WFraud Alert Sebagai Prediksi Pesan Penipuan WhatsApp Menggunakan Naïve Bayes

[1]Ganis Sanhaji, Julian[3], [2]Herlan Syah
[1,2,3]Fakultas Teknik, Program Studi Teknik Elektro, Universitas Islam Nusantara, Kota Bandung, Indonesia
Email:ganissanhaji90@gmail.com, julianfikoma@gmail.com, syahherlan13333333@gmail.com

Abstrak− Penelitian ini menggambarkan dampak serius kejahatan siber dalam bentuk penipuan online terhadap masyarakat dan ekonomi Indonesia, penelitian ini juga membahas risiko dan kerugian finansial yang diakibatkan oleh penipuan online. Tujuan utama dari penelitian ini adalah membuat aplikasi WFraud Alert yang memiliki tujuan khusus untuk mengidentifikasi pesan WhatsApp dengan membedakan antara pesan normal, pesan penipuan, dan pesan judi online. Pada pendeskripsian masalah yang di teliti dapat terlihat bahwa penelitian ini berfokus pada dampak serius kejahatan siber, khususnya penipuan online, terhadap masyarakat dan ekonomi Indonesia. Tujuan penelitian ini adalah untuk mengidentifikasi dan mengukur dampak penipuan online dengan menggunakan data yang dikeluarkan oleh Kementerian Komunikasi dan Informatika (Kominfo). Berdasarkan laporan dari Agustus 2018 hingga 16 Februari 2023, teridentifikasi sebanyak 1.730 konten penipuan online. Selama lima tahun sebelumnya, kerugian akibat penipuan online di Indonesia mencapai total sekitar Rp 18,7 triliun. Aplikasi WFraud Alert ini memiliki tujuan untuk mengidentifikasi pesan WhatsApp yang terdiri dari pesan normal, pesan penipuan dan pesan judi online. Penelitian ini menggunakan metode kuantitatif dengan data primer sebanyak 156 data, yang terdiri dari pesan normal, penipuan, dan judi online. Algoritma Naive Bayes digunakan untuk melatih model aplikasi WFraud Alert. Algoritma Naive Bayes tetap relevan dan efektif dalam menangani masalah klasifikasi dalam berbagai konteks, terutama dalam Natural Language Processing. Hasil penelitian menunjukkan bahwa aplikasi ini memiliki performa yang kuat, dengan presisi mencapai 91%, *recall* mencapai 91%, dan *F1-score* mencapai 91. Hasil dari aplikasi WFraud Allert telah berhasil menjadi solusi yang efektif dalam mengidentifikasi penipuan dalam pesan WhatsApp.

Kata Kunci: Naïve Bayes, WhatsApp, Klasifikasi

Abstract− *This research illustrates the serious impact of cybercrime in the form of online fraud on Indonesian society and the economy. This research also discusses the risks and financial losses caused by online fraud. The main objective of this research is to create a WFraud Alert application which has the specific aim of identifying WhatsApp messages by distinguishing between normal messages, fraudulent messages and online gambling messages. In the description of the problem examined, it can be seen that this research focuses on the serious impact of cybercrime, especially online fraud, on Indonesian society and the economy. The aim of this research is to identify and measure the impact of online fraud using data released by the Ministry of Communication and Information (Kominfo). Based on reports from August 2018 to February 16 2023, 1,730 pieces of online fraudulent content were identified. Over the previous five years, losses due to online fraud in Indonesia reached a total of around IDR 18.7 trillion. The WFraud Alert application aims to identify WhatsApp messages consisting of normal messages, fraudulent messages and online gambling messages. This research uses quantitative methods with 156 primary data, consisting of normal messages, fraud and online gambling. The Naive Bayes algorithm is used to train the WFraud Alert application model. The Naive Bayes algorithm remains relevant and effective in dealing with classification problems in various contexts, especially in Natural Language Processing. The research results show that this application has strong performance, with precision reaching 91%, recall reaching 91%, and F1-score reaching 91. The results of the WFraud Alert application have succeeded in becoming an effective solution in identifying fraud in WhatsApp messages.*

Keywords*: Naïve Bayes, WhatsApp, Classification*

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah menciptakan konektivitas global yang sebelumnya tak terbayangkan, membuka pintu bagi pertukaran informasi yang cepat dan efisien di seluruh dunia melalui berbagai platform media sosial yang saat ini telah menjadi bagian integral dari kehidupan kita. Media sosial telah memainkan peran kunci dalam mengubah paradigma berkomunikasi dengan menghapuskan hambatan geografis, waktu, dan jarak secara tegas, memberikan kita kemampuan untuk berinteraksi dan berkomunikasi dengan siapa pun, di mana pun, dan kapan pun, tanpa perlu melakukan pertemuan tatap muka fisik. Dengan kemampuan ini, kebutuhan akan efisiensi dalam interaksi sosial pun semakin terpenuhi dengan sangat baik.[1]

Sebagai contoh kasus, dalam situasi ketika seorang teman mengalami sakit dan tidak dapat hadir di sekolah, mereka dapat dengan mudah mengejar pelajaran yang telah mereka lewatkan melalui berbagai platform media sosial seperti WhatsApp. Bahkan dalam kasus yang lebih mendalam, ketika kita berhadapan dengan situasi di mana pertemuan fisik dengan keluarga, kerabat, teman lama, atau seseorang yang telah berpisah selama satu dekade atau lebih seakan menjadi hal yang mustahil, media sosial menjadi jembatan yang sangat berharga, memungkinkan kita untuk menjaga komunikasi dan memelihara hubungan dengan orang-orang tersebut tanpa batasan geografis dan waktu. Sehingga dengan kejadian tersebut media sosial khususnya Whatsapp ini dapat menimbulkan indikasi atau peluang terjadinya penipuan atau tindak kejahatan oleh pihak yang tidak bertanggung jawab melalui pesan WhatsApp. Dari beberapa kasus yang beredar, pesan penipuan dalam konteks WhatsApp telah menjadi isu yang semakin memprihatinkan. Penipuan sering kali disamarkan sebagai komunikasi resmi,

memerlukan keahlian untuk mengidentifikasinya. Kementerian Komunikasi dan Informatika telah berupaya melawan penipuan online dengan pemblokiran ribuan situs berita palsu. Permasalahan semakin kompleks dengan munculnya pesan yang mengelabui penerima, seperti permintaan pulsa atau tindakan kejahatan lainnya.

Kementerian Komunikasi dan Informatika telah mengidentifikasi sekitar 800 ribu situs berita palsu yang tersebar luas di berbagai wilayah Indonesia. [2]Menariknya, pada tahun 2016, Kemkominfo juga melakukan pemblokiran terhadap 773 ribu situs web yang terbagi dalam 10 kategori yang berbeda. Kategori-kategori tersebut mencakup pornografi, narkoba, perjudian, isu-isu SARA (Suku, Agama, Ras, dan Antargolongan), penipuan/dagang ilegal, radikalisme, kekerasan anak, keamanan internet, dan pelanggaran hak kekayaan intelektual (HKI). Berdasarkan laporan dari Agustus 2018 hingga 16 Februari 2023, teridentifikasi sebanyak 1.730 konten penipuan online. Selama lima tahun sebelumnya, kerugian akibat penipuan online di Indonesia mencapai total sekitar Rp 18,7 triliun. Tindakan ini merupakan langkah signifikan dalam menjaga integritas serta keamanan informasi di ranah digital di Indonesia. Angka-angka ini mencerminkan dampak serius yang ditimbulkan oleh kejahatan siber dalam bentuk penipuan online terhadap masyarakat dan ekonomi Indonesia. Oleh karena itu, sangatlah penting bagi masyarakat untuk memprediksi konten isi pesan WhatsApp agar terhindar dari penipuan yang sering terjadi didalam masyarakat.[3]

Persepsi masyarakat terhadap pesan penipuan di WhatsApp telah menjadi isu yang semakin memprihatinkan. Pesan-pesan semacam ini sering kali mengecoh penerima dengan menyamar sebagai komunikasi resmi, mengingatkan kita akan pentingnya mengenali contoh-contoh pesan penipuan WhatsApp yang beragam. Dalam kasus pertama, pesan-pesan yang meminta pulsa telah menjadi permasalahan yang berulang selama beberapa tahun terakhir. Penipu sering kali mengambil identitas seseorang yang kita kenal untuk meminta pulsa, bahkan mereka dapat mengaku sebagai orang terdekat yang membutuhkan transfer uang atau diperintahkan untuk klik link dengan website yang tidak jelas. Sejumlah alasan seperti kecelakaan, pengobatan, biaya sekolah, atau alasan lainnya seringkali menjadi bahan penipuan yang digunakan.

Kajian mengenai klasifikasi pesan spam dalam bahasa Indonesia telah menjadi subjek penelitian yang cukup signifikan di masa lalu. Penelitian-penelitian sebelumnya telah mencoba berbagai metode pembelajaran mesin, termasuk namun tidak terbatas pada *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier (NBC)*, yang telah menunjukkan performa yang baik dalam melakukan klasifikasi, meskipun masih ada ruang untuk peningkatan lebih lanjut guna mencapai tingkat performa yang dianggap sebagai "ideal". Naive Bayes adalah suatu teknik klasifikasi yang mendasarkan pada perhitungan kemungkinan yang relatif sederhana, namun memiliki kemampuan untuk menggambarkan seluruh spektrum kemungkinan dengan cara yang memadukan beragam kombinasi dan frekuensi nilai-nilai yang terdapat dalam basis data yang dianalisis. Secara lebih mendalam, kinerja klasifikasi pesan spam dalam bahasa Indonesia masih memiliki potensi pengembangan yang signifikan. Salah satu pendekatan yang dapat dieksplorasi adalah melalui penerapan metode yang mampu menjalankan tugas klasifikasi teks dengan lebih efektif dan efisien secara keseluruhan. Dalam konteks ini, teknik-teknik terbaru dalam pembelajaran mesin seperti jaringan saraf tiruan (*neural networks*), yang melibatkan arsitektur seperti LSTM (*Long Short-Term Memory*) atau CNN (*Convolutional Neural Networks*), mungkin bisa menjadi solusi yang menjanjikan untuk mengatasi kompleksitas bahasa dalam pesan SMS berbahasa Indonesia.

Algoritma Naïve Bayes memiliki kelebihan yaitu tidak perlu memerlukan banyak data dalam pemprosesan datanya sehingga pada kemampuan ini untuk memberikan performa yang baik bahkan ketika data yang tersedia terbatas. Algoritma ini mampu menghasilkan estimasi yang cukup akurat meskipun hanya dengan dataset yang terbatas. Hal ini membuatnya menjadi pilihan yang menarik dalam situasi di mana ketersediaan data sangat terbatas atau sulit diperoleh. Selain kelebihan Naïve Bayes juga memiliki kekurangan yaitu Algoritma ini sering diakui sebagai estimator yang kurang baik, sehingga diperlukan kehati-hatian dalam menafsirkan hasil probabilitas dari metode 'predict_proba'. Hal ini disebabkan oleh asumsi bahwa semua fitur dalam model dianggap sebagai fitur yang independen satu sama lain. Namun, dalam konteks situasi nyata, mencari rangkaian fitur yang benar-benar independen bisa menjadi tugas yang rumit dan sulit ditemukan. [4]

Selain itu, dalam upaya meningkatkan kualitas klasifikasi spam, penting untuk mempertimbangkan pendekatan lintas-disiplin, yang mungkin melibatkan metode pengolahan bahasa alami (*Natural Language Processing* - NLP). Dengan menerapkan teknik-teknik NLP seperti *Tokenisasi*, *Stemming*, Lemmatisasi, dan penggunaan *Word Embeddings* seperti Word2Vec atau GloVe, model klasifikasi dapat lebih baik memahami struktur dan makna teks bahasa Indonesia, yang seringkali kaya akan variasi. Selain pemrosesan teks itu sendiri, aspek lain yang dapat memperbaiki performa klasifikasi termasuk penyaringan dan pembersihan data pelatihan untuk mengeliminasi noise dan sampel yang tidak relevan. Demikian pula, pertimbangan untuk menambahkan fitur tambahan seperti informasi pengirim, waktu pengiriman, atau atribut lainnya yang dapat memberikan wawasan tambahan dalam klasifikasi spam.[5]

Dalam upaya ini, penggabungan hasil dari beberapa model pembelajaran mesin menggunakan teknik ensemble learning dapat menjadi pendekatan yang efektif untuk meningkatkan akurasi dan konsistensi hasil klasifikasi. Pembaruan dan pengembangan dataset yang lebih besar dan bervariasi juga merupakan langkah penting dalam menghadapi permasalahan klasifikasi pesan spam dengan lebih efektif dan akurat. Selanjutnya, dalam mengukur kemajuan dan hasil penelitian ini, evaluasi dan validasi yang ketat menggunakan metrik yang

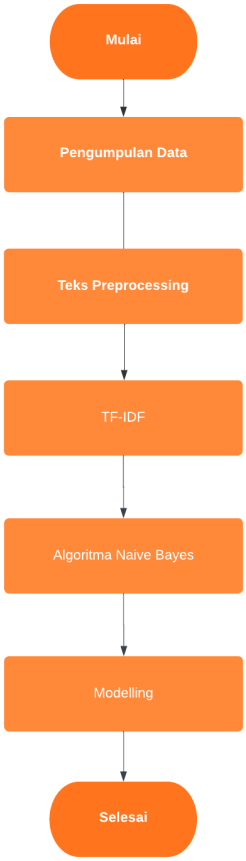
sesuai merupakan aspek yang tidak dapat diabaikan. Dengan pendekatan holistik yang mencakup teknologi pembelajaran mesin, pengolahan bahasa alami, dan pengembangan dataset, diharapkan akan ada perbaikan yang signifikan dalam klasifikasi pesan spam berbahasa Indonesia, mendekati atau bahkan mencapai performa yang ideal.[6]

Berdasarkan uraian di atas, penelitian ini akan memanfaatkan Algoritma Naive Bayes sebagai algoritma utama untuk melakukan pengklasifikasian pesan WhatsApp yang menggunakan bahasa Indonesia. Metode ini akan digunakan dalam rangka mengidentifikasi dan memisahkan pesan-pesan berdasarkan konteks dan kategori tertentu, sehingga memungkinkan untuk memahami isi pesan WhatsApp secara lebih mendalam dan efisien.

2. METODE PENELITIAN

2.1 Tahap Penelitian

Tahapan ini merupakan suatu pendekatan yang secara terperinci menjelaskan seluruh prosedur yang terlibat dalam pembuatan program yang bertujuan untuk melakukan prediksi pesan penipuan di *platform* Whatsapp. Proses ini melibatkan peneliti dalam mengumpulkan data yang relevan dan signifikan untuk kemudian diolah menjadi sebuah informasi yang dapat digunakan untuk pengklasifikasian pesan tersebut. Seluruh langkah-langkah dalam proses ini, sebagaimana diilustrasikan dalam Gambar 1 yang terdapat dalam *flowchart* penelitian.



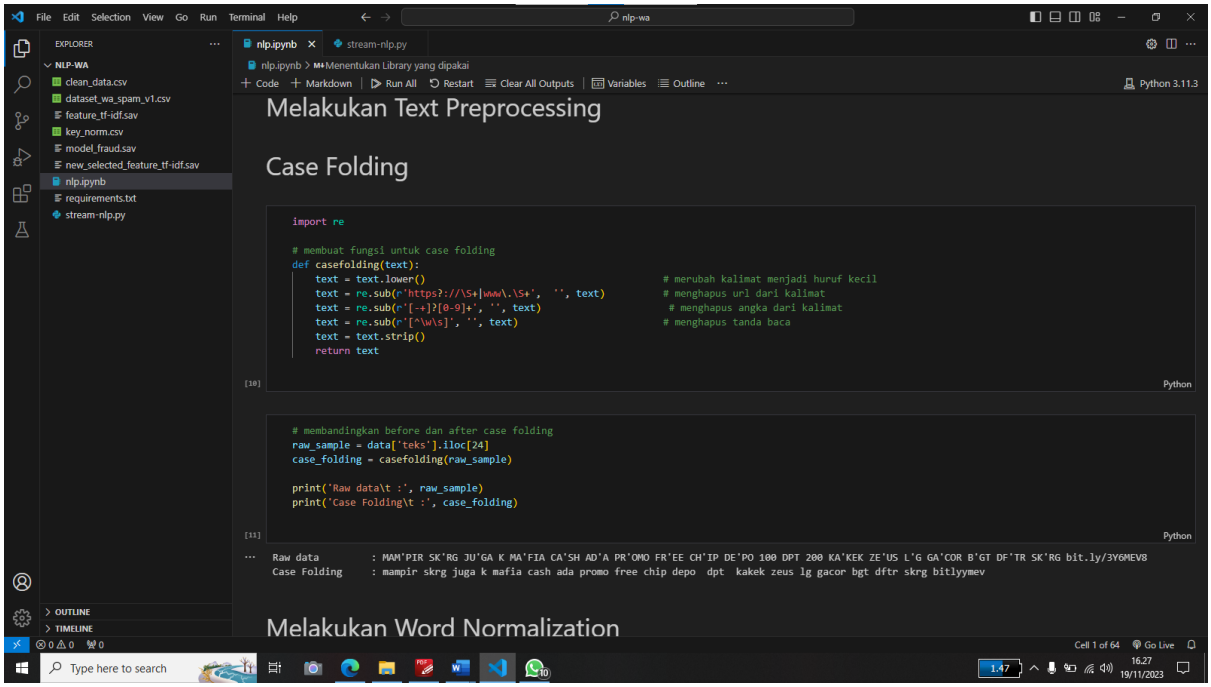
Gambar 1. *Flowchart* Penelitian

Dari Gambar 1 kita dapat pahami bahwa tahap mulai adalah proses menentukan ide penelitian yang akan dilakukan dan dilanjutkan dengan proses pencarian data yang sesuai dengan ide penelitian. Setelah kita menentukan ide lanjut ketahap selanjutnya yaitu pengumpulan data. Kami mengumpulkan dari aplikasi WhatsApp masing-masing yang disatukan ke Microsoft Excel lalu kami ubah dalam bentuk CSV. Data yang kami peroleh berupa data normal sebanyak 52 data, data promosi judi online sebanyak 52 data, dan data penipuan sebanyak 52 data. Jika dijumlahkan maka semuanya ada 156 data. Selanjutnya yaitu tahap teks preprocessing, Proses preprocessing dalam analisis data atau pembelajaran mesin (machine learning) melibatkan dua tahap yang dilakukan secara terpisah pada data training dan data testing. Tujuan utama dari pemisahan ini adalah untuk memudahkan dan memastikan integritas dalam seluruh proses analisis data dan pembelajaran mesin. Proses ini juga memiliki tujuan lainnya, yaitu melakukan berbagai langkah transformasi dan penyiapan data guna mencapai

dua sasaran utama: pertama, membersihkan data teks dari berbagai noise dan gangguan yang mungkin terkandung di dalamnya, dan kedua, mempersiapkan data teks sedemikian rupa sehingga memungkinkan pemrosesan selanjutnya untuk menghasilkan hasil yang lebih akurat. Teks preprocessing yang kami gunakan ada 4 yaitu, case folding, word normalization, filtering, dan stemming.

1. Case Folding

Case folding adalah langkah yang hampir selalu diterapkan dalam pemrosesan teks, dan ini dilakukan dengan alasan yang sangat penting. Data teks yang kita peroleh seringkali tidak terstruktur dan memiliki konsistensi yang rendah dalam penggunaan huruf kapital. Selain mengubah huruf, case folding juga dapat melibatkan langkah-langkah tambahan, seperti penghapusan URL yang terdapat dalam teks, serta eliminasi tanda baca yang tidak relevan. Kami disini merubah huruf kapital menjadi huruf kecil, menghapus url dari pesan yang didapatkan, menghapus angka dari pesan, dan menghapus tanda baca. [7]Berikut proses case folding yang telah kami buat dapat dilihat dari Gambar 2.



Gambar 2. Proses case folding

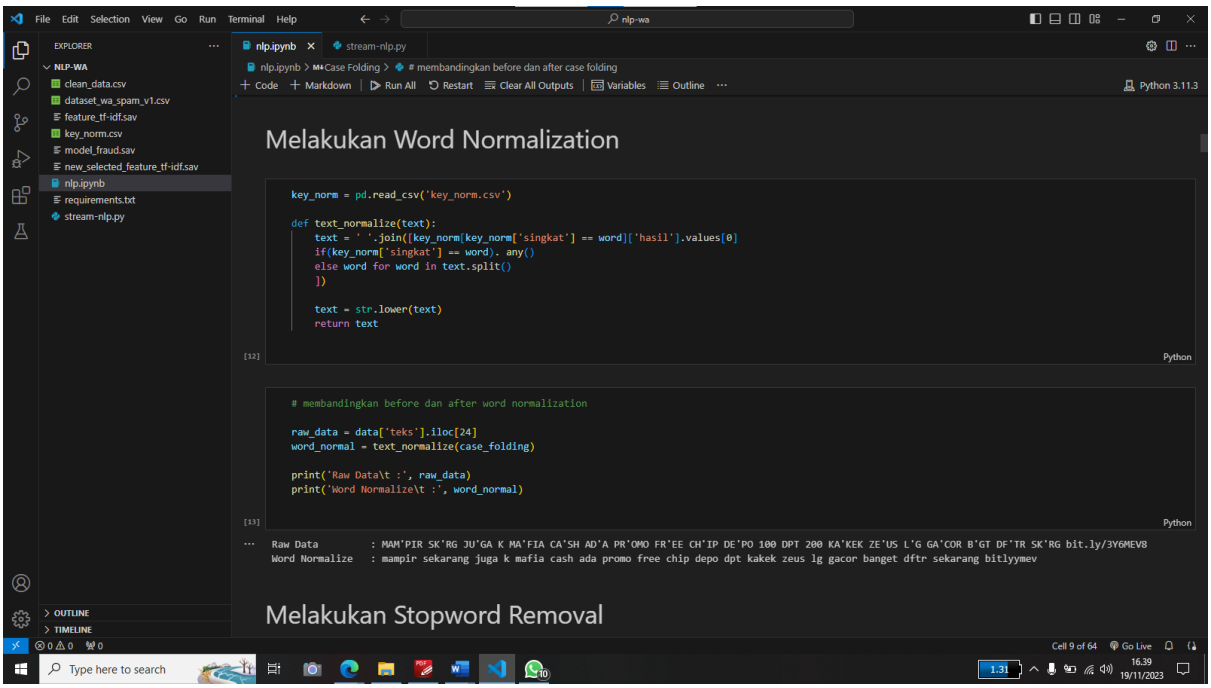
Dalam proses ini dapat kita lihat bahwa proses case folding yang telah kami lakukan berjalan dengan baik. Proses ini sudah merubah huruf kapital menjadi huruf kecil, menghapus url dari pesan yang diterima, menghapus angka dari pesan dan menghapus tanda baca. Sebagai lebih jelasnya dapat dilihat dari Tabel 1 proses case folding yang telah kami proses.

Tabel 1. Proses *case folding*

Pesan Asli	*Case Folding*
MAM'PIR SK'RG JU'GA K MA'FIA CA'SH AD'A PR'OMO FR'EE CH'IP DE'PO 100 DPT 200 KA'KEK ZE'US L'G GA'COR B'GT DF'TR SK'RG bit.ly/3Y6MEV8	mampir skrg juga k mafia cash ada promo free chip depo dpt kakek zeus lg gacor bgt dftr skrg bitlyymev

2. Word Normalization

Word normalization adalah satu dari sekian banyak tahapan esensial dalam proses teks preprocessing yang dilakukan untuk mempersiapkan data teks sebelum analisis lebih lanjut. Fokus utama dari word normalization adalah mengatasi variasi bentuk kata-kata yang mungkin ada dalam teks dan mengubahnya menjadi bentuk yang lebih konsisten serta seragam. Seringkali pesan yang diterima masih berupa singkatan yang membuat para pengguna aplikasi WhatsApp kebingungan, dengan proses word normalization kita dapat mengubah dari kata singkat menjadi kata yang seharusnya. [8]Berikut proses word normalization yang telah kami buat dapat dilihat dari Gambar 3.



Gambar 3. Proses word normalization

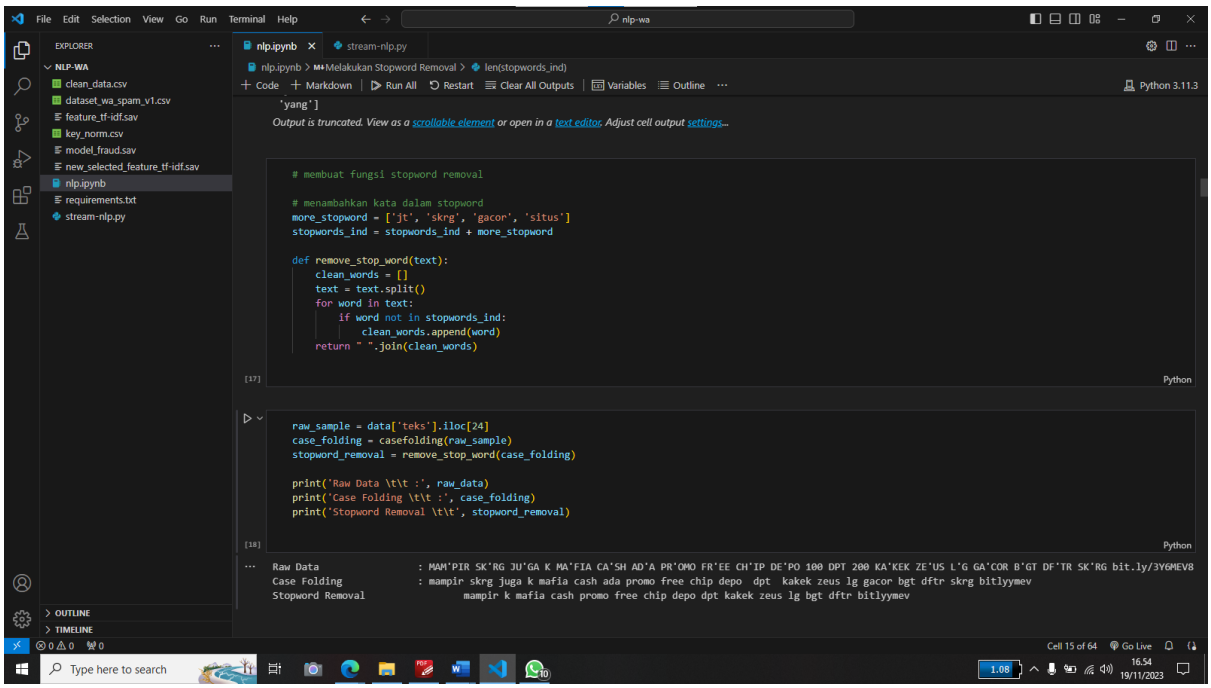
Dalam proses ini dapat kita lihat bahwa proses word normalization yang telah kami lakukan berjalan dengan baik. Proses ini sudah sudah melakukan tugasnya yaitu membuat kata singkatan menjadi kata yang seharusnya. Sebagai lebih jelasnya dapat dilihat dari Tabel 2 proses word normalization yang telah kami proses.

Tabel 2. Proses *word normalization*

Case Folding	*Word Normalization*
mampir skrg juga k mafia cash ada promo free chip depo dpt kakek zeus lg gacor bgt dftr skrg bitlyymev	mampir sekarang juga k mafia cash ada promo free chip depo dapat kakek zeus lagi gacor banget daftar sekarang bitlyymev

3. Filtering (Stopword Removal)

Menurut Triawati (2009), konsep stoplist atau stopword adalah komponen penting dalam pemrosesan teks, terutama dalam pendekatan bag-of-words. Stopword merupakan suatu proses dalam pemrosesan bahasa alami yang melibatkan pembuangan atau penghilangan imbuhan-imbuhan yang terdapat dalam kata-kata yang ada dalam sebuah teks, dan seringkali dikenal sebagai stoplist. [9]Hal ini dilakukan karena kata-kata yang mengandung imbuhan tersebut seringkali tidak memiliki nilai informatif yang signifikan dalam konteks analisis teks atau klasifikasi teks. Tahapan pembersihan stopword ini sebenarnya merupakan langkah awal yang signifikan dalam proses pengolahan teks. Dengan menghilangkan stopwords, kita dapat mengurangi noise dan meningkatkan ketepatan analisis teks. Pada proses ini kami menghilangkan beberapa kata yang sering muncul dalam pesan diantaranya yaitu skrg, jt, gacor, dan situs. Berikut proses stopword removal yang telah kami buat dapat dilihat dari Gambar 4.



Gambar 4. Proses stopword removal

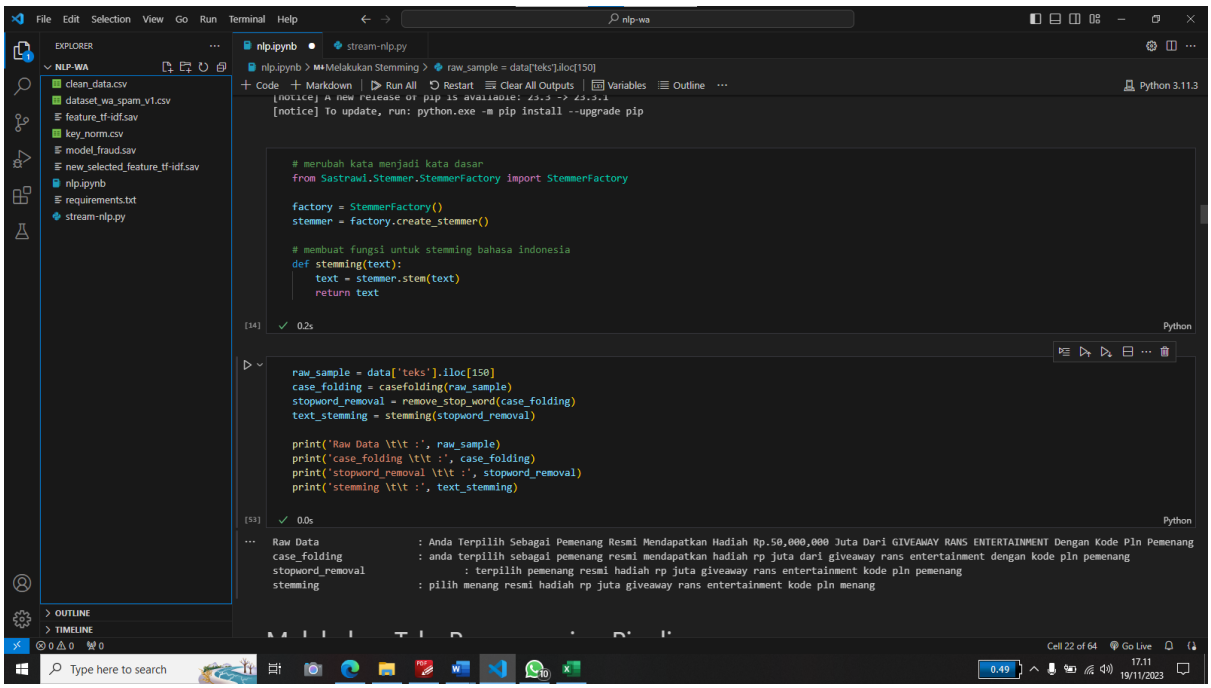
Dalam proses ini dapat kita lihat bahwa proses stopword removal yang telah kami lakukan berjalan dengan baik. Proses ini sudah sudah melakukan tugasnya yaitu menghapus beberapa kata yang ingin kita hapus. Sebagai lebih jelasnya dapat dilihat dari Tabel 3 proses stopword removal yang telah kami proses.

Tabel 3. Proses *stopword removal*

Pesan Asli	*Filtering*
MAM'PIR SK'RG JU'GA K MA'FIA CA'SH AD'A PR'OMO FR'EE CH'IP DE'PO 100 DPT 200 KA'KEK ZE'US L'G GA'COR B'GT DF'TR SK'RG bit.ly/3Y6MEV8	mampir k mafia cash promo free chip depo dpt kakek zeus lg bgt dftr bitlyymev

4. Stemming

Stemming merupakan salah satu teknik penting dalam teks preprocessing yang digunakan untuk mengubah kata-kata dalam teks ke bentuk dasarnya dengan menghapus akhiran kata. Stemming adalah salah satu teknik penting dalam preprocessing data teks yang digunakan untuk menghilangkan infleksi kata sehingga kata-kata tersebut dapat direduksi ke bentuk dasarnya. Teknik stemming sangat berguna dalam analisis teks karena memungkinkan kata-kata yang berasal dari akar yang sama dianggap sebagai bentuk yang sama, sehingga mengurangi kompleksitas dalam pemahaman teks. [10]Berikut salah satu contoh proses pesan yang telah melalui proses stemming yang terdapat pada Gambar 5.



Gambar 5. Proses stemming

Dalam proses ini dapat kita lihat bahwa proses stemming yang telah kami lakukan berjalan dengan baik. Proses ini sudah sudah melakukan tugasnya yaitu menghilangkan infleksi kata sehingga kata-kata tersebut dapat direduksi ke bentuk dasarnya. Sebagai lebih jelasnya dapat dilihat dari Tabel 4 proses stemming yang telah kami proses.

Tabel 4. Proses *stemming*

Pesan Asli	*Stemming*
Anda Terpilih Sebagai Pemenang Resmi Mendapatkan Hadiah Rp.50,000,000 Juta Dari GIVEAWAY RANS ENTERTAINMENT Dengan Kode Pln Pemenang	pilih menang resmi hadiah rp juta giveaway rans entertainment kode pln menang

2.2 Analisis Data

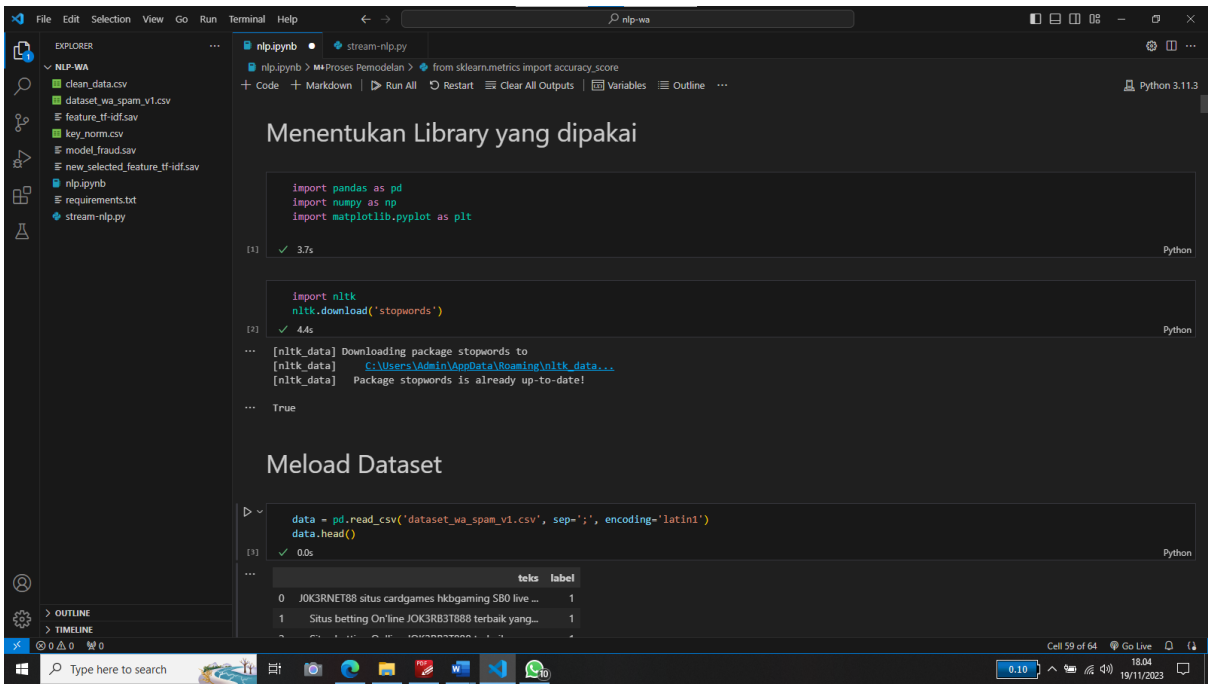
2.2.1. Teknik Pengumpulan Data

Teknik pengumpulan data yang kami lakukan merujuk pada beragam strategi dan metode yang diterapkan untuk menghimpun informasi yang relevan dengan tujuan studi. Kami menggunakan teknik mengumpulkan data Primer. Proses pengumpulan data primer adalah tahapan yang berkaitan dengan pengambilan informasi atau data secara langsung dari sumber asli atau subjek yang relevan dengan penelitian atau studi yang sedang berlangsung. Data yang kami miliki sebanyak 156 pesan yang mencakup pesan normal, penipuan, dan judi online yang kita peroleh dari aplikasi WhatsApp masing-masing yang kami kumpulkan lalu kami rekap di Microsoft Excel untuk dibuat file CSV. [11]Berikut tabel rincian jenis pesan yang digunakan pada Tabel 5.

Tabel 5. Jenis-jenis pesan yang digunakan

Jenis Pesan	Label	Jumlah
Normal	0	52
Judi Online	1	52
Penipuan	2	52

Setelah kami kumpulkan datanya dan telah di konversi ke file CSV. Kami menentukan library yang akan dipakai dan meload dataset yang telah kami buat. Berikut langkah pemanggilan library dan meload dataset yang dilakukan terdapat pada Gambar 6.



Gambar 6. Memanggil library dan meload dataset

2.2.2. Pelabelan Data

Pelabelan data dari penelitian ini peneliti melakukan pelabelan secara manual dengan menyesuaikan label data yaitu label 0,1, dan 2. Label 0 dikategorikan sebagai pesan WhatsApp yang normal, label 1 dikategorikan sebagai pesan WhatsApp promosi judi online, dan label 2 dikategorikan sebagai pesan WhatsApp penipuan. Pelabelan ini sudah dilakukan dan dapat dilihat teks beserta labelnya di tabel 2. Pada bagian ini juga merupakan salah satu proses untuk mendapatkan hasil representasi corpus yang diharapkan. Kami melakukannuya secara manual melalui Microsoft Excel lalu di konversi ke file CSV. Karena jika memakai otomatis harus menggunakan aplikasi tambahan. [12]Berikut contoh teks beserta labelnya sebagai identifikasi data yang tercantum pada Tabel 6.

Tabel 6. Teks beserta labelnya

Teks	Label
Assalamualaikum warahmatullahi wabarakatuh pak selamat siang saya ijin bertanya pak apakah hari ini bapak sedang di kampus pak?	0
Berani depo minimal 20rb aja ak bantu kknya maxwin 1juta skrg jga di game yang kk kuasai ak setting maxwin di bet 200 lngsung di buy freespin pertama ak berani jaminkan kemenangan 100% akun kamu di situs akuLINK : https://bit.ly/Linkgacorgressya	1
Nasabah Yth" Harap ke kantor Hari ini karena REKENING tabungan Anda ERROR & harus Dimonitor ulang. Sebelum Ke kantor harap hubungi WAWAN 081221188837 tks	2

2.3. Naïve Bayes

Naive Bayes, dalam hakikatnya, menggambarkan suatu kerangka kerja matematis yang tumbuh dari akar-akar teori statistik dan probabilitas. Meskipun telah menjadi landasan yang cukup mapan dalam dunia komputasi dan ilmu data, algoritma ini terus mempertahankan relevansinya dalam era *machine learning (ML)* yang tengah berkembang pesat, terutama ketika dikaitkan dengan masalah-masalah yang terus berubah dalam domain NLP atau *Natural Language Processing*. [13]Salah satu keunggulan utama dari Naive Bayes adalah sifatnya yang sederhana, namun memiliki kemampuan bersaing yang cukup kuat dengan berbagai model algoritma lainnya dalam berbagai aplikasi analisis data. Sebagai ilustrasi, kita dapat melihat bagaimana aplikasi Naive Bayes telah membawa manfaat besar dalam hal seperti klasifikasi dokumen, di mana algoritma ini dapat secara efisien mengelompokkan dan mengidentifikasi dokumen berdasarkan ciri-ciri tertentu. Selain itu, dalam pertempuran melawan banjir pesan spam yang mengganggu, Naive Bayes telah menjadi senjata utama. Yang membuatnya semakin menonjol adalah kemampuannya untuk beroperasi dengan cukup baik bahkan saat hanya memiliki

jumlah data pelatihan yang terbatas, karena mampu memperkirakan parameter-parameter yang diperlukan dengan efisien. Dengan demikian, Naive Bayes telah membuktikan diri sebagai alat yang tak ternilai dalam menangani berbagai permasalahan klasifikasi dalam berbagai konteks dunia nyata.
Dasar dari penggunaan Naïve Bayes dalam pemrograman adalah rumus Bayes, di mana peluang kejadian h sebagai D ditentukan oleh peluang D saat h terjadi, peluang h, dan peluang D. Persamaan yang digunakan dapat diungkapkan sebagai berikut :

$$P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$$

Keterangan :
h : Hipotesis data dari suatu kelas yang telah ditentukan
D : Data yang belum ada kelasnya
P(h) : Probabilitas dari suatu analisis
P(D) : Probabilitas dari data D
P(h|D) : Probabilitas data h berdasarkan kondisi D
P(D|h) : Probabilitas D berdasarkan hipotesis A

Model klasifikasi Naïve Bayes yang dikenal sebagai Multinomial Naïve Bayes Classifier (NBC) merupakan sebuah pendekatan yang disederhanakan dari Metode Bayes. Model ini secara khusus cocok untuk tugas klasifikasi teks atau dokumen sebagai berikut :

$$V_{MAP} = \arg\max P\,(v_j \mid a_1, a_2, a_3, \ldots a_n)$$

Dengan dua rumus tersebut dapat kita gabungkan menjadi :

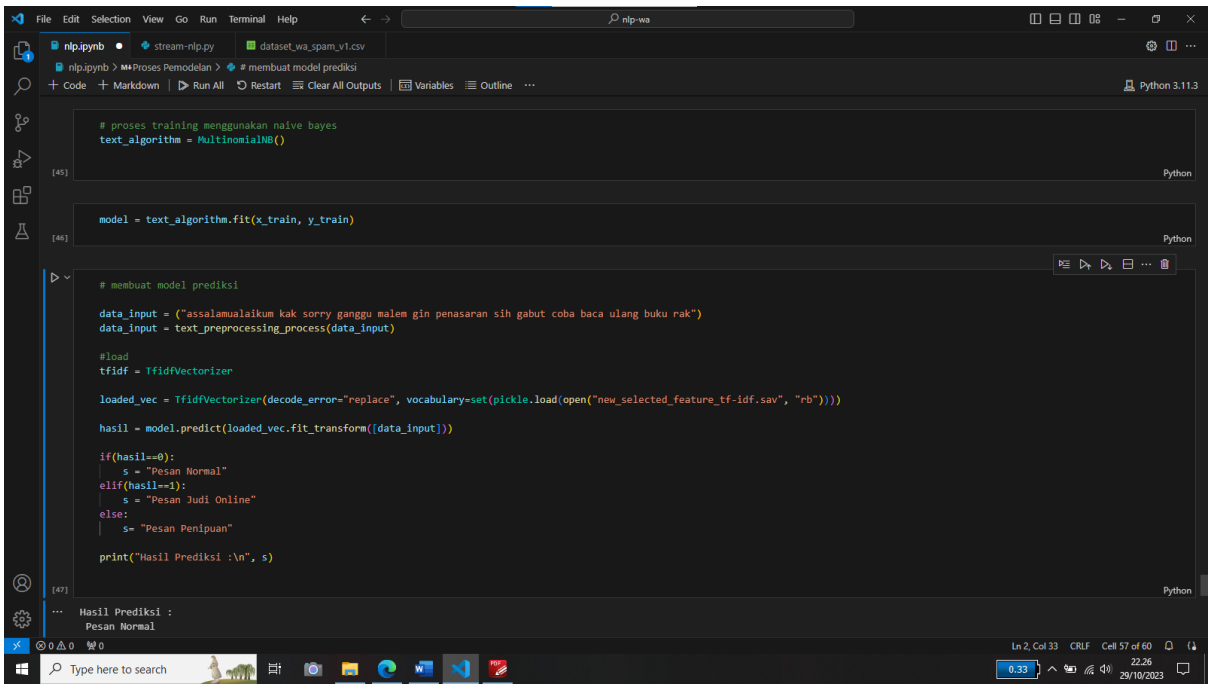
$$\text{VMAP} = \text{agr max (Vj eV)}\ \frac{\text{P (a1,a2,a3,... an | Vj)}}{P\,(a1,a2,a3,\ldots an)}$$

3. HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan temuan hasil penelitian yang terkait dengan aplikasi yang telah kami kembangkan, yang kami beri nama WFraud Alert, yang dirancang untuk memprediksi pesan penipuan WhatsApp menggunakan metode klasifikasi Naïve Bayes.

3.1 Modelling

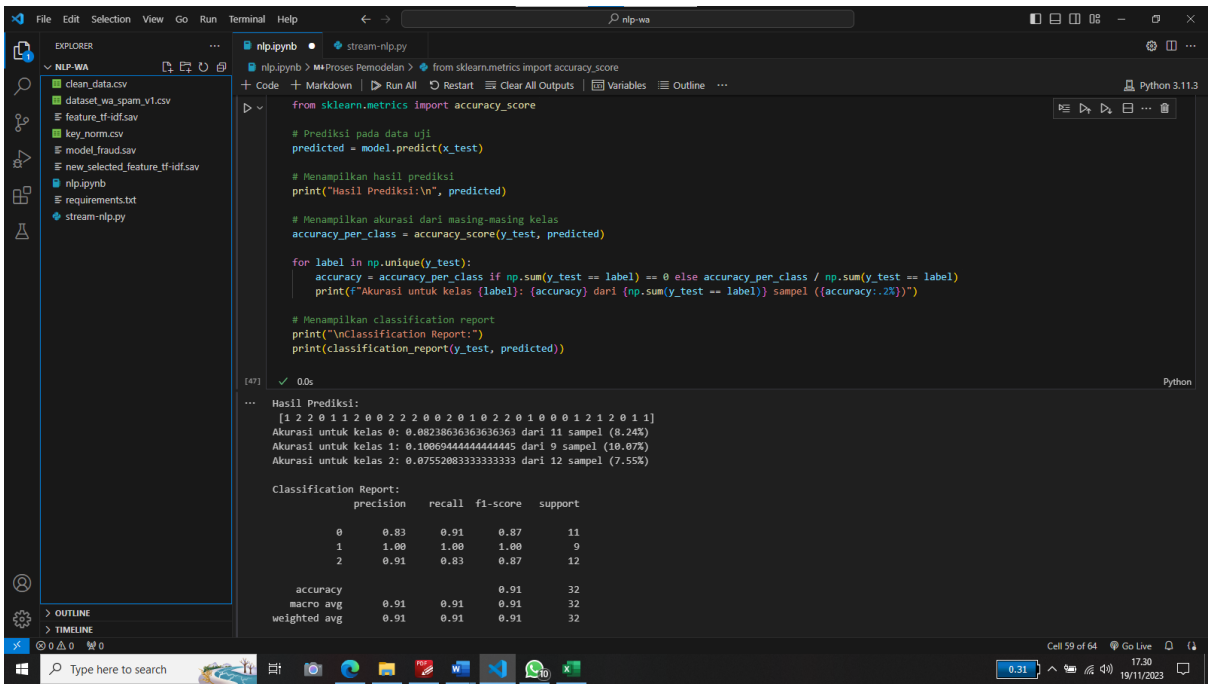
Melalui pelatihan model Naive Bayes, hal ini membuka pintu untuk mengimplementasikan pendekatan yang sangat berguna dalam dunia analisis data. Dengan mendasarkan diri pada teorema Bayes dan mengasumsikan independensi antara variabel-variabel dalam data, model Naive Bayes memungkinkan kita untuk memanfaatkan probabilitas dalam upaya mengklasifikasikan data ke dalam kategori-kategori yang sesuai. Berikut *modelling* untuk aplikasi *WFraud Allert* yang kami buat terdapat pada Gambar 7.



Gambar 7. *Modelling* Aplikasi

3.2 Evaluasi Model

Aplikasi *WFraud Allert* telah membuktikan dirinya sebagai sebuah solusi pendeteksi penipuan pesan WhatsApp yang sangat efektif. Dalam serangkaian pengujian dan evaluasi, performa aplikasi kami telah menunjukkan hasil yang sangat baik. Presisi aplikasi mencapai tingkat 91%, yang mengindikasikan bahwa sebagian besar dari pesan yang diklasifikasikan sebagai penipuan adalah benar-benar penipuan. Selain itu, *Recall* aplikasi mencapai tingkat 91%, yang berarti bahwa sebagian besar pesan penipuan yang ada berhasil dideteksi oleh aplikasi. Selain itu, *F1-score* aplikasi mencapai tingkat 91%, yang merupakan kombinasi yang seimbang antara presisi dan *recall*. Hasil ini menunjukkan bahwa aplikasi kami memiliki kemampuan yang kuat dalam membantu pengguna WhatsApp melindungi diri dari penipuan dan pesan berbahaya. Selain itu tingkat hasil akurasi dari kelas kami juga cukup baik yaitu untuk kelas 0 memiliki akurasi 8,24% dari 11 sampel, kelas 1 memiliki 10,7% dari 9 sampel dan kelas 2 memiliki 7,55% dari 12 sampel. Kami berkomitmen untuk terus meningkatkan dan mengembangkan aplikasi ini untuk memberikan perlindungan yang lebih baik kepada pengguna kami dalam menghadapi beberapa ancaman penipuan pesan WhatsApp yang terus berkembang dan terjadi di masyarakat. Berikut Evaluasi Model untuk aplikasi *WFraud Allert* yang telah kami buat terdapat pada Gambar 8.



Gambar 8. Evaluasi model dan tingkat akurasi setiap kelas

3.3. Confussion Matrix

Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Tabel ini secara rinci memisahkan hasil prediksi model menjadi empat kategori utama, yakni True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).[14] Dengan menyajikan informasi ini, confusion matrix memberikan wawasan mendalam tentang performa model dalam mengklasifikasikan instance-instance pada setiap kelas. Meskipun umumnya diterapkan dalam konteks masalah klasifikasi biner, confusion matrix juga sangat bermanfaat dalam evaluasi model klasifikasi multikelas, membantu mengidentifikasi sejauh mana model dapat mengenali berbagai kelas target dengan akurasi dan memberikan pemahaman yang lebih komprehensif terkait jenis kesalahan prediksi yang mungkin terjadi. Kami menggunakan 3 kelas yaitu kelas pertama pesan normal, kelas kedua pesan judi online dan pesan ketiga yaitu pesan penipuan.

Rumus confusion matrix untuk menghitung accuracy, precision, dan recall keadaan 3 kelas seperti berikut.

$$Accuracy = \frac{Rasio\ True\ Positive\ (TP)}{Total\ jumlah\ data}$$

$$Precision = \frac{Rasio\ True\ Positive\ (TP)}{TP + FP}$$

$$Recall = \frac{Rasio\ True\ Positive\ (TP)}{TP + FN}$$

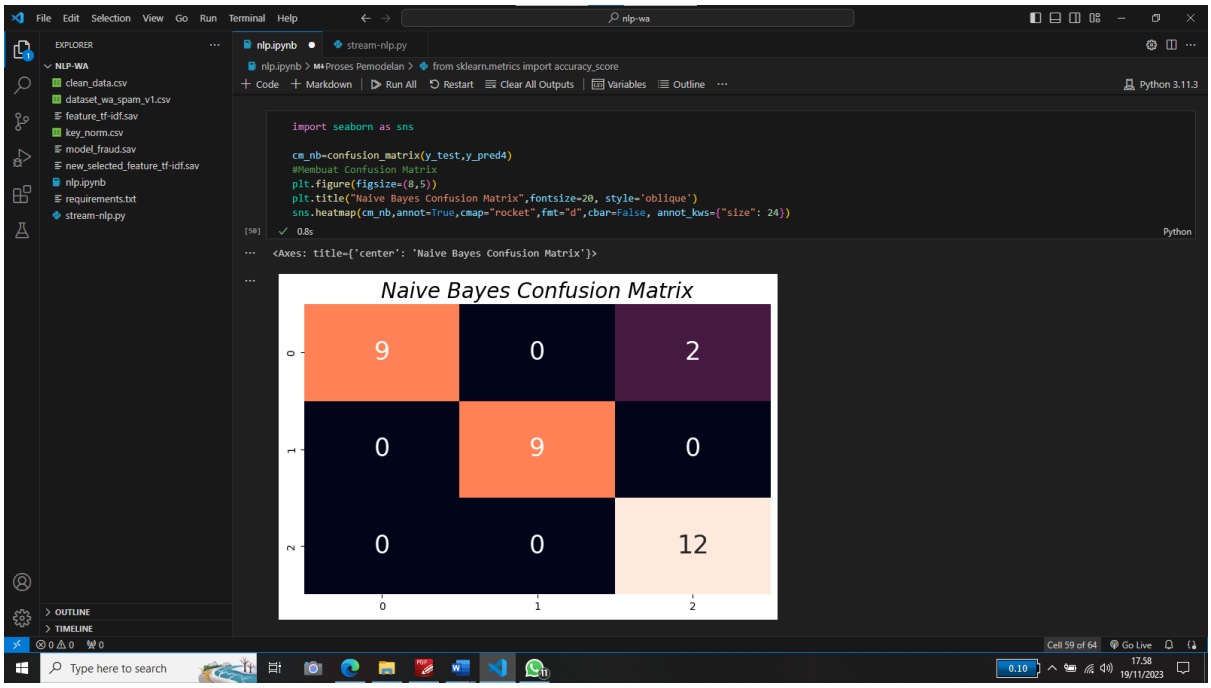
Keterangan:

TP (True Positive) = ini adalah jumlah dari satu kelas true yang bisa di prediksi dengan benar pada kelas true.

False Positive (FP) = ini adalah kondisi dimana kelas true yang prediksinya salah pada kelas false, sedangkan

False Negatif (FN) = ini adalah dimana kondisi pada kelas false yang di prediksi salah pada kelas true.

Berikut data confussion matrix yang kami proses dalam aplikasi pendeteksi penipuan pesan WhatsApp yang kami buat terdapat pada Gambar 8.



Gambar 9. Confussion matrix

4. KESIMPULAN

Dapat disimpulkan, Aplikasi WFraud Alert menunjukkan kualitas unggul dalam mengidentifikasi dan mengklasifikasikan pesan WhatsApp, terutama terkait dengan penipuan dan promosi judi online. Performa aplikasi ini tidak hanya mencerminkan keefektifan algoritma Naive Bayes, tetapi juga menyoroti peran penting langkah-langkah teks preprocessing dalam mempersiapkan data teks untuk analisis lebih lanjut. Penting untuk diakui bahwa aplikasi ini tidak hanya memberikan kontribusi dalam mendeteksi penipuan online, tetapi juga memiliki dampak praktis dalam melindungi pengguna dari ancaman keamanan siber. Keakuratannya dalam mengklasifikasikan pesan WhatsApp menggambarkan sebuah solusi yang efektif dalam menghadapi permasalahan yang muncul dari komunikasi digital, di mana serangan siber melalui pesan penipuan semakin canggih. Penerapan algoritma Naive Bayes bukan hanya sekadar pilihan yang cerdas, tetapi juga merupakan keputusan strategis yang mempertahankan relevansinya dalam era yang terus berkembang. Dalam konteks Natural Language Processing, di mana bahasa terus berubah dan berkembang, Naive Bayes terbukti efektif dalam menangani klasifikasi, bahkan dalam situasi di mana data pelatihan terbatas.

Hasil performa Aplikasi WFraud Alert yang mencapai presisi 91%, recall 91%, dan F1-score 91% menciptakan dasar yang kuat bagi aplikasi ini sebagai alat yang andal. Presisi yang tinggi menunjukkan kemampuan aplikasi untuk meminimalkan jumlah positif palsu, sedangkan recall yang tinggi menandakan kemampuannya untuk mengidentifikasi sebagian besar kasus positif. F1-score yang cukup tinggi mencerminkan keseimbangan yang baik antara presisi dan recall.

Sebagai tindak lanjut, penelitian ini memberikan pemahaman yang lebih dalam tentang pengembangan aplikasi sejenis, serta potensi integrasi teknik-teknik terbaru dalam pembelajaran mesin, seperti jaringan saraf tiruan dan Word Embeddings. Seiring dengan pertumbuhan kompleksitas ancaman siber, langkah-langkah inovatif seperti ini menjadi kunci untuk menjaga relevansi dan efektivitas solusi keamanan. Dengan kata lain, Aplikasi WFraud Alert bukan hanya mencapai tujuannya dalam menghadapi tantangan penipuan online, tetapi juga membuka pintu untuk pengembangan dan peningkatan lebih lanjut dalam bidang keamanan siber dan Natural Language Processing.

UCAPAN TERIMAKASIH

Kami ingin mengungkapkan rasa terima kasih yang tulus kepada sejumlah pihak yang telah memberikan dukungan yang tak ternilai dalam proses penelitian yang penulis lakukan. Pertama-tama, ucapan terima kasih kepada orang tua kami, yang telah memberikan dukungan moral dan motivasi yang konsisten sepanjang perjalanan ini. Selanjutnya, kami juga ingin menyampaikan penghargaan kepada Ketua Prodi Teknik Elektro di Universitas Islam Nusantara, yang telah memberikan bimbingan dan arahan yang berharga dalam pengembangan penelitian ini. Tidak kalah penting, terima kasih kepada teman-teman yang telah berperan sebagai pilar dukungan

sosial dan akademis, memberikan wawasan, masukan, dan semangat yang membantu kami meraih hasil yang bermakna dalam penelitian ini. Semua dukungan ini telah menjadi fondasi yang kuat bagi kemajuan dan keberhasilan penelitian kami, dan kami sangat bersyukur atas kontribusi mereka.

REFERENCES

[1] Veronika, "Perkembangan Teknologi Informasi dan Komunikasi: Sejarah, Dampak, Tantangan, dan Peluang. ," Kompasiana.

[2] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *TEKNO KOMPAK*, vol. 15, no. 1, 2019.

[3] Lenny, "Kominfo Catatkan 1.730 Kasus Penipuan Online, Kerugian Ratusan Triliun," Kata Data.

[4] LERAVIO, "Naive Bayes: Pengertian, Kelebihan, dan Implementasinya," ADMIN LERAVIO.

[5] R. Rachman, R. N. Handayani, and I. Artikel, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[6] R. S. Stmik and N. Mandiri, "Komparasi Algoritma Support Vector Machine, Naïve Bayes Dan C4.5 Untuk Klasifikasi SMS," *IJCIT (Indonesian Journal on Computer and Information Technology*, vol. 2, no. 2, 2017.

[7] D. Delvia Arifin and Ma. Bijaksana, "SMS Filtering Menggunakan Naive Bayes Classifier dan FP-Growth Algorithm Frequent Itemset," *e-Proceeding of Engineering*, 2019, [Online]. Available: http://www.ranks.nl/stopwords.

[8] R. Dwiyansaputra, G. Satya Nugraha, F. Bimantoro, and A. Aranta, "DETEKSI SMS SPAM BERBAHASA INDONESIA MENGGUNAKAN TF-IDF DAN STOCHASTIC GRADIENT DESCENT CLASSIFIER (Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent Classifier)," *JTIKA*, 2021, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/

[9] L. D. Utami, L. Yusuf, and D. Nurlaela, "Komparasi Algoritma Naïve Bayes dan Support Vectors Machine pada Analisis Sentimen SMS HAM dan SPAM," *Jurnal Informatika dan Teknologi*, vol. 4, no. 2, 2021, doi: 10.29408/jit.v4i2.3665.

[10] U. Banten Jaya, J. Syeh Nawawi Albantani, and S. -Banten, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI SMS SPAM BERBAHASA INDONESIA," *SAINTEK | Jurnal Sains & Teknologi* , 2019.

[11] H. Herwanto, N. L. Chusna, and M. S. Arif, "Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, p. 1316, Oct. 2021, doi: 10.30865/mib.v5i4.3119.

[12] Zia Ayu Nuansa Gumilang, "IMPLEMENTASI NAÏVE BAYES CLASSIFIER DAN ASOSIASI UNTUK ANALISIS SENTIMEN DATA ULASAN APLIKASI E-COMMERCE SHOPEE PADA SITUS GOOGLE PLAY," YOGYAKARTA, 2018.

[13] M. Ibrahim, E. Bu, and I. Lubis, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Naive Bayes Classifier Untuk Mendeteksi Tingkat Krediblitas Hoax News/ Fake News Pada Sosial Media Di Indonesia Berbasis Android (Studi Kasus : Kantor Tribun Medan)," *Media Online)*, vol. 1, no. 1, 2020, [Online]. Available: https://djournals.com/resolusi

[14] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.