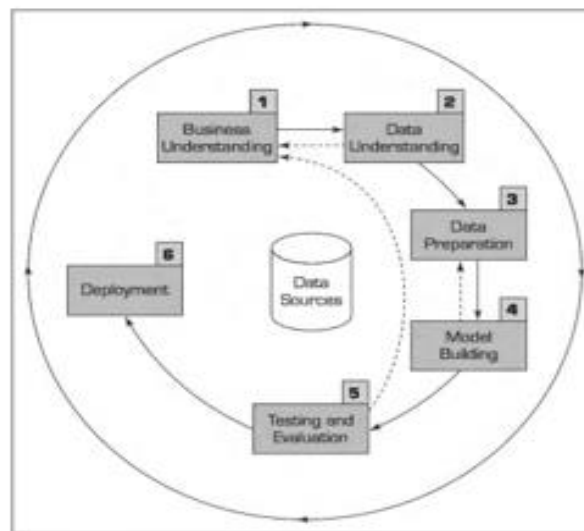
$\bar{X}_i$  adalah variabel yang didapatkan dari proses ekstraksi data atau dokumen yang diklasifikasikan, sedangkan  $y_i$  adalah hasil dari proses klasifikasi dokumen atau data sebelumnya. Selanjutnya, algoritma LR melakukan perhitungan terhadap nilai  $\bar{A}$  dan  $b$  yang berdasarkan pada data *training* atau data pelatihan [32].

#### F. Tahapan Penelitian

Alur atau tahapan penelitian ini mengadopsi pendekatan CRISP-DM melalui enam tahapan. Pertama adalah pemahaman bisnis, kedua adalah pemahaman data, ketiga adalah persiapan data, keempat adalah pemodelan data, kelima adalah evaluasi, dan terakhir adalah tahap deployment. Keenam tahapan tersebut dapat dilihat pada Gambar 3 berikut ini [33].

##### 1) Pemahaman Kebutuhan Bisnis

Tahap ini dilakukan wawancara dengan seorang staff pranata ahli komputer dan beberapa stakeholders Perpustakaan Nasional Republik Indonesia (PNRI) secara langsung dan melalui pengisian formulir secara online



Gambar. 3. Diagram CRISP-DM

atau daring, tujuannya untuk menggali kebutuhan bisnis organisasi. Pengisian formulir online dilakukan untuk mendapatkan kategori-kategori yang dibutuhkan developer aplikasi iPusnas berdasarkan referensi penelitian sebelumnya [25].

##### 2) Pemahaman Data

Tahap ini melakukan pengambilan dataset secara parsing html dengan alat bantu kodifikasi secara otomatis, yaitu google play scrapper yang disediakan python. Selanjutnya, menganalisis struktur data hasil pengambilan dataset tersebut, sehingga didapatkan bagian kolom review untuk digunakan dalam proses klasifikasi ulasan pengguna aplikasi.

##### 3) Persiapan Data

Ada 2.430 dataset ulasan pengguna dari total keseluruhan ulasan sebanyak 26.091 data yang didapatkan merupakan data ulasan teks pengguna aplikasi iPusnas di Google Playstore selama satu tahun sejak bulan Januari 2021 sampai Desember 2021. Dataset tersebut hanya dipilih sebagian dari jumlah seluruhnya secara manual, karena tidak dapat menganalisis data dalam jumlah yang besar secara manual. Penelitian ini menerapkan metode olah data ulasan teks berupa satuan kalimat seperti yang dilakukan oleh penelitian [34]. Ulasan yang dipilih harus memenuhi kriteria antara lain ulasan tidak kosong atau hanya berupa emoticon, menggunakan bahasa indonesia, dan tidak bermakna ambigu.

Anotasi dataset dilakukan secara manual oleh dua orang annotator dan masing-masing annotator tidak diperbolehkan memberi label lebih dari satu jenis kategori pada setiap teks ulasan. Ulasan yang memiliki dua label



berbeda dari kedua annotator harus disepakati satu label terpilih. Jika masing-masing jenis kategori memiliki jumlah data ulasan yang berbeda, maka dilakukan teknik down-sampling berdasarkan saran dari referensi penelitian terdahulu [35].

Prapemrosesan data dilakukan dengan empat proses yang berbeda seperti *cleaning*, *case folding*, *stopword*, dan *stemming*. Proses *Cleaning* dapat melakukan penghapusan tanda baca, emoticon, simbol, url, dan angka sehingga didapatkan hasil data teks yang siap diolah pada proses berikutnya [29]. *Case folding* bertujuan mengubah dokumen teks berupa huruf besar ke dalam dokumen teks huruf kecil [29]. *Stopwords* melakukan penghilangan kata yang tidak diperlukan pada suatu dokumen teks [29]. *Stemming* merupakan proses pengubahan suatu kata menjadi bentuk dasar dari kata tersebut [29]. Pada penelitian ini digunakan library sastrawi untuk proses stemming yang disediakan oleh scikitlearn [36].

#### 4) Pemodelan Data

Proses pemodelan data melakukan ekstraksi fitur yang memanfaatkan beberapa jenis fitur dari tf-idf *n-gram*. Tf-idf menganalisis frekuensi kata untuk menerapkan metode pembobotan kata dengan perhitungan yang menghasilkan frekuensi kata paling penting dan kata yang muncul dalam setiap dokumen. Rumus TF-IDF ditunjukkan pada persamaan (3) berikut ini [29]:

$$W_{dt} = TF_{dt} \times \log \left( \frac{d}{df} \right) \quad (3)$$

Keterangan:

*TF* = frekuensi kata dalam dokumen.

*IDF* = frekuensi kata dalam kumpulan dokumen.

*d* = jumlah total dokumen.

*df* = banyaknya jumlah dokumen yang mengandung *term* atau kata.

*N-gram* merupakan suatu teknik pemisahan kata berdasarkan kategori atau jenis token yang diterapkan. Beberapa jenis fitur *n-gram* tersebut adalah *unigram*, *bigram*, dan *trigram*. *Unigram* mengandung rangkaian kata  $n = 1$  dalam teks ulasan, *bigram* mengandung rangkaian kata  $n = 2$  dalam teks ulasan, dan *trigram* mengandung rangkaian kata  $n = 3$  dalam teks ulasan [23]. Tahap klasifikasi menggunakan kombinasi ketiga fitur *n-gram* dengan algoritma klasifikasi *naïve bayes*, *support vector machine*, dan *logistic regression*.

#### 5) Evaluasi

*Confusion matrix* bertujuan untuk mengevaluasi metode klasifikasi algoritma *support vector machine*, *logistic regression*, dan *naïve bayes*. *Confusion matrix* digunakan untuk mengukur tingkat keakuratan sebuah metode melalui proses evaluasi [29]. Beberapa istilah dalam proses evaluasi yang digunakan secara umum antara lain *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), dimana TP jumlah data prediksi benar positif pada kenyataannya, FP jumlah data prediksi positif bernilai salah, TN jumlah kebenaran nilai negatif pada kelas sesungguhnya, serta FN merupakan jumlah data dengan prediksi negatif yang tidak sesuai hasil kenyataannya [37]. Parameter pengukuran evaluasi metode yang digunakan adalah sebagai berikut:

- A. *Precision* merupakan representasi hasil nilai dari jumlah dokumen TP dengan FP. Pada persamaan (4) menunjukkan rumus *precision*.

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

- B. *Recall* adalah representasi hasil nilai dari jumlah dokumen TP dengan FN. Pada persamaan (5) menunjukkan rumus *recall*.

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

- C. *F-Measure* merupakan hasil pembagian nilai jumlah *recall* dan *precision*. Pada persamaan (6) menunjukkan rumus *precision*.

$$F-Measure = \frac{(2 \times recall \times precision)}{(recall + precision)} \quad (4)$$

#### 6) Evaluasi

Tahap deployment pada penelitian ini berupa penulisan ilmiah atau penelitian.

### III. HASIL DAN PEMBAHASAN

#### A. Dataset Ulasan

Dataset sebanyak 2.430 ulasan teks dilakukan proses pemisahan kalimat teks sehingga menjadi 2.817 data ulasan dan dianotasi secara manual oleh dua orang *annotator* dengan hanya satu label. Hasilnya didapatkan 1.609 dataset terpilih yang memenuhi kriteria-kriteria ulasan dalam penelitian ini. Kategori ulasan mengadopsi dasar-dasar kategori pada penelitian sebelumnya [25], kemudian dikelompokkan berdasarkan kata kunci sejenis, ketiga kategori tersebut adalah kategori problem yang mengandung kata kunci seperti *bug report*, *feature shortcoming*, dan *complaint*. Kategori kedua adalah request berisi kata kunci yang berkaitan dengan *user request* dan *usage scenario*. Kategori ketiga adalah *other* berkaitan dengan *feature strength* dan *praise*.

Salah satu contoh masing-masing ulasan dari ketiga kategori tersebut dapat dilihat pada tabel I berikut ini.

TABEL I  
CONTOH ULASAN JENIS-JENIS KATEGORI

Kategori	Contoh Ulasan
<i>Problem</i>	aplikasinya masih banyak bug, variasi buku masih minim.
<i>Request</i>	tambahkan daftar buku favorit juga.
<i>Other</i>	bagus aplikasinya.

Didapatkan hasil jumlah data ulasan setiap kategori antara lain *problem* 808 data, *request* 282 data, dan *other* 519 data. Jumlah dataset yang tidak seimbang dari ketiga kategori tersebut harus dilakukan *down-sampling* seperti yang dianjurkan oleh penelitian [35] agar nilai hasil klasifikasi ulasan lebih optimal, sehingga didapatkan jumlah total dataset sebanyak 846 data ulasan yang mana masing-masing kategori memiliki 282 data ulasan. Pada tahap berikutnya, data ulasan diolah menggunakan teknik-teknik prapemrosesan data, yaitu *cleaning*, *case folding*, *stopword*, dan *stemming*. Keempat teknik tersebut dapat dilihat pada tabel II berikut ini.

TABEL II  
PRAPEMROSESAN DATA

Jenis	Ulasan	Hasil
<i>Cleaning</i>	Tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	Tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya
<i>Case Folding</i>	Tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya
<i>Stopword</i>	Tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	['kadang', 'error', 'bikin', 'kesal', 'tolong', 'diperbaiki']
<i>Stemming</i>	Tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	['kadang', 'error', 'bikin', 'kesal', 'tolong', 'baik']

Dataset yang telah melalui tahap prapemrosesan data selanjutnya diolah dalam proses ekstraksi fitur *tf-idf n-gram* antara lain *unigram*, *bigram*, dan *trigram*. Tabel III menampilkan contoh ekstraksi fitur-fitur *tf-idf n-gram* terhadap ulasan.

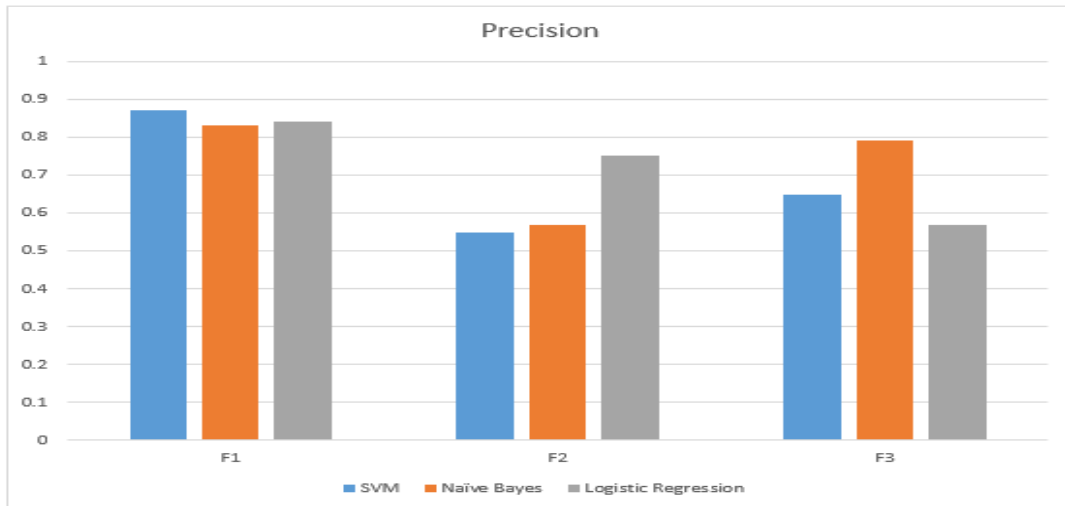
TABEL III  
HASIL N-GRAM

N-Gram	Ulasan	Hasil
<i>Unigram</i>	tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	['tapi', 'kadang', 'error', 'sampai', 'bikin', 'kesal', 'jadi', 'tolong', 'diperbaiki', 'ya']
<i>Bigram</i>	tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	['tapi kadang', 'kadang error', 'error sampai', 'sampai bikin', 'bikin kesal', 'kesal jadi', 'jadi tolong', 'tolong diperbaiki', 'diperbaiki ya']
<i>Trigram</i>	tapi kadang error sampai bikin kesal jadi tolong diperbaiki ya	['tapi kadang error', 'kadang error sampai', 'error sampai bikin', 'sampai bikin kesal', 'bikin kesal jadi', 'kesal jadi tolong', 'jadi tolong diperbaiki', 'tolong diperbaiki ya']

### B. Hasil Penelitian

Dataset terpilih selanjutnya diolah ke dalam tahap klasifikasi ulasan pengguna dengan tiga algoritma klasifikasi yang digunakan pada penelitian ini, serta dikombinasikan dengan tiga fitur n-gram, yaitu *unigram* (selanjutnya disebut F1), *bigram* (selanjutnya disebut F2), dan *trigram* (selanjutnya disebut F3). Evaluasi pada penelitian ini juga menggunakan 10k-fold cross validation dan perbandingan data latih dan data uji adalah 80:20.

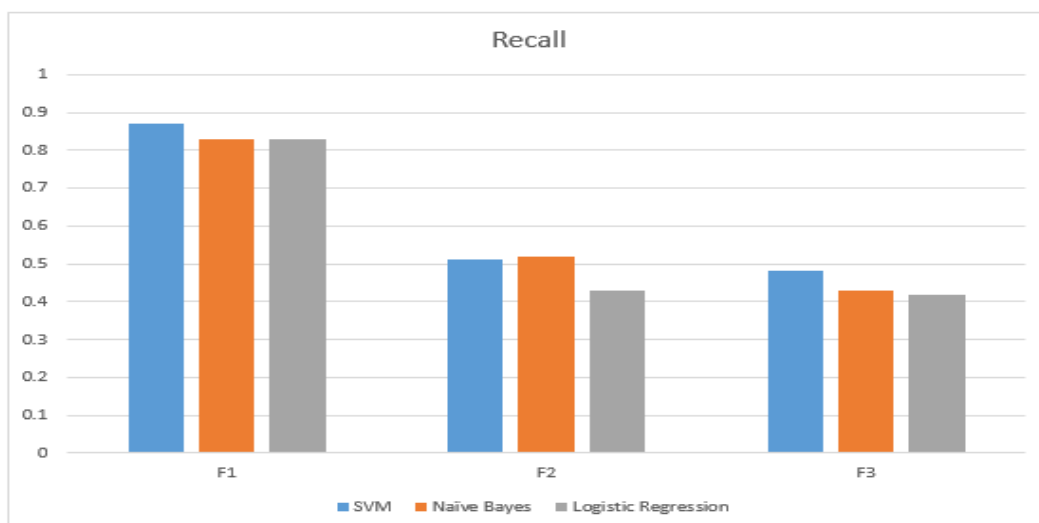
Pada Gambar 4 menampilkan grafik hasil evaluasi klasifikasi *precision*, dimana terlihat jelas bahwa fitur F1 memberikan nilai tertinggi dibandingkan F2 dan F3. Kombinasi fitur F1 dan algoritma SVM menghasilkan grafik tertinggi untuk nilai *precision*.



Gambar. 4. Grafik *Precision*

Pada Gambar 4, kombinasi algoritma SVM dan fitur F1 mencapai nilai terbaik sebesar 87%, sedangkan untuk naïve bayes 83% dan logistic regression 84%. Untuk fitur F2 didapatkan hasil terbaik adalah logistic regression 75%, kemudian diikuti naïve bayes 57% dan SVM 55%. Fitur F3 mendapatkan nilai terbaik dari naïve bayes 79%, kedua SVM 65%, dan terendah logistic regression 57%. Pada evaluasi *precision*, fitur F1, F2, dan F3 mengalami perubahan nilai evaluasi yang tidak konsisten.

Gambar 5 juga menunjukkan bahwa kombinasi fitur F1 dan algoritma SVM menghasilkan grafik tertinggi untuk nilai *recall* dibandingkan fitur F2 dan F3.



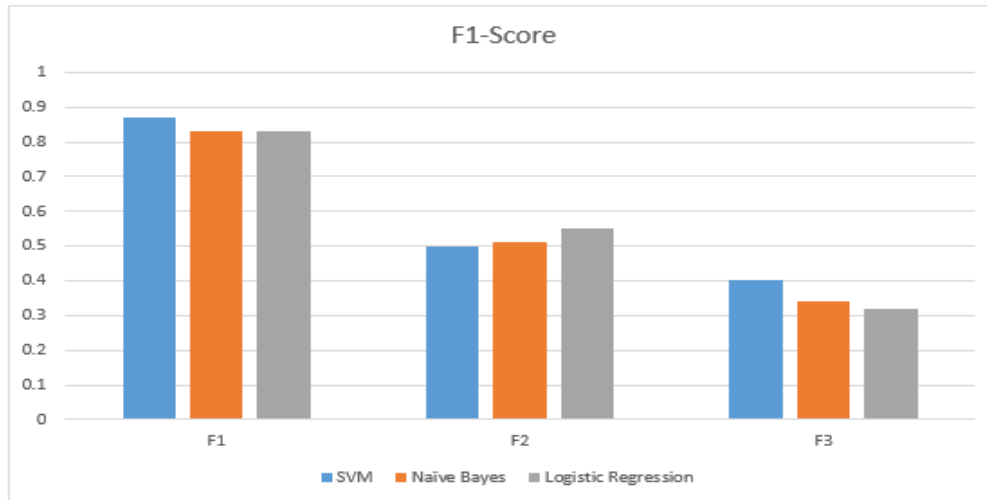
Gambar. 5. Grafik *Recall*

Pada Gambar 5, kombinasi algoritma SVM dan fitur F1 juga mendapatkan nilai terbaik sebesar 87% dan hasil seimbang diperoleh algoritma naïve bayes 83% dan logistic regression 83%. Untuk fitur F2 didapatkan hasil terbaik dari naïve bayes 52%, diikuti oleh SVM 51% dan logistic regression 43%. Fitur F3 memperoleh nilai terbaik dari



SVM 48%, kedua naïve bayes 43%, dan terendah logistic regression 42%. Pada evaluasi recall, fitur F1, F2, dan F3 mengalami penurunan nilai evaluasi secara konsisten dari ketiga algoritma.

Gambar 6 juga menggambarkan kombinasi fitur F1 dan algoritma SVM juga menghasilkan nilai grafik tertinggi untuk pada *f1-score* dibandingkan dengan fitur F2 dan F3.

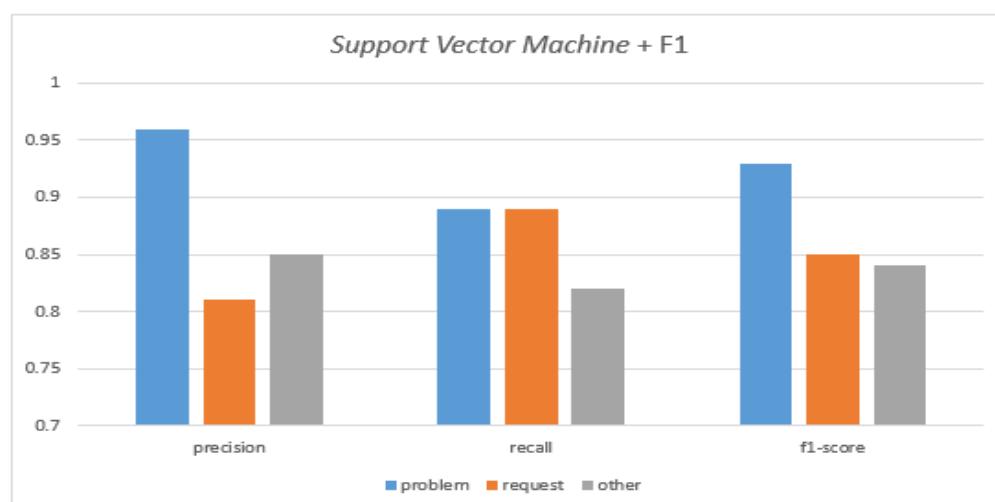


Gambar. 6. Grafik *F1-Score*

Pada Gambar 6, kombinasi algoritma SVM dan fitur F1 juga mendapatkan nilai terbaik sebesar 87% dan hasil seimbang juga diperoleh algoritma naïve bayes 83% dan logistic regression 83%. Untuk fitur F2 didapatkan hasil terbaik dari logistic regression 52%, selanjutnya algoritma naïve bayes 51% dan SVM 50%. Perolehan nilai terbaik fitur F3 berasal dari SVM 40%, kedua *naïve bayes* 34%, dan terendah *logistic regression* 32%. Pada evaluasi *f1-score*, fitur F1, F2, dan F3 juga mengalami penurunan nilai evaluasi secara konsisten untuk setiap algoritma.

Berdasarkan hasil penelitian didapatkan setelah melalui tahap evaluasi, diketahui bahwa model klasifikasi ulasan terbaik diperoleh dari kombinasi fitur F1 dengan algoritma SVM. Rincian hasil evaluasi nilai *precision*, *recall*, *f1-score* untuk masing-masing kategori pada kombinasi SVM dan fitur F1 dapat dilihat pada grafik gambar 7 berikut ini.

Gambar 7 menunjukkan bahwa kategori problem lebih dominan mencapai nilai grafik tertinggi dibandingkan



Gambar. 7. Grafik Kategori Klasifikasi

kategori lainnya. Pada precision kategori *problem* mencapai nilai 96%, sedangkan *request* 81% dan *other* 85%. Sementara itu, nilai recall terbaik dihasilkan oleh kategori *problem* dan *request*, yaitu 89%. Nilai *f1-score* terbaik juga dicapai oleh kategori *problem* 93%, kedua adalah *request* 85%, dan terendah dicapai kategori *other* 84%.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa model klasifikasi ulasan pengguna aplikasi terbaik dihasilkan oleh algoritma *Support Vector Machine* (SVM) yang dikombinasikan dengan fitur *tf-idf unigram* (F1), dimana nilai evaluasi *precision*, *recall*, dan *f1-score* bernilai sama sebesar 87%. Kemudian untuk perolehan nilai evaluasi tertinggi dari tiga kategori ulasan, kategori *problem* merupakan kategori yang menghasilkan lebih banyak nilai evaluasi terbaik dibandingkan kategori *request* dan *other*, yaitu *precision* 96%, *recall* 89%, dan *f1-score* 93%. Hasil evaluasi terendah *precision* 55% dihasilkan dari kombinasi fitur F2 dengan algoritma SVM, sedangkan *recall* 42% dan *f1-score* 32% dari hasil kombinasi fitur F3 dengan *logistic regression*.

#### DAFTAR PUSTAKA

- [1] S. Wahyu Handani, D. Intan Surya Saputra, Hasirun, R. Mega Arino, and G. Fiza Asyrofi Ramadhan, "Sentiment analysis for go-jek on google play store," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012032.
- [2] Google, "Rating konten Aplikasi & Game di Google Play - Bantuan Google Play." [Online]. Available: <https://support.google.com/googleplay/answer/6209544?hl=id>. [Accessed: 01-Nov-2020].
- [3] Perpustakaan Nasional, "Peraturan Perpustakaan Nasional Republik Indonesia Nomor 7 Tahun 2020 Tentang Rencana Strategis Perpustakaan Nasional Tahun 2020-2024," no. Agustus 2020, 2020.
- [4] L. Kinerja, I. Pemerintah, P. Jasa, I. Perpustakaan, D. A. N. Pengelolaan, and N. Nusantara, "Perpustakaan Nasional."
- [5] "Renstra Pujasintara 2020-2024 21-Jan-3."
- [6] "IPusnas Dashboard." [Online]. Available: <https://dashboard-ipusnas.perpusnas.go.id/>. [Accessed: 15-Jul-2022].
- [7] I. K. Mensah, "Perceived Usefulness and Ease of Use of Mobile Government Services," *Int. J. Technol. Diffus.*, vol. 11, no. 1, pp. 1–16, 2019, doi: 10.4018/ijt.d.2020010101.
- [8] H. Yang and P. Liang, "Identification and classification of requirements from app user reviews," *Proc. Int. Conf. Softw. Eng. Knowl. Eng. SEKE*, vol. 2015-Janua, pp. 7–12, 2015, doi: 10.18293/SEKE2015-063.
- [9] "Google Play Buku - Aplikasi di Google Play." [Online]. Available: <https://play.google.com/store/apps/details?id=com.google.android.apps.books>. [Accessed: 15-Jul-2022].
- [10] "Amazon Kindle - Aplikasi di Google Play." [Online]. Available: <https://play.google.com/store/apps/details?id=com.amazon.kindle>. [Accessed: 15-Jul-2022].
- [11] I. Almarashdeh and M. K. Alsmadi, "How to make them use it? Citizens acceptance of M-government," *Appl. Comput. Informatics*, vol. 13, no. 2, pp. 194–199, 2017, doi: 10.1016/j.aci.2017.04.001.
- [12] A. Mahmood, "Identifying the influence of various factor of apps on google play apps ratings," 2019.
- [13] A. Khairunnisa and E. Prasoj, "the Effect of Technology Acceptance Model, Electronic Word of Mouth (Ewom) and Service Quality in Application Contact ...," *PROCEEDING ICTESS (Internasional ...)*, vol. 2020, pp. 91–102, 2020.
- [14] N. Donthu, S. Kumar, N. Pandey, N. Pandey, and A. Mishra, "Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis," *J. Bus. Res.*, vol. 135, no. July, pp. 758–773, 2021, doi: 10.1016/j.jbusres.2021.07.015.
- [15] M. R. Islam, "Numeric Rating of Apps on Google Play Store by Sentiment Analysis on User Reviews," pp. 1–4, 2014.
- [16] K. Sharma and K. Lin, "Review Spam Detector with Rating Consistency Check," 2013.
- [17] D. Lo, M. R. Lyu, and I. King, "Automating App Review Response Generation," *2019 34th IEEE/ACM Int. Conf. Autom. Softw. Eng.*, pp. 163–175, 2019, doi: 10.1109/ASE.2019.00025.
- [18] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," *2013 21st IEEE Int. Requir. Eng. Conf. RE 2013 - Proc.*, pp. 125–134, 2013, doi: 10.1109/RE.2013.6636712.
- [19] C. Tao, H. Guo, and Z. Huang, "Identifying security issues for mobile applications based on user review summarization," *Inf. Softw. Technol.*, vol. 122, no. October 2019, p. 106290, 2020, doi: 10.1016/j.infsof.2020.106290.
- [20] L. V. G. Carreno and K. Winbladh, "Analysis of user comments: An approach for software requirements evolution," *Proc. - Int. Conf. Softw. Eng.*, pp. 582–591, 2013, doi: 10.1109/ICSE.2013.6606604.
- [21] D. Tuhenay, "Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM)," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 105–111, 2021, doi: 10.33387/jiko.v4i2.2958.
- [22] Y. Ekanata and I. Budi, "Mobile application review classification for the Indonesian language using machine learning approach," *2018 4th Int. Conf. Comput. Technol. Appl. ICCTA 2018*, pp. 117–121, 2018, doi: 10.1109/CATA.2018.8398667.
- [23] T. Widiyaningtyas, I. A. Elbaiti Zaeni, and R. Al Farisi, "Sentiment Analysis Of Hotel Review Using N-Gram And Naive Bayes Methods," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, pp. 4–8, 2019, doi: 10.1109/ICIC47613.2019.8985946.
- [24] E. Suprayogi, I. Budi, and R. Mahendra, "Information extraction for mobile application user review," *2018 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2018*, pp. 343–348, 2019, doi: 10.1109/ICACSIS.2018.8618164.
- [25] E. Guzman, M. El-Haliby, and B. Bruegge, "Ensemble methods for app review classification: An approach for software evolution," *Proc. - 2015 30th IEEE/ACM Int. Conf. Autom. Softw. Eng. ASE 2015*, pp. 771–776, 2015, doi: 10.1109/ASE.2015.88.
- [26] "Fision - Aplikasi di Google Play." [Online]. Available: <https://play.google.com/store/apps/details?id=mam.reader.ipusnas>. [Accessed: 19-Oct-2021].
- [27] J. Xu and Y. Cai, "Incorporating context-relevant knowledge into convolutional neural networks for short text classification\*," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 10067–10068, 2019, doi: 10.1609/aaai.v33i01.330110067.
- [28] M. Allahyari et al., "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," 2017.
- [29] F. R. Irawan et al., "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors Sentiment Analysis of Gojek Users Using K-Nearest Neighbor," vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.
- [30] Dimsyiar M Al Hafiz, Khoiril Amaly, Javen Jonathan, M Terangono Rachmatullah, and Rosidi, "Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes," *J. Rekayasa Elektro Sriwij.*, vol. 2, no. 2, pp. 151–157, 2021, doi: 10.36706/jres.v2i2.29.
- [31] M. M. Mala Olhang, S. Achmadi, and F. . A. Wibisono, "Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (Nbc)," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 4, no. 2, pp. 214–221, 2020, doi: 10.36040/jati.v4i2.2695.
- [32] C. C. Aggarwal and C. X. Zhai, *Mining text data*, vol. 9781461432. 2013.
- [33] S. A. Lee, *Business Intelligence and Analytics Ramesh Sharda Dursun Delen Efraim Turban Tenth Edition*. 2014.
- [34] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can i improve my app? Classifying user reviews for software maintenance and evolution," *2015 IEEE 31st Int. Conf. Softw. Maint. Evol. ICSME 2015 - Proc.*, no. September, pp. 281–290, 2015, doi: 10.1109/ICSM.2015.7332474.



- [35] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 4, pp. 42–47, 2012.
- [36] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *M.Sc. Thesis, Append. D*, vol. pp, pp. 39–46, 2003.
- [37] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Comput. Surv.*, vol. 49, no. 2, 2016, doi: 10.1145/2938640.