Klasifikasi Ulasan Pengguna Aplikasi Wondr Berdasarkan *ISO* 25010 Menggunakan *Multinomial Naive Bayes*

1st Achmad Chairul Irfansyah
Program Studi Informatika
Universitas Telkom, Kampus Surabaya
Surabaya. Indonesia
chairul@student.telkomuniversity.ac.id

2nd Alqis Rausanfita, S.Kom., M.Kom.
Program Studi Informatika
Universitas Telkom, Kampus Surabaya
Surabaya. Indonesia
alqisfita@telkomuniversity.ac.id

3rd Dr. Dyah Putri Rahmawati, S.Stat.
Program Studi Informatika
Universitas Telkom, Kampus Surabaya
Surabaya. Indonesia
dyahputri@telkomuniversity.ac.id

***Abstrak* — Penelitian ini bertujuan mengklasifikasikan ulasan pengguna aplikasi *Wondr by BNI* berdasarkan kategori ISO 25010 Quality in Use menggunakan algoritma Multinomial Naive Bayes. Data diperoleh melalui *scraping* ulasan Google Play Store periode Juli 2024 hingga April 2025. Ulasan diproses melalui tahap *preprocessing*, pelabelan manual ke dalam enam kategori (Satisfaction, Effectiveness, Efficiency, Freedom from Risk, Context Coverage, dan Uncategorized), serta direpresentasikan menggunakan metode TF-IDF. Evaluasi model dilakukan menggunakan K-Fold Cross Validation (K-Fold 5 sampai 10) dengan penerapan teknik oversampling SMOTE dan ADASYN pada rasio 60%, 80%, dan 100% untuk mengatasi ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa oversampling mampu meningkatkan performa model secara signifikan. Performa terbaik diperoleh pada rasio 60%–80%, dengan kombinasi *ADASYN* 80% menghasilkan nilai tertinggi, yaitu accuracy 0,8516, precision 0,8548, recall 0,8516, dan F1-score 0,8501. Penelitian ini diharapkan dapat memberikan informasi bagi pengguna serta menjadi masukan bagi pengembang dalam meningkatkan kualitas aplikasi *Wondr by BNI*.**

***Kata kunci* — Multinomial Naïve Bayes, Text Classification, ISO 25010 Quality in Use, Oversampling, TF-IDF, Wondr by BNI, Google Play Store.**

I. PENDAHULUAN

Perkembangan teknologi digital yang sangat pesat dalam satu dekade terakhir telah mendorong transformasi besar di sektor perbankan. Salah satu institusi perbankan yang gencar melakukan inovasi dalam mendukung transformasi digital ini adalah Bank Negara Indonesia (BNI). Bentuk perwujudan inovasi dalam mendukung transformasi digital, BNI meluncurkan aplikasi terbarunya bernama Wondr by BNI pada 5 Juli 2024[1].

Ulasan merupakan salah satu bentuk evaluasi yang diberikan oleh individu terhadap suatu produk atau layanan [2]. Ulasan yang diperoleh dari *Google Play Store* umumnya berjumlah besar dan bersifat tidak terstruktur, sehingga diperlukan teknik tertentu untuk memahami persepsi pengguna terhadap aplikasi tersebut [3]. Oleh karena itu, dibutuhkan suatu metode yang dapat mengklasifikasikandataset yang bersumber dari ulasan aplikasi *Wondr by BNI* di *Google Play Store*. Salah satu pendekatan yang relevan untuk digunakan adalah teknik klasifikasi teks, yaitu metode yang bertujuan untuk mengidentifikasi atau mengorganisasi kelas-kelas dalam suatu dokumen [4].

klasifikasi teks yang mampu untuk mengkategorikan data review aplikasi *Wondr by BNI* ke dalam satu atau gabungan dari beberapa kategori yang relevan secara otomatis, yakni dengan melakukan metode klasifikasi *Multi-Class Classification*. Penelitian tentang klasifikasi teks pada *Multi-Class Classification* ini banyak yang sudah melakukan penelitian. Salah satu penelitian yang dilakukan oleh [5] dengan judul" Unbalanced *Multi-Class Classification Classification with Adaptive Synthetic Multinomial Naive Bayes Approach*" merupakan penelitian tentang *Multi-Class Classification* dengan menggunakan metode *Multinomial Naive Bayes*, dengan dataset yang didapat dari hasil crawling twitter untuk mencari data tentang kenaikan BBM. Hasil akurasi yang didapat bahwa *Multinomial Naive Bayes* memiliki akurasi 88,2%. Salah satu penelitian tambahan selanjutnya [6]dengan judul "*Klasifikasi Berita Menggunakan Metode Multinomial Naïve Bayes*" merupakan penelitian tentang klasifikasi *Multi-Class Classification* dengan topik berita, dataset yang digunakan pada penelitian ini diambil dari portal web berita Kompas dan Tempo dengan kategori Teknologi, otomotif, dan Kesehatan. Penelitian ini juga melakuakn pemrosesan kata dengan TF-IDF serta pemrosesan dataset dengan preprocessing. Akurasi yang didapat pada penelitian ini akurasi 96%, precession 96%, *recall* 96% dan *f1-score* 96% dengan 10.500 dataset yang digunakan.

II. KAJIAN TEORI

A. Wondr by BNI

Wondr merupakan sebuah aplikasi *mobile banking* yang dimiliki oleh bank BUMN yakni bank BNI. Bank BNI selaku pemilik wondr mengembangkan sebuah platfrom yang dirancang sebagai ekosistem digital dengan menggabungkan layanan keuangan, hiburan, gaya hidup, dan program loyalitas dalam satu aplikasi yang dimana wondr sendiri

dikembangkan untuk menyasar generasi mudah, khususnya Gen Z dan milenial[7].

B. *Scraping* Data

Scraping merupakan merupakan sebuah teknik pengumpulan data web terstruktur secara otomatis menggunakan sebuah aplikasi ataupun ekstension web serta bisa menggunakan kode program yang biasanya menggunakan Bahasa pemrogaman python. Data yang terkumpul bisa berasal dari ribuan data, jutaan bahkan miliaran data pada dunia maya[8]. Scraping, atau bisa disebut *web scraping*, merupakan teknik pengambilan data secara otomatis dari satu situs website menggunakan program atau sebuah skrip tertentu, scraping bisa digunakan untuk mengambil informasi aplikasi dari Google Play Store menggunakan sebuah alat Bernama *Google-play-sraper*. Dengan teknik ini memungkinkan peneliti mengekstraksi data dalam jumlah besar [9].

C. *Multi-class* Classification

Klasifikasi Multi-class Classification merupakan sebuah metode machine learning untuk mengklasifikasikan sebuah data ke dalam lebih dari satu atau dua kategori yang berbeda. Terdapat beberapa pendekatan untuk mengkasifikasikan Multi-Class Classification, yakni One Against One dan One Against All. Dari kedua pendekatan tersebut dapat membantu dalam proses mengklasifikasikan Multi-Class Classification [10]. Multi-Class Classification bisa digunakan untuk mengklasifikasikan beberapa jenis-jenis serangan siber berdasarkan data yang diperoleh dari sistem deteksi intrusi. Klasifikasi tidak hanya terbatas pada dua kelas (seperti serangan atau bukan serangan), tetapi juga melibatkan beberapa kelas serangan yang berbeda, menjadikan studi Multi-Class Classification yang kompleks [11].

D. *Multinomial Naive Bayes*

Multinomial *Naive Bayes* merupakan bentuk dari Naive Bayes yang sering digunakan untuk klasifikasi teks dan analisis sentimen. Metode probabilitas ini memanfaatkan teorema Bayes dan sangat populer dalam pemrosesan bahasa alami (Natural Language Processing/NLP). Algoritma ini bekerja dengan menggunakan Term Frequency, yaitu jumlah kemunculan sebuah kata di dalam dokumen. Dalam proses pemodelannya, terdapat dua elemen utama yang diukur: adanya kata dalam dokumen (presence) dan seberapa sering kata itu muncul (frequency). [12]. Fase inisial meliputi komputasi *prior probability* tiap label menggunakan formulasi matematis yang mengacu pada persamaan (1) berikut ini [13].

$$P(c) = \frac{N_C}{N} \quad 1$$

Keterangan :
$P(c)$: Prior Probability Kelas C.
N_C : Jumlah Kelas *C* Pada Seluruh Dokumen.
N : jumlah seluruh dokumen.

Tahapan berikutnya melibatkan komputasi probabilitas term kondisional untuk setiap kategori label dengan memperhitungkan nilai pembobotan TF-IDF masing-masing term. Implementasi Multinomial Naive Bayes dalam penelitian ini mengintegrasikan skema pembobotan TF-IDF melalui modifikasi persamaan dasar, seperti contoh persamaan berikutini

$$\mathrm{P}(t_n \mid \mathrm{c}) = \frac{w_{ct} + 1}{(\sum w' \in VW'_{ct}) + B'} \quad 2$$

Keterangan :
w_{ct} : Prior Probability Kelas C.
$\Sigma \mathrm{w}' \in \mathrm{VW}'_{ct}$: Jumlah Total Bobot Dari Keseluruhan Term Yang Berada Di Kelas C.
B' : Jumlah Bobot Dari Term Yang Unik Pada Seluruh Dokumen.

E. ISO 25010 Quality in Use

ISO merupakan sebuah lembaga global yang fokus pada pembuatan standar internasional. Salah satu inovasinya terwujud dalam model ISO 9126, yang ditujukan untuk menilai sejauh mana produk atau perangkat lunak dapat memenuhi kebutuhan penggunanya. Evaluasi ini meliputi faktor-faktor efektivitas, efisiensi, keamanan, serta kepuasan pengguna dalam mencapai target tertentu dengan menggunakan perangkat lunak tersebut. [14]. *Quality In Use* adalah salah satu kelompok dari ISO (*International Organization of Standardization*) dimana merupakan salah satu organisasi standar internasional. Model *Quality in Use* salah satu kelompok yang berfokus pada perspektif penggunaan dengan memperhatian 5 aspkeh yakni, *Effectiveness*, *Efficiency*, *Freedom from Risk*, *Satisfaction*, *Context Coverage* [15], Dan salah satu label yang tidak menjadi bagian Quality in Use yakni Uncategorized.

F. *Pre-processing*

Tahap preprocessing dilakukan setelah pengumpulan data mentah. Ulasan yang terkumpul kemudian melalui serangkaian proses pra-pemrosesan meliputi pembersihan data, penyederhanaan, dan transformasi untuk memudahkan pengolahan serta meningkatkan kualitas data [16].

G. K-Fold Cross Validation

K-Fold Cross Validation merupakan metode yang digunakan untuk mengevaluasi kinerja sebuah model pembelajaran mesin secara objektif. Pada metode ini, data terlebih dahulu akan diacak secara acak (*randomly shuffled*) agar terhindar dari bias urutan data, kemudian dibagi menjadi *k* bagian atau *fold* yang berukuran hamper sama. Proses pelatihan dan pengujian model dilakukan sebanyak *k* kali: pada setiap iterasi, satu *fold* digunakan sebagai data uji (*validation set*), sedangkan k-1-fold lainya digunakan untuk melatih model (*training set*) nilai kinerja model, seperti akurasi atau metrik lain yang sesuai, dihitung di setiap iterasi, lalu dirata-ratakan untuk memperoleh extimasi performa keseluruhan yang lebih stabil[17].

H. Term Frequency – Inverse Document Frequency

Metode TF-IDF adalah teknik yang digunakan untuk mengukur nilai setiap kata berdasarkan frekuensi kemunculan kata itu, terutama dalam konteks pengambilan informasi. Pendekatan ini dikenal karena efektivitasnya, kesederhanaan dalam penerapannya, serta hasil yang sangat tepat. TF-IDF menerapkan metode statistik untuk menilai pentingnya suatu kata[18]. Term Frequency (TF) adalah suatu ukuran yang menggambarkan seberapa sering suatu kata (term) muncul dalam sebuah dokumen. Penghitungan

frekuensi kata ini mempertimbangkan panjang keseluruhan dokumen, karena setiap dokumen memiliki jumlah kata yang bervariasi. Oleh karena itu, nilai TF diperoleh dengan membagi jumlah kemunculan kata tertentu dengan total jumlah kata dalam dokumen tersebut. ditunjukkan pada Persamaan dibawah

$$TF = \frac{Jumlah\ kemunculan\ kata\ dalam\ dokument}{Jumlah\ kata\ dalam\ dokumen} \quad 3$$

Keterangan :
TF : frekuensi kemunculan kata pada sebuah dokumen.

IDF (*Inverse Document Frequency*) Setelah menghitung nilai TF, langkah berikutnya adalah menghitung nilai IDF yang bertujuan untuk menilai seberapa signifikan sebuah kata. Oleh karena itu, jika nilai IDF yang dihasilkan semakin rendah, maka kata tersebut menjadi semakin kurang berarti. Dapat ditunjukkan pada persamaan dibawah

$$\text{IDF} = \log\frac{N}{n} \quad 4$$

Keterangan :
IDF : Mengukur penting/tidak sebuah kata dalam dokument.
N : Total jumlah dokument dalam *corpus*
n : Merupakan jumlah dokument dalam *corpus*
log : Merupakan *logoritma*

Tahapan berikutnya melibatkan komputasi probabilitas term kondisional untuk setiap kategori label dengan memperhitungkan nilai pembobotan TF-IDF masing-masing term. Implementasi Multinomial Naive Bayes dalam penelitian ini mengintegrasikan skema pembobotan TF-IDF melalui modifikasi persamaan dasar, seperti contoh persamaan berikutini

$$\mathrm{P}(t_n \mid \mathrm{c}) = \frac{w_{ct} + 1}{(\sum w' \in VW'_{ct}) + B'} \quad 5$$

Keterangan :
w_{ct} : Prior Probability Kelas C.
$\Sigma w' \in VW'_{ct}$: Jumlah Total Bobot Dari Keseluruhan Term Yang Berada Di Kelas C.
B' : Jumlah Bobot Dari Term Yang Unik Pada Seluruh Dokumen.

I. Oversampling

Metode yang sering digunakan untuk mengatasi ketidakseimbangan pada sata *multi-Class* adalah *SMOTE* (Synthetic Minority *Oversampling* Technique) dan *ADASYN* (Adaptive Synthetic Sampling). Pada data multiclass, dua metode ini diterapkan pada masing-masing kelas minoritas secara berulang dengan tujuan menambhak jumlah sampel pada sesmua kelas yang memiliki frekuensi rendah. *SMOTE* membantu data sintetis dengan mengambil sastu titik minoritas, lalu mencari beberapa tetangga terdekatnya, kemudian membuat titik baru ditengah-tengah melalui proses interpolasi. Metode ini membuat data minoritas menjadi lebih beragam tanpa melakukan duplikasi

J. *K-Fold Cross Validation*

K-Fold Cross Validation merupakan metode yang digunakan untuk mengevaluasi kinerja sebuah model pembelajaran mesin secara objektif. Pada metode ini, data terlebih dahulu akan diacak secara acak (*randomly shuffled*) agar terhindar dari bias urutan data, kemudian dibagi menjadi *k* bagian atau *fold* yang berukuran hamper sama. Proses pelatihan dan pengujian model dilakukan sebanyak *k* kali: pada setiap iterasi, satu *fold* digunakan sebagai data uji (*validation set*), sedangkan k-1-fold lainya digunakan untuk melatih model (*training set*) nilai kinerja model, seperti akurasi atau metrik lain yang sesuai, dihitung di setiap iterasi, lalu dirata-ratakan untuk memperoleh extimasi performa keseluruhan yang lebih stabil[17].

K. Evaluasi Model

Evaluasi model merupakan proses penilaian efektifitas dari sebuah algoritma dalam mengkategorikan *Multi-Class Classification*, dengan memanfaatkan beberapa kategori *matrix*, *Accuracy*, *Precision*, *Recall*, *F1-score* dan *Confusion matrix*. *Accuracy* menentukan seberapa banyak data yang berhasil diklasifikasikan dengan benar, *Precision* mengukur kemampuan model dalam mengidentifikasi prediksi yang dihasilkan oleh sistem. *Recall* mengukur keberhasilan suatu sistem untuk mengenali kelas yang menjadi target. *F1-Score* memberikan gambaran pada kinerja model.

LABEL 1
(EVALUASI MODEL)

		Actual Label			
		kelas 1	kelas 2	…	kelas M
Predicted Label	Kelas 1	TP	FN	…	FN
	Kelas 2	FP	TN	…	
	Kelas 3	…	…	…	…
	Kelas M	FP	TN	…	TN

Keterangan :
TP (True Positive) : Prediksi positif dan itu benar
TN (True Negatif) : Prediksi negatif dan itu benar
FP (False Positive) : Prediksi positif dan itu salah
FN (False Negative) : Prediksi negatif dan itu salah

Persamaan matrik-matrik dirumuskkan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad 6$$

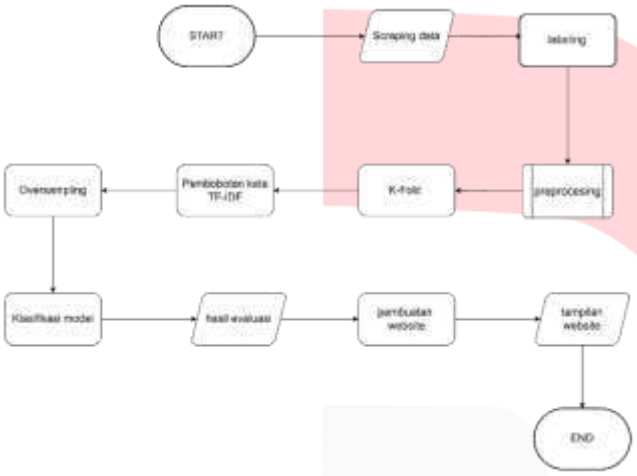
$$Precision = \frac{TP}{TP + FP} \quad 7$$

$$Accuracy = \frac{TP}{TP + FN} \quad 8$$

$$Accuracy = 2X\frac{Precision\ x\ Recall}{Precision\ +\ Recall} \quad 9$$

III. METODE

Gambar rancangan dari sistem yang akan dibangun pada penelitian ini terdapat beberapa proses, muai dari Scraping data dari google play store, lalu labeling secara manual, preprocessing untuk membersihkan dataset, k-fold untuk membagi dataset, selanjutnya pembobotan kata dengan TF-IDF, oversampling untuk pemproses data yang tidak seimbang, klasifikasi model dengan multinomial naive bayes, hasil evaluasi untuk menentukan Accuracy, Precision, Recall, F1-score dari hasil pelatihan model, lalu pembuatan web dengan flask dan html biasa, selanjutnya menampilkan hasil dari pelatihan model ke dalam web. Untuk alur penelitian dapat dilihat pada gambar 1.



GAMBAR 1
(ALUR PENELITIAN)

A. Pengumpulan Data

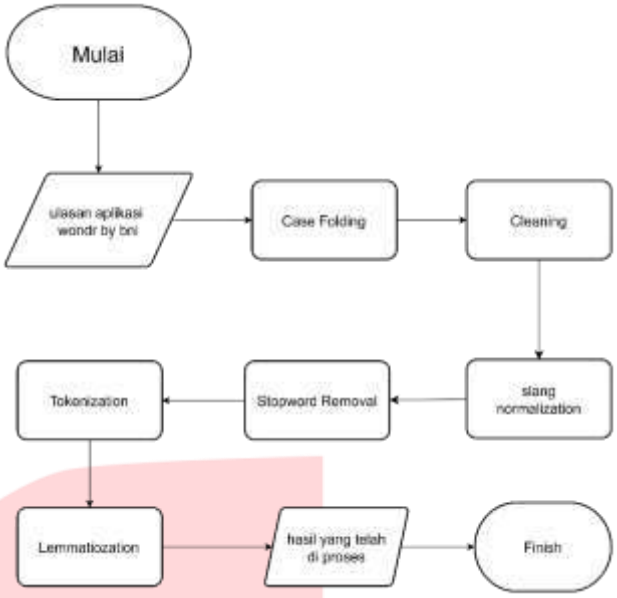
Pada penelitian ini, pengumpulan data review aplikasi wondr by BNI pada platfrom *google play store* dilakukan dengan metode *scraping* dengan pustaka python *google-play-scraper*. Setelah melakukan scraping, data yang didapat untuk penelitian ini sebanyak 16.001. Data yang didapat berisi tentang review aplikasi mulai dari kenyamanan pengunaan, keamanan penggunaan, dan lain sebagainya. Jumlah total perkelas Quality in Use sebagai berikut *Context Coverage* berjumlah 820, kelas *Freedom from Risk* berjumlah 1051, kelas *Effectiveness* berjumlah 2539, kelas *Efficiency* berjumlah 2710, kelas *Satisfaction* berjumlah 5888, dan kelas uncategoriz berjumlah 2993 data.

B. Pelabelan

Setelah mendapatkan data, selanjutnya adalah proses labeling, diaman proses ini dilakukan secara manual menggunakan pendekatan multi-class dengan total 5 label Quality in Use dan 1 label Uncategorized. Pelabelan manual pada penelitian ini berfungsi untuk memastikan bahwa setiap dataset review aplikasi diberikan label dengan akurat dan sesuai dengan konteks sesuai dengan Quality in Use.

C. Preprocessing data

Tahapan preprocesing data merupakan proses dari pembersihan data. Pada tahapan preprocessing data terdapat beberapa proses mulai dari *Case folding*, *Cleaning*, *Stopword removal*, *Tokenization*, *Lemmatization*. Proses ini berfungsi untuk mempermudah algoritma dalam memproses data serta mempermudah TF-IDF untuk mengubah dalam bentuk numerik. Alur dari preprocessing data ini ditunjukkan pada gambar



GAMBAR 2
(ALUR PREPROCESSING DATA)

D. Pembagian Data

Pembagian data dilakukan agar mempermudah algoritma untuk melatih data, dimana nantinya data akan dibagi menjadi 2 data uji dan data latih. Akan tetapi, pembagian data ini dilakukan dengan menggunakan teknik *K-Fold Cross Validation* yang berfungi untuk membagi data dengan beberapa nilai K. pada penelitian ini, proses pembagian data dilakukan secara berulang dengan beberapa variasi nilai K mulai dari K 5 hingga K 10, proses pengujian dengan variasi nilai K bertujuan untuk menguji peforma model secara keseluruhan serta mendapatkan hasil yang optmila dari setiap nilai K.

E. Pembobotan TF-IDF

Setelah proses pembagian data menggunakan teknik K-Fold Cross Validation, selanjutnya adalah proses dari pembobotan kata dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dimanan nantinya setiap kata akan dijadikan dalam bentuk numerik agar mempermudah pelatihan data dengan algoritma Multinomial Naive Bayes.

F. *Oversampling*

Setelah proses ekstraksi fitur teks, proses selanjutnya adalah oversampling. Pada penelitian ini, proses oversampling di gunakan untuk mengatasi ketidakseimbangan pada kelas (*Class Imbalance*) pada ISO Quality in Use, dimana metode oversampling ini hanya di gunakan pada data latih. Teknik *oversampling* ini menggunakan metode *SMOTE* dan *ADASYN*, yang mana dapat mengatasi ketidak seimbangan pada data *multi-class*. Proses *oversampling* pada penelitian ini menggunakan rasio 60%, 80%, dan 100%, dimana rasio ini sudah di rancang sedemikian rupa agar memberikan proporsi yang cukup untuk mengatasi ketidak seimbangan pada model.

G. Pembagian Data

Setelah proses oversampling pada data latih, selanjutnya proses klasifikasi model dengan algoritma Multinomial Naive Bayes. Agar proses pelatihan model ini lebih optimal, model algoritma Multinomial Naive Bayes ditambahkan parameter (*alpha*) terbaik melalui *Grid Search* dengan nilai kandidat yang sudah di tentukan mulai dari [0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1.0, 1.5] pada setiap iterasi. Model dilatih

menggunakan data latih yang telah melewati proses *Oversampling*.

H. Evaluasi Hasil

Tahapan selanjutnya adalah evaluasi hasil dari pelatihan model untuk mengevaluasi kinerja algoritma *Multinomial Naive Bayes*. Evaluasi hasil ini menggunakan teknik *Confusion Matrix*, metode ini dipilih agar menampilkan akurasi dari setiap label Multi-Class ISO Qualtiy in Use secara keseluruhan.

I. Skenario Pengujian

Pada penelitian ini, dilakukan dua skenario pengujian yang digunakan untuk mengevaluasi kinerja multi-class terhadap data label review aplikasi Wondr by BNI. Adapun skenario yang digunakan adalah sebagai berikut:

1. Penerapan *K-Fold Cross Validation* dalam membagi data uji dan data latih pada dataset multiclass review Wondr by Bni menggunakan *ISO Quality in Use* dengan algoritma *Multinomial Naive Bayes*.Pengujian ini, model diuji dengan *K-Fold Cross Validation* dengan berbagai nilai K (5,6,7,8,9, dan 10) untuk mengevaluasi model dalam membagi data secara proprosional. Pengujian ini bertujuan untuk mendapatkan nilai K yang menghasilkan performa terbiak berdasarkan matrix evaluasi.
2. Pengujian selanjutnya adalah menerapkan metode *oversampling* pada data latih menggunakan metode *SMOTE* dan *ADASYN* dengan beberapa kelipatan rasio dari 60%, 80% dan 100%. Output dari pengujian ini adalah mendapatkan rasio tuningan yang paling terbaik, serta membandingkan hasil pengujian sebelum *oversampling* dan sesudah *oversampling*.

J. Pembangunan Website

Setelah hasil evaluasi didapat, proses selanjutnya adalah pembaginan website. Website pada penelitian ini berfungsi untuk menampilkan hasil dari evaluasi model serta menampilkan *Confusion Matrix*, wordcloud dan menmpilkan input data dimana pengguna web dapat memasukkan kata dalam web dan akan ditampilkan preprocessing dan juga hasil label dari kalimat tersebut. Website ini dibangun dengan menggunakan fremwork Flask, yang mana dapat di integrasikan antar komponen *backand* dan *frontend*.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Proses awal, melakukan *scraping* untuk pengumpulan data. Proses ini mengambil data review aplikasi Wondr by BNI dari google play store pada rentan waktu mulai dari bulan Juli 2024-April 2025 dengan total hasil crawling medapatkan 16.001 data review aplikasi Wondr by BNI.

B. Pelabelan

Setelah data didapat, langkah selanjutnya adalah proses pelabelan pada data review aplikasi menggunakan ISO Qualtiy in Use 25010 yang mana diklasifikasikan menurut label ISO Quality in Use seperti Efficiency, Effectivenes, Satisfaction, Freedom from Risk, dan Context Coverage, lalu label dilaur Quality in Use untuk melabeli komentar yang tidak mengarah disalah satu Quality in Use yakni Uncategorized. Labeling dilakukan dengan 3 orang dengan sistem mayoritas voting.

C. Preprocessing Data

Preprocessing data dilakukan untuk membersihkan data agar memudahkan algoritma untuk melakukan pelatihan multi-class classification. Proses dari pembersihan data ini ada beberapa tahapan mulai dari *Case folding*, *Cleaning*, *Slank normalization*, *Tokenisasi*, *Stopword Removal*, *Lemmatization*. Proses ini dimulai dengan mengubah kalimat yang terdapat huruf kapital menjadi kecil agar tidak menggaggu proses pelatihan model. Lalu proses selanjutnya adalah membersihkan kata-kata dari elemen yang tidak dibutuhkan seperti, titik koma, tanda baca, angka, dan lain sebagainya. Lalu proses selanjutnya adalah mengubah kata yang tidak baku menjadi kata baku, dengan bantuan kamus yang telah di sediakan. Selanjutnya, mengurai kata-kata menjadi beberapa token melalui proses tokenisasi. Selanjutnya menghapus kata-kata umum dan kurang bermakna dalam teks yang dapat mempengaruhi kerja model di hapus, setelah itu proses terakhir, setiap kata akan digabungkan dengan proses linguistrik komputasional dalam bentuk dasar atau dengan kamus (lema).

D. Pembagian Data

Setelah tahapan preprocessing data, selanjutnya membagi data bersih menggunakan metode K-Fold Cross-Validation dengan beberapa jumlah fold. Dalam penelitian ini, penerapan nilai K-Fold 5 hingga 10, dimana setiap fold nya akan dibagi menjadi data latih dan data uji.

E. TF-IDF

Setelah proses pembagian data dengan K-Fold, selanjutnya proses ekstraksi fitur pada teks menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Proses ini mengubah sebuah teks menjadi bentuk numerik, proses ini digunakan agar dapat dipahami oleh algoritma machine learning.

F. *Oversampling*

Setelah transformasi teks menggunakan proses TF-IDF, dataset teridentifikasi adanya ketidakseimbangan distribusi antar kelas pada data latih, label ISO Quality in Use pada kelas *Satisfaction* memiliki jumlah lebih banyak ketimbang kelas-kelas Quality in Use lainnya. Ketimpangan ini menjadikan dataset akan bias serta, ketika diproses dengan algoritma pelatihan akan condong pada kelas mayoritas. Teknik *oversampling* digunakan untuk mengatasi hal ini, dengan cara menambah jumlah kelas Quality in Use lainya seperti *Efficiency*, *Effectivenes*, *Freedom from Risk*, *Context Coverage* serta kelas di luar Quality in Use yakni *Uncategorized* agar seimbang dengan kelas *Satisfaction*. Setiap kelas yang diproses K-Fold Cross Validation menerapkan proses ini.

G. Klasifikasi Model

Setelah data pelatihan dilakukan pengumbahan teks menjadi bentuk numerik dan penyeimbangan kelas melalui *oversampling*, model dilatih menggunakan algoritma Multinomial Naive Bayes yang dipilih karena mudah untuk memproses data multi-class. Agar mendapatkan hasil dan performa yang optimal, dilakukan pencarian nilai parameter alpha terbaik melalui Grid Search dengan kandidat [0.01, 0.05, 0.1, 0.3, 0.5, 0,7, 1.0, 1.5].Setelah itu, setiap iterasi pada model dilatih menggunakan data latih yang seimbang dan diuji dengan data uji, selanjutnya akurasi yang didapat akan dievaluasi.

H. Evaluasi Model

Tahapan awal, dilakukan pengujian model dengan berbagai variasi nilai K 5 hingga 10 untuk menentukan pembagian data latih dan data uji paling efektif pada pelatihan model. Tujuan dari penerapan nilai K-Fold ini untuk mengindentifikasi nilai mana yang menghasilkan prediksi yang paling optimal. Proses selanjutnya adalah mencari parameter *best alpha* yang optimal pada algoritma *Multinomial Naive Bayes* melalui *Grid Search*. Nilai *alpha* yang diujikan mulai dari [0.01, 0.05, 0.1, 0.3, 0.5, 0,7, 1.0, 1.5]. Setiap kombinasi antar *fold* dan alpha serta rasio teknik *oversampling ADASYN* dan *SMOTE* disetiapn *fold* nya akan dievaluasi berdasarkan matrix untuk menilai dan efektivitas model.

I. Pembahasan

Pada tahapan ini, dilakukan sebuah analisis terhadap performa model klasifikasi yang sudah dilakukan. Fungsi dari evaluasi ini untuk melihat kinerja seberapa baik model *Multinomial Naive Bayes* dalam mengenali pola data dan melakukan prediksi *multi-class* dengan tepat, terutama pada beberapa kelas *Quality in Use* dengan jumlah kelas Satisfaciton menjadi kelas mayoritas ketimbang kelas *Freedom from Risk, Efficiency, Context Coverage, Uncategorized,* dan *Effectiveness* dengan jumlah minoritas. Ketidak seimbangan jumlah antar kelas pada *Quality in Use* sangat berpotensi membuat model akan mempelajari pola kelas mayoritas, sehingga pola kelas dengan jumlah minoritas akan teralihkan.

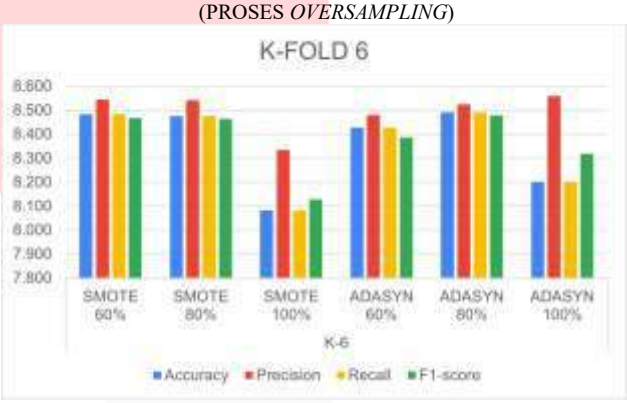
Pada proses evaluasi dilakukan dengan beberapa sampel, pertama menggunakan data asli yang tidak melewati proses *oversampling*. Lalu kedua, menggunakan data yang telah melewati proses *oversampling* dengan dua metode *oversampling* yang cocok untuk *multi-class*, yaitu *SMOTE* dan *ADASYN*. Masing-masing memiliki rasio yang berbeda mulai dari kelipatan 60%, 80%, 100%. Metode ini digunakan untuk memberikan hasil evaluasi dengan rasio *oversampling* mana yang kinerja terbaik.

Dari beberapa percobaan, *SMOTE* dengan rasio 60% memberikan hasil hampir sempurna, pada rasio ini *SMOTE* menghasilkan akurasi terbaiknya. Ini menendakan bahwa rasio 60% mampu bekerja dengan maksimal. *SMOTE* 80% juga tidak jauh berbeda dengan rasio 60%, *SMOTE* dibeberapa pengujian K-fold juga memberikan hasil terbaik. *SMOTE 100%* justru menjadi teknik *oversampling* dengan hasil yang kurang begitu maksimal, karena rasio ini justru model terlalu agresif dalam melakukan *oversampling* sehingga hasil yang diberikan kurang begitu baik.

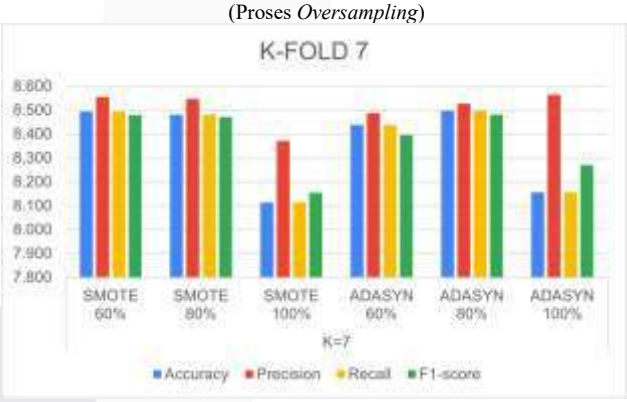
Pada *oversampling ADASYN* 60%, model tidak melakukan *oversampling*. Dari hasil yang di dapat, pengaruh dari karakter *ADASYN* yang adaptif ditambah dengan rasio *oversampling* yang terlalu rendah di 60%, serta strategi penentuan target per kelas menyebabkan tidak mengalami penambahan sampel sintetis untuk memproses data *multi-class*

ADASYN 80%, menjadikan metode *oversampling* yang bisa memberikan akurasi terbaik disetiap foldnya. Ini menandakan, bahwa rasio 80% bisa membuat dataset diproses secara maksimal, dibeberapa percobaan nilai K-Fold rasio ini menghasilkan akurasi diatas 84%. *ADASYN* 100%, menjadi teknik *oversampling* dengan hasil yang kurang maksimal, dibeberapa K-Fold justru hasil yang didapat menurun. Dapat disimpulkan, terlalu tinggi rasio yang diberikan, model akan bekerja lebih ekstra dan hasil yang didapat kurang begitu bagus.

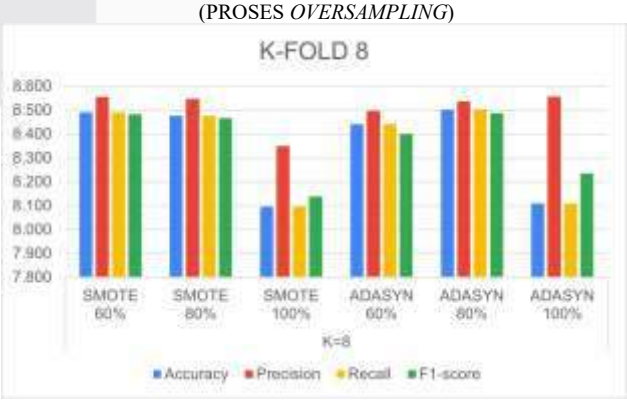
GAMBAR 3
(PROSES *OVERSAMPLING*)



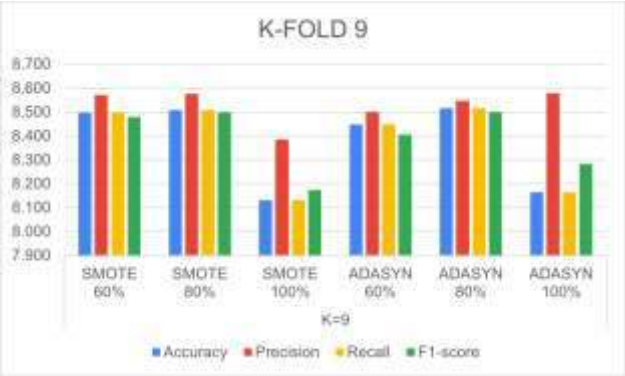
GAMBAR 4
(Proses *Oversampling*)



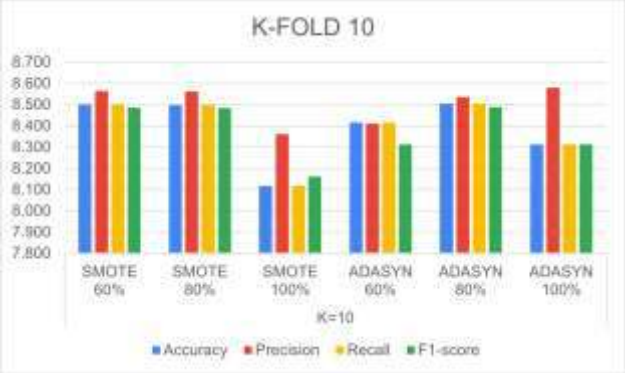
GAMBAR 5
(PROSES *OVERSAMPLING*)



GAMBAR 6
(PROSES *OVERSAMPLING*)



GAMBAR 7
(PROSES *OVERSAMPLING*)



GAMBAR 8
(PROSES *OVERSAMPLING*)

Dari hasil pengujian yang telah dilakukan mulai dari proses pengambilan data, pelabelan data, preprocesing data, TF-IDF atau mengubah teks dalam bentuk numerik, oversampling data, pelatihan model dengan algoritma *Multinomial Naive Bayes*. menunjukkan bahwa nilai K = 9 dengan Oversampling 80% *ADASYN* memberikan performa terbaik dan optimal dari segi matrix, ditunjukkan perolehan rata-rata Accuracy 0,8592, Precision 0,8590, Recall 0,8592, F1-Score 0,8572. Dari beberapa nilai K-Fold, K=9 memberikan kinerja pada model lebih unggul dan stabil dibandingkan nilai K lainnya

J. Visualisasi Website

Visualisasi hasil dari penelitian ini difungsikan unguk memaparkan hasil dari pelatihan model dengan Multinomial Naive Bayes. Dalam hal ini, website di bangun menggunakan Framework Flask sebagai backand, serta HTML, Tailwind. Dalam website penelitian ini, model yang digunakan untuk prediksi multi-class adalah model dari hasil yang terbaik yaitu model dari K-Fold 9

GAMBAR 9
(MODEL K-FOLD 9)

GAMBAR 10
(MODEL K-FOLD 9)

V. KESIMPULAN

Penelitian ini berhasil menciptakan model klasifikasi *multi-class* menggunakan algoritma *Multinomial Naive Bayes* dengan jumlah data 16.200 komentar review aplikasi Wondr by BNI yang berasal dari *Google Play Store*. Dengan berbagai proses mulai mencari data dengan teknik *scraping*, labeling, *preprocessing teks*, lalu *K-Fold CV*, serta proses pembobotan kata dengan TF-IDF, dan *overssampling* data untuk mengatasi ketidakseimbangan antar kelas. Model mampu mengklasifikasikan antar kelas *Quality in Use* dengan cukup baik. Hasil dari pengujian model menggunakan *K-Fold Cross Validation* dengan berbagai variasi nilai K (5 hingga 10) serta berbagai rasio *oversampling SMOTE* dan *ADASYN* (60%, 80%, 100%) menunjukkan bahwa nilai K=9 dengan rasio *oversampling ADASYN* 80% *Accuracy* 0,8516, *Precision* 0,8548, *Recall* 0,8516, dan *F1-score* 0,8501. Selain itu, hasil *alpha* terbaik menunjukkan bahwa *alpha* optimal bervariasi pada tiap K-Fold CV per rasio, dengan *alpha* 0.10 paling sering muncul di semua nilai *accuracy* tertinggi baik *SMOTE* maupaun *ADASYN*.

REFERENSI

[1] berita BNI, “wondr by BNI Catatkan 2 Juta Unduhan Sejak Diluncurkan, Jadi Game Changer di Industri Perbankan - Berita | BNI.” Accessed: Jun. 12, 2025. [Online]. Available: https://www.bni.co.id/id-id/beranda/kabar-bni/berita/articleid/23850

[2] E. Hasibuan and E. A. Heriyanto, “ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER,” *Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13–24, Oct. 2022, doi: 10.56127/JTS.V1I3.434.

[3] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, “ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE SENTIMENT ANALYSIS OF MOLA APPLICATION REVIEWS ON GOOGLE PLAY STORE USING SUPPORT VECTOR MACHINE ALGORITHM,” *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 5, no. 1, 2022.

[4] N. Khoirunnisaa, K. Nabila, N. Kesuma, S. Setiawan, A. Yunizar, and P. Yusuf, "KLASIFIKASI TEKS ULASAN APLIKASI NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SVM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 1, pp. 64–73, Jan. 2024, doi: 10.36080/SKANIKA.V7I1.3138.

[5] F. Fauzi, Ismatullah, and I. M. Nur, "Unbalanced multiclass classification with adaptive synthetic multinomial naive Bayes approach," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. T. 13, nr 3, no. 3, pp. 64–70, 2023, doi: 10.35784/IAPGOS.3740.

[6] F. Dewi and F. K. S. Dewi, "KLASIFIKASI BERITA MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES," *Scan : Jurnal Teknologi Informasi dan Komunikasi*, vol. 16, no. 3, pp. 1–8, Oct. 2021, doi: 10.33005/scan.v16i3.2870.

[7] C. V. Putra *et al.*, "ANALISIS DIGITAL MARKETING ANTARA APLIKASI MYBCA DAN WONDR BY BNI," *Jurnal Ilmiah Manajemen dan Akuntansi*, vol. 1, no. 6, pp. 24–33, Nov. 2024, doi: 10.69714/1QRT1R98.

[8] D. Rudini, D. G. Purnama, and A. A. Khan, "Penggunaan Teknik Web Scraping dalam Aplikasi Pengambilan Data dari Google Maps untuk Menunjang Digital Marketing," *Lentera: Multidisciplinary Studies*, vol. 2, no. 1, pp. 10–19, Nov. 2023, doi: 10.57096/LENTERA.V2I1.61.

[9] Rana M. Amir Latif, Syed Umair Aslam Shah, Farah Ijaz, M. Talha Abdullah, Muhammad Farhan, and Abdul Karim, "Data scraping from google play store and visualization of its content for analytics," Jan. 2019.

[10] R. Silviany Tantika, P. Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and U. Islam Bandung, "Bandung Conference Series: Statistics Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," 2022, doi: 10.29313/bcss.v2i2.3590.

[11] A. L. Pomerantsev and O. Y. Rodionova, "New trends in qualitative analysis: Performance, optimization, and validation of multi-class and soft models," Oct. 01, 2021, *Elsevier B.V.* doi: 10.1016/j.trac.2021.116372.

[12] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.

[13] L. Mayasari and D. Indarti, "KLASIFIKASI TOPIK TWEET MENGENAI COVID MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES DENGAN PEMBOBOTAN TF-IDF," *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 1, pp. 43–53, 2022, doi: 10.35760/IK.2022.V27I1.6184.

[14] M. N. Zahra, K. Kraugusteeliana, and P. Korespodensi, "ANALISIS KUALITAS PERFORMA APLIKASI DIGITAL BANKING X MENGGUNAKAN FRAMEWORK ISO 25010 EVALUATION OF QUALITY OF PERFORMANCE DIGITAL BANKING X APPLICATION USING FRAMEWORK ISO 25010," vol. 10, no. 3, 2023, doi: 10.25126/jtiik.2023106326.

[15] ISO/IEC 25010:2011, "Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models", Accessed: May 04, 2025. [Online]. Available: www.iso.org

[16] A. Tholib and Z. Arifin, "Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode TF-IDF dan Long Short-Term Memory (LSTM)," *Journal homepage: Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, 2024, doi: 10.33650/jeecom.v4i2.

[17] S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning," *Sensors 2023, Vol. 23, Page 2333*, vol. 23, no. 4, p. 2333, Feb. 2023, doi: 10.3390/S23042333.

[18] U. Negeri *et al.*, "Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN) Nanang Husin," *Jurnal Esensi Infokom*, vol. 7, no. 1, p. 75, 2023.