

## Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Pada Play Store Menggunakan Metode Naïve Bayes

Ahmad Komarudin<sup>1</sup>, Atiqah Meutia Hilda<sup>2\*</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka  
Jl. Tanah Merdeka, No. 6, Kampung Rambutan, Ciracas, Jakarta Timur, Indonesia

e-mail: <sup>1</sup>ahmadkomarudin1205@gmail.com, <sup>2</sup>atiqahmeutiahilda@uhamka.ac.id

(\* ) Corresponding Author

Artikel Info : Diterima : 17-11-2023 | Direvisi : 13-01-2024 | Disetujui : 16-01-2024

**Abstrak** - Salah satu inovasi baru pemerintah yaitu diadakannya Identitas Kependudukan Digital yang mana hal ini untuk membawa transformasi dalam cara masyarakat membuktikan identitasnya sendiri, membuatnya lebih mudah, aman, dan efisien dalam berbagai konteks kehidupan sehari-hari. Diketahui pada 20 Oktober 2023 aplikasi Identitas Kependudukan Digital hanya memperoleh 3,3 bintang dari 32,7 ribu ulasan, hal ini menunjukkan bahwa aplikasi masih belum mencapai tingkat kepuasan yang maksimal dalam melayani kebutuhan masyarakat. Oleh karena itu diperlukan analisis sentimen ulasan pengguna untuk mendapatkan wawasan lebih mendalam tentang bagaimana masyarakat merespons aplikasi ini, sehingga penelitian ini bertujuan untuk mendapatkan sebuah data sentimen dari aplikasi Identitas Kependudukan Digital pada *Play Store* berupa ulasan-ulasan dari pengguna serta hasil dari analisis dapat dijadikan sebagai bahan evaluasi bagi pengembang aplikasi. Pengambilan data dilakukan menggunakan metode *scrapping* melalui *Google Colab* sebanyak 1000 ulasan yang hasilnya akan dilabeli negatif dan positif. Kemudian data akan dibersihkan dan disederhanakan melalui *preprocessing*, dan akan di klasifikasikan dengan algoritma *Naïve Bayes* dengan data uji 90% dan data latih 10% per-iterasinya. Hasil klasifikasi selanjutnya dilakukan perhitungan *performance* dengan *confusion matrix* dan memperoleh nilai *accuracy* sebesar 85,06%, *precision* sebesar 80,31%, dan *recall* sebesar 92,89%.

Kata Kunci : Identitas Kependudukan Digital, Analisis sentimen, *Naïve bayes*

**Abstract** - One of the government's new innovations is the implementation of Digital Population Identity which is to bring transformation in the way people prove their own identity, making it easier, safer, and more efficient in various contexts of everyday life. It is known that on October 20, 2023 the Digital Population Identity application only received 3.3 stars from 32.7 thousand reviews, this shows that the application still has not reached the maximum level of satisfaction in serving the needs of the community. Therefore, sentiment analysis of user reviews is needed to gain deeper insight into how people respond to this application, so this research aims to obtain sentiment data from the Digital Population Identity application on the *Play Store* in the form of reviews from users and the results of the analysis can be used as evaluation material for application developers. Data collection is done using the *scrapping* method through *Google Colab* as many as 1000 reviews whose results will be labeled negative and positive. Then the data will be cleaned and simplified through *preprocessing*, and will be classified with the *Naïve Bayes* algorithm with 90% test data and 10% training data for each iteration. The classification results are then calculated *performance* with *confusion matrix* and obtained an *accuracy* value of 85,06%, *precision* of 80,31%, and *recall* of 92,89%.

Keywords : Digital ID, Sentiment Analysis, *Naïve Bayes*

## PENDAHULUAN

Transformasi digital telah menjadi tren global yang mempengaruhi berbagai aspek kehidupan, termasuk pelayanan administrasi publik. Fenomena ini telah menciptakan langkah baru yang mengubah cara pemerintah berinteraksi dengan masyarakat dan bagaimana layanan publik diakses dan dikelola. Di era digital ini, teknologi informasi telah membuka pintu bagi pemerintah untuk memberikan pelayanan yang lebih efisien dan terjangkau dengan memanfaatkan platform digital (Wiranti & Frinaldi, 2023). Ini termasuk pengembangan portal pelayanan



*online*, penggunaan aplikasi *mobile*, dan implementasi identitas digital yang memungkinkan warga untuk mengakses layanan publik dengan lebih cepat dan mudah. Salah satu inovasi penting dalam upaya mempermudah akses warga terhadap pelayanan publik, seperti perizinan, pelayanan kesehatan, dan pendidikan adalah aplikasi Identitas Kependudukan Digital.

Tahun 2024 merupakan puncak dari pemilihan umum (pemilu) presiden baru Negara Kesatuan Republik Indonesia (NKRI) yang penjadwalan tahapan pentingnya sudah dibuat sejak tahun 2022 (Devit & Ridwan, 2023). Pemilu adalah sebuah sarana pelaksanaan dari demokrasi yang menjembatani suara rakyat sebagai kedaulatan dalam memilih seorang pemimpin (Yusrin & Salpina, 2023). Pada 19 Oktober – 25 November 2023 telah diumumkan secara resmi pasangan Capres dan Cawapres 2024, pengambilan suara akan dilakukan pada 14 Februari 2024 (Devit & Ridwan, 2023). Dalam konteks ini, salah satu prosedur pencoblosan yaitu pemilih membawa dan menunjukkan Identitas Kependudukan Digital atau e-KTP saat datang ke tempat pemungutan suara. Namun, ada beberapa hal yang perlu diperhatikan dalam implementasi kebijakan ini.

Salah satu aspek yang perlu diperhitungkan adalah kesiapan infrastruktur dan teknologi yang mendukung penggunaan Identitas Kependudukan Digital (IKD). IKD *mobile* atau aplikasi Identitas Kependudukan Digital adalah sebuah aplikasi yang bertujuan untuk memberikan solusi digital dalam mengelola dan mengakses informasi identitas kependudukan. Aplikasi ini dirancang untuk membantu warga negara dalam berbagai aspek yang terkait dengan identitas mereka, seperti dokumen identifikasi pribadi, data demografis, dan informasi kependudukan lainnya. Aplikasi Identitas Kependudukan Digital umumnya memiliki fitur-fitur seperti pengelolaan kartu identitas digital, akses cepat ke data kependudukan, kemampuan untuk melakukan verifikasi identitas secara *online*, serta integrasi dengan layanan publik dan sektor swasta. Aplikasi ini dapat diunduh masyarakat secara gratis melalui *Play Store* yaitu sebuah *marketplace* aplikasi dimana pengguna dapat mengunduh aplikasi sekaligus dapat menyampaikan ulasan dan penilaian terhadap aplikasi Identitas Kependudukan Digital.

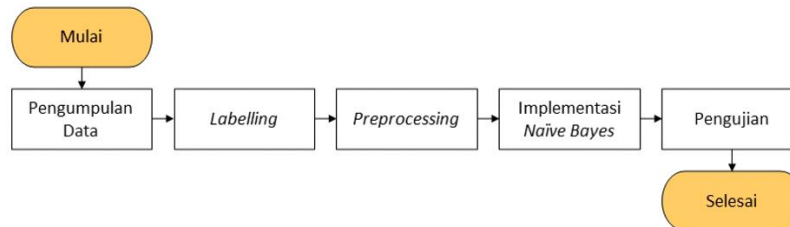
Adapun ulasan pengguna dapat mencerminkan persepsi dan pengalaman pengguna mengenai aplikasi, serta menjadi salah satu sumber informasi yang paling terpercaya bagi calon pengguna lainnya yang ingin mengevaluasi kegunaan dan kualitas aplikasi (Muhammadin & Sobari, 2021). Namun berdasarkan rating hasil ulasan pengguna diketahui bahwa rating aplikasi Identitas Kependudukan Digital pada 20 Oktober 2023 hanya memperoleh rating 3,3 bintang dari 5 bintang dan 32,7 ribu ulasan. Hal ini menunjukkan bahwa aplikasi Identitas Kependudukan Digital masih belum mencapai tingkat kepuasan yang maksimal dalam melayani kebutuhan masyarakat. Oleh karena itu diperlukan analisis sentimen ulasan pengguna untuk mendapatkan wawasan lebih mendalam tentang bagaimana masyarakat merespons aplikasi ini. Hasil analisis sentimen ulasan pengguna juga dapat membantu pengembang dalam mengidentifikasi area-area yang memerlukan perbaikan atau peningkatan, baik dari segi teknis maupun pengalaman pengguna, sehingga aplikasi Identitas Kependudukan Digital dapat terus berkembang dan memenuhi kebutuhan warga negara secara efektif.

Penelitian terdahulu oleh (Fikri *et al.*, 2020) dengan membandingkan *Naïve Bayes* dan *Support Vector Machine* pada *analisis sentimen Twitter* yang menggunakan 2654 *tweets* diperoleh akurasi 73,63% untuk *Naïve Bayes*, 70,20% untuk *Support Vector Machine* (SVM). Klasifikasi SVM dipilih karena kemampuannya dalam menemukan *hyperplane* terbaik sebagai pemisah antara dua kelas data. *Hyperplane* ini juga dikenal sebagai “*support vector*” yang dapat meminimalkan jarak antara data terdekat dari setiap kelas ke *hyperplane* tersebut (Yousef & ALali, 2022). Sedangkan *Naïve Bayes Classifier* (NBC) adalah salah satu metode klasifikasi probabilitas yang sederhana namun efektif dalam mengklasifikasikan data. Metode ini didasarkan pada teori probabilitas dan prinsip *Bayes*, dan mengasumsikan bahwa setiap atribut dalam data independen satu sama lain (*naive assumption*). Metode ini diasumsikan bahwa setiap fitur (atribut) pada data adalah independen satu sama lain, meskipun sebenarnya tidak selalu terjadi dalam dunia nyata. Meskipun demikian, NBC memiliki keuntungan utama dalam penggunaannya yang sederhana dan cepat untuk dilatih pada data yang sangat besar. Pada penelitian lain oleh (Normah *et al.*, 2022) terkait analisis sentimen terkait perkembangan *Vtube* dengan *Support Vector Machine* dan menerapkan *SMOTE* sebagai penyeimbangan data yang mampu meningkatkan nilai keakurasian dari sebuah model sehingga diperoleh akurasi 88,18% dari 321 data.

Berdasarkan hasil kedua penelitian sebelumnya, dapat disimpulkan bahwa metode NBC memiliki kinerja yang lebih baik dalam melakukan klasifikasi data ulasan dan metode *SMOTE* merupakan penyeimbang data untuk meningkatkan performance dari model. Maka dari itu penelitian ini bertujuan untuk menggali lebih dalam bagaimana pengguna merespon aplikasi IKD dengan tingkat akurasi yang baik dengan menerapkan *SMOTE* pada pengklasifikasian dengan model *Naïve Bayes*. Selain itu, metode ini NBC juga digunakan untuk memahami faktor-faktor yang mempengaruhi sentimen pengguna dengan melakukan analisis probabilitas terhadap kata-kata kunci dan konteks ulasan. Dengan demikian, peneliti dapat mengidentifikasi pola-pola yang mungkin mempengaruhi pandangan positif atau negatif terhadap aplikasi Identitas Kependudukan Digital.

## METODE PENELITIAN

Pada penelitian ini pengolahan data dilakukan dengan beberapa proses yang harus dilalui yaitu dapat dilihat pada Gambar 1.



Sumber: Penelitian (2023)

Gambar 1. Tahapan penelitian

### 1. Pengumpulan Data

Peneliti mengumpulkan data ulasan aplikasi Identitas Kependudukan Digital pada *Play Store* dengan cara *scrapping data*. *Scrapping* adalah salah satu teknik untuk mengumpulkan data pada sebuah *website* secara otomatis, lalu data dapat disimpan dalam format tertentu (Sulahmi *et al.*, 2023). Perolehan dari *scrapping* ini berupa data yang belum memiliki label (positif dan negatif).

### 2. Labelling

Proses melabelkan atau menentukan kelas (positif, negatif) dari ulasan pada dataset. Memberikan label pada dataset memiliki tujuan untuk menciptakan representasi dari objek data yang terdapat dalam ulasan, serta membantu dalam memahami cara mengenali sentimen data (Setya Ananto & Hasan, 2023). Pada *labelling* dapat dilakukan secara otomatis dan manual, data dari *Play Store* dapat dilakukan dengan melihat angka rating yang terkandung dari setiap ulasan biasanya rating 1-2 untuk negatif dan 3-5 untuk positif. Namun terkadang para pengguna memberikan antara ulasan dengan rating ulasan tidak sesuai, sehingga cara manual merupakan cara yang bisa tervalidasinya data pada tahap *labelling*.

### 3. Preprocessing

Fase pertama adalah *preprocessing*, data yang tidak terstruktur akan ditransformasikan menjadi data yang lebih terstruktur sehingga menghasilkan data yang siap dilakukan analisis lebih lanjut (Yousef & ALali, 2022). Ada lima tahapan *preprocessing* pada penelitian ini meliputi:

#### a. Cleansing

Proses membersihkan dokumen dari kata-kata yang tidak diperlukan atau disebut juga sebagai tahapan *cleaning*. Tujuannya agar dokumen yang akan diolah menjadi lebih bersih dan relevan. Salah satu cara untuk membersihkan dokumen adalah dengan menghapus simbol dan tanda baca yang tidak berkontribusi dalam analisis teks (Villavicencio *et al.*, 2021).

#### b. Transform Case

*Transform case* merupakan rangkaian daripada analisis teks, dengan tujuan mengubah huruf kapital pada kalimat-kalimat dalam sebuah dokumen menjadi huruf kecil (Md Saad *et al.*, 2023). Sehingga semua data yang diperoleh menjadi sama yaitu berhuruf kecil.

#### c. Tokenize

Tokenisasi adalah proses mengubah dokumen teks menjadi serangkaian token atau unit-token, seperti kata-kata atau frasa-frasa, yang lebih mudah diolah oleh komputer (Jaya Hidayat *et al.*, 2021).

#### d. Filter Stopword

*Filter stopwords* merupakan rangkaian dari analisis teks yang berfungsi untuk mengeliminasi kata-kata yang dianggap kurang memberikan kontribusi signifikan terhadap makna dari isi ulasan dalam dokumen. Kata yang dieliminasi disebut kata penghubung atau *stopword* (Halim & Andri Safuwani, 2023). Keluaran dari fase ini yaitu data akan semakin ringkas karena penghilangan semua kata penghubung.

#### e. Filter Token by Length

*Filter token by length* yaitu metode pada analisis teks yang digunakan sebagai penghapus kata-kata dengan jumlah minimal maksimal huruf tertentu (Yin *et al.*, 2022). Data yang telah melalui fase ini kemudian akan dilakukan ketahap implementasi model.

### 4. Implementasi

Pada tahapan ini dilakukan proses pemodelan klasifikasi teks pada data sudah dapat diolah. Sebelum pemodelan data yang telah di *preprocessing* akan melalui tahapan pembobotan TF-IDF terlebih dahulu yaitu sebuah metode pembobotan yang digunakan dalam analisis teks untuk mengevaluasi seberapa penting sebuah kata atau frasa dalam sebuah dokumen (Dwi Antonio *et al.*, 2022), setelah itu dilakukan penyeimbangan data dengan *SMOTE Upsampling*. Kemudian dilakukan pembagian data menjadi 90% data uji dan 10% data latih per-iterasi.

Lalu dilakukan pemodelan menggunakan algoritma *Naïve Bayes*.

### 5. Pengujian

Kemudian pengujian *performance* dari sebuah model yang telah dibuat memakai *confusion matrix*. Perhitungan yang dikenal sebagai "*confusion matrix*" membandingkan dataset dengan hasil klasifikasi yang sesuai dengan jumlah data sebenarnya. Hasil yang diperoleh dari tahap ini yaitu akurasi, *precision*, dan *recall* dari model dengan satuan persen (%) (Normawati & Prayogi, 2021).

## HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data

Peneliti melakukan pengumpulan dataset ulasan aplikasi Identitas Kependudukan Digital pada *Play Store* menggunakan *scrapping data* dengan *Google Colab*. Gambar 2 adalah Ilustrasi dari pengumpulan data pada penelitian ini.



Sumber: Penelitian (2023)

Gambar 2. Ilustrasi pengumpulan data

Dapat dilihat bahwa proses penarikan dimulai dengan pengambilan ulasan dari *Play Store* dengan kode aplikasi "*gov.dukcapil.mobile\_id*". Proses ini memilih ulasan terbaru (*NEWEST*) sebanyak 1000 yang diekspor ke dalam bentuk CSV. Adapun data yang diambil merupakan data terbaru pada 20 Oktober 2023.

1	userName	score	at	content
2	Ferawati Tanjung	1	10/20/2023 11:46	Sumpah yah bingung banget sama nih aplikasi. Disuruh bikin ktp barcode tapi aplikasinya kaya
3	Ujat Sutisna	3	10/20/2023 10:43	App bagus bisa digunakan tapi tidak lengkap menu dokumen cuma ada KTP ga ada kk
4	Rani Purnama	4	10/20/2023 9:39	Aplikasinya keren bgt
5	Yandri Prima Dani	1	10/20/2023 7:23	Dah 2 bulan tuh dokumen belum terintegrasi Mulu heran dah
6	Wandi Vivo	5	10/20/2023 6:16	Dukcapil metro keren
7	Yadi Tarmidi	1	10/20/2023 5:45	Tertap saja harus datang ke kantor disdukcapil, untuk scan QR untuk aktivasi akun. Kalu untuk
8	Sunardi	1	10/20/2023 3:53	Masih dipake gak si aplikasinya? Buat pengajuan gak pernah di proses,
9	Yeni Yeni tamala	5	10/20/2023 3:10	Dokumen kaga bisa terbuka tolong perbaiki
10	Baron Stretfire	1	10/20/2023 1:23	sangat disayangkan apk sgt bruk skli.. buka apk kok stngh hari

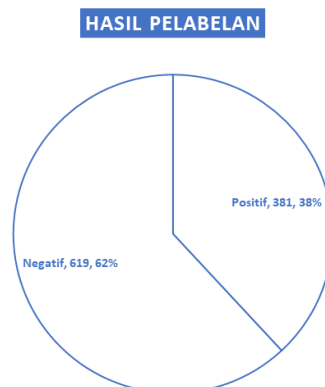
Sumber: Penelitian (2023)

Gambar 3. Hasil pengumpulan data

Gambar 3 menunjukkan hasil sebagian data ulasan yang telah dikumpulkan berkaitan dengan pendapat, saran, dan kritik masyarakat mengenai aplikasi IKD.

### 2. Labelling

Proses pelabelan data pada data ulasan secara manual oleh peneliti dengan menggunakan bantuan aplikasi *Microsoft Excel*. Gambar 4 merupakan visualisasi dari hasil pelabelan.



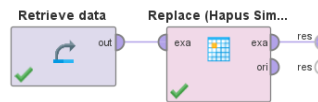
Sumber: Penelitian (2023)

Gambar 4. Visualisasi hasil pelabelan

Setelah dilakukan proses pelabelan diperoleh hasil keseluruhan total data sebanyak 1000 data dengan jumlah kategori ulasan terbanyak adalah data ulasan berkategori negatif sebanyak 619 data, sementara jumlah data ulasan berkategori positif sebanyak 381 data.

### 3. Preprocessing

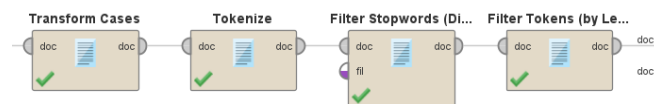
Selanjutnya membersihkan dokumen dari kata-kata yang tidak diperlukan atau disebut juga sebagai tahapan *cleaning*. Proses *cleansing* pada penelitian ini menggunakan *tool RapidMiner*. Pada tahap ini akan dihapus simbol pada ulasan yang tidak berkontribusi dalam analisis teks.



Sumber: Penelitian (2023)

Gambar 5. *Cleansing*

Operator yang digunakan adalah operator *retrieve* atau data ulasan aplikasi dihubungkan dengan operator *replace*. operator *replace* digunakan untuk menggantikan nilai-nilai tertentu dalam data dengan nilai baru. Pada proses *cleansing* ini terdapat 1 operator *replace* yang mempunyai *regular expression* berbeda yaitu *replace* hapus simbol.



Sumber: Penelitian (2023)

Gambar 6. Proses *preprocessing* menggunakan *tool RapidMiner*

Gambar 6 menggambarkan metode prapemrosesan lanjutan yang dimulai dengan *transform case* sampai *filter token by length*. Pada prosedur *filter stopwords*, peneliti melakukan pencarian mencari kamus *stopword* bahasa Indonesia yang terdiri dari kata penghubung. Peneliti menemukan kamus tersebut pada internet, yaitu di <https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist>. Setelah itu kamus tersebut dimasukkan pada parameter *filter stopwords*. Pada *filter token by length* parameter yang digunakan yaitu *min chars* 4 dan *max chars* 25. Hasil yang diperoleh dari *preprocessing* tersebut sebagai berikut:

Tabel 1. Hasil dari preprocessing

Tahapan <i>preprocessing</i>	Hasil
Salah satu data ulasan	Gak bisa buka Document.. kl blm bagus aplikasi.nya d ada kan lg Cetak KTP Elektriikk. Semua urusan keganggu gara <sup>2</sup> ini
<i>Cleansing</i>	Gak bisa buka Document kl blm bagus aplikasinya d ada kan lg Cetak KTP Elektriikk Semua urusan keganggu gara ini
<i>Transform Case</i>	gak bisa buka document kl blm bagus aplikasinya d ada kan lg cetak ktp elektriikk semua urusan keganggu gara ini
<i>Tokenize</i>	gak, bisa, buka, document, kl, blm, bagus, aplikasinya, d, ada, kan, lg, cetak, ktp, elektriikk, semua, urusan, keganggu, gara, ini
<i>Filter Stopword</i>	gak, buka, document, kl, blm, bagus, aplikasinya, d, lg, cetak, ktp, elektriikk, urusan, keganggu, gara
<i>Filter Token by Length</i>	buka, document, bagus, aplikasinya, cetak, elektriikk, urusan, keganggu, gara

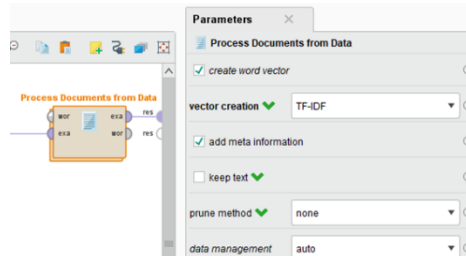
Sumber: Penelitian (2023)



#### 4. Implementasi

##### a. Pembobotan TF-IDF

Tahap awal implementasi adalah melakukan pembobotan TF-IDF untuk menghitung bobot kata dalam dokumen. Pada platform *RapidMiner*, salah satu metode untuk melakukan pembobotan TF-IDF adalah dengan memanfaatkan operator *Process Documents from Data*. Dalam operator tersebut peneliti menempatkan beberapa tahapan *preprocessing*. Berikut adalah prosedur pembobotan TF-IDF dengan menggunakan *RapidMiner*:



Sumber: Penelitian (2023)

Gambar 7. Proses TF-IDP

Gambar 8 adalah hasil dari proses pembobotan TF-IDF yang menunjukkan nilai bobotan kata pada dataset.

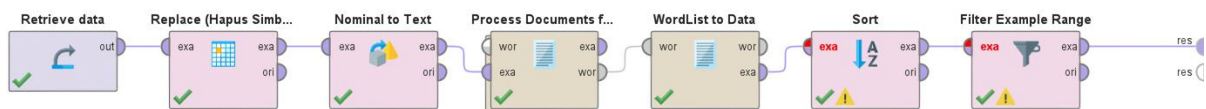
text	abalabal	abdet	abis	acount	adaasal
bermanfaat praktis	0	0	0	0	0
terimakasih disdukcapil muba	0	0	0	0	0
terkesan ribetpadahal sampe menit digital.	0	0	0	0,280	0
tolong upgrade verifikasi scan kode masy.	0	0	0	0	0
sebener kasih refiew rating jelek dibaca t..	0	0	0	0	0
perbaiki aplikasi kurangnya	0	0	0	0	0
gimana ceritanya coba digital buka	0	0	0	0	0
mantap	0	0	0	0	0
pemerintah kerja aplikasi bermutu kesala..	0	0	0	0	0
aplikasi kntrl boomer melek teknologi ribet	0	0	0	0	0

Sumber: Penelitian (2023)

Gambar 8. Hasil TF-IDP

##### b. Visualisasi Data

Kemudian operator *Process Documents from Data* dihubungkan dengan operator *WordList to Data*. Operator tersebut digunakan untuk menentukan bobot nilai serta banyaknya frekuensi kemunculan dari setiap kata pada dataset yang telah melakukan *preprocessing*.



Sumber: Penelitian (2023)

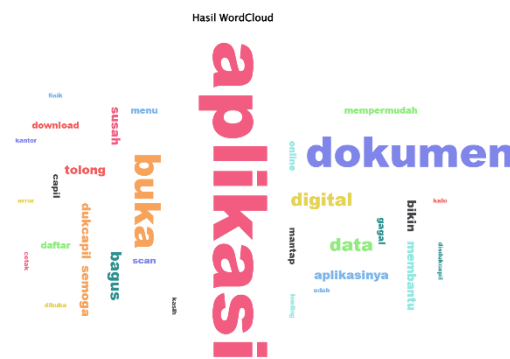
Gambar 9. Visualisasi data dengan wordcloud

Gambar 9 menunjukkan proses visualisasi data dengan *wordcloud*. Adapun operator yang digunakan yaitu operator data ulasan yang telah dilakukan *preprocessing* kemudian dihubungkan dengan operator *WordList to Data* yang digunakan untuk mengkonversi daftar kata menjadi format data yang dapat dipergunakan untuk analisis lanjutan. Operator ini perlu untuk dihubungkan lagi dengan operator *Sort* dan *Filter Example Range* untuk mengurutkan hasil berdasarkan jumlah kata terbanyak yang muncul dalam dataset.

Row No.	word	in documents	total	in class (Neg...	in class (Pos...
1	aplikasi	210	264	224	40
2	dokumen	140	152	139	13
3	buka	118	136	119	17
4	digital	62	67	46	21
5	data	56	64	52	12
6	bagus	57	58	20	38
7	aplikasinya	47	51	34	17
8	dukcapil	49	51	33	18
9	tolong	48	49	43	6
10	susah	45	48	47	1

Sumber: Penelitian (2023)

Gambar 10. Hasil urutan jumlah kemunculan kata



Sumber: Penelitian (2023)

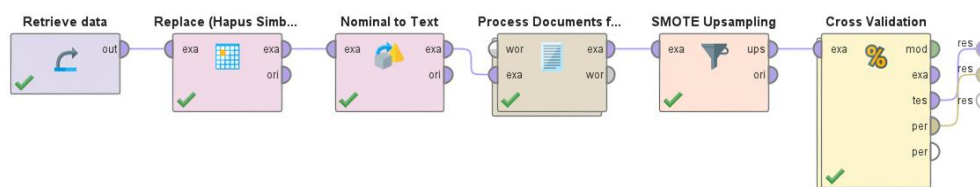
Gambar 11. Hasil wordcloud

Hasil ini menunjukkan bahwa semakin besar total kata pada dokumen maka akan semakin besar juga kata yang ditampilkan. Dari Gambar 11 diketahui kata yang sering muncul pada data ulasan aplikasi IKD yaitu kata “aplikasi”, “dokumen”, “buka”, “digital”, “data”, “bagus”, “aplikasinya”, “dukcapil”, “tolong”, “susah”.

Kata-kata yang muncul tersebut dapat disimpulkan bahwa sentimen masyarakat terhadap aplikasi Identitas Kependudukan Digital yaitu negatif. Karena berdasarkan Gambar 10 menunjukkan sebagian kata-kata yang muncul kebanyakan berasal dari kelas negatif. Beberapa kata kunci yang sering muncul seperti “susah”, “tolong”, berasal dari kelas sentimen negatif, menunjukkan adanya masalah atau ketidakpuasan dalam penggunaan aplikasi.

### c. Klasifikasi *Naïve Bayes*

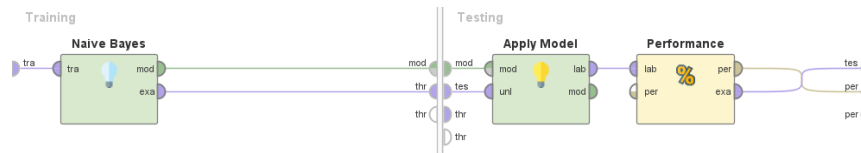
*Naïve Bayes* digunakan untuk pemodelan klasifikasi teks selama tahap implementasi. Dimana peneliti menggunakan operator *Cross Validation* dengan *folds cross validation* sebanyak 10, sehingga data uji sebesar 90% dan data latih 10% per-iterasinya. Sebelum melakukan klasifikasi diperlukan penyeimbangan data dengan operator *SMOTE Upsampling* untuk memperoleh akurasi yang lebih baik (Astuti & Lenti, 2021).



Sumber: Penelitian (2023)

Gambar 12. Pemodelan dengan *RapidMiner*

Dalam operator *Cross Validation* dilakukan implementasi algoritma dengan memasukan operator *Naïve Bayes* yang disambungkan dengan operator *Apply Model* dan operator *Performance*.



Sumber: Penelitian (2023)

Gambar 13. Operator dalam *Cross Validation*

Operator *Naive Bayes* sebagai algoritma klasifikasi. Lalu dihubungkan dengan operator *Apply Model* untuk mengaplikasikan model tersebut pada data *testing* untuk menghasilkan prediksi sentimen. Dan terakhir dihubungkan dengan operator *Performance* untuk memberikan metrik evaluasi kinerja model pada penelitian ini yaitu akurasi, *precision* dan *recall*.

**d. Confusion Matrix**

*Performance* dari model yang dibuat, akan dilakukan pengujian dengan metode *confusion matrix* dengan tujuan mengukur seberapa keakuratan dari sebuah model pada klasifikasi sebuah dataset. Hasil dari *confusion matrix* sebagai berikut:

	true Negatif	true Positif
pred. Negatif	478	44
pred. Positif	141	575

Sumber: Penelitian (2023)

Gambar 14. *Confusion Matrix*

Berdasarkan Gambar 14 diketahui *True Negative* (TN) sebesar 478, *False Positive* (FP) sebesar 141, *False Negative* (FN) sebesar 44, dan *True Positive* (TP) sebesar 575. Dalam penelitian (Fauziningrum & Suryaningsih, 2021) untuk menghitung *accuracy*, *precision*, dan *recall* menggunakan *confusion matrix* dilakukan rumus berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{575 + 478}{575 + 478 + 141 + 44} \times 100\% = 85.06\%$$

$$Precision = \frac{TP}{FP + TP} \times 100\% = \frac{575}{141 + 575} \times 100\% = 80.31\%$$

$$Recall = \frac{TP}{FN + TP} \times 100\% = \frac{575}{44 + 575} \times 100\% = 92.89\%$$

Hasil evaluasi dari analisis sentimen tersebut memperoleh nilai *accuracy* sebesar 85.06% menunjukkan bahwa model mampu mengklasifikasikan data secara keseluruhan dengan tingkat baik, *precision* sebesar 80.31% menunjukkan model mampu mengidentifikasi hasil positif dengan benar, dan *recall* sebesar 92.89% menandakan bahwa model menemukan sebagian besar *instance* positif yang sebenarnya. Dengan *recall* yang tinggi, model tersebut cenderung berhasil mengidentifikasi sebagian besar kasus positif.

**KESIMPULAN**

Dari hasil analisis sentimen diperoleh informasi bahwa terdapat lebih banyak sentimen negatif daripada sentimen positif yang menunjukkan bahwa ulasan pengguna mengenai aplikasi Identitas Kependudukan Digital pada *Play Store* cenderung kurang baik. Hasil tersebut menyatakan bahwa pemerintah masih belum siap dalam menyediakan layanan identitas kependudukan digital dengan baik. Kemudian hasil dari klasifikasi data dengan menerapkan *SMOTE Upsampling* sangat mempengaruhi dari peningkatan *performance Naive Bayes* dengan diperoleh nilai *accuracy* sebesar 85.06%, *precision* sebesar 80.31%, dan *recall* sebesar 92.89%, sehingga dapat disimpulkan bahwa model klasifikasi yang diperoleh dapat mengklasifikasikan dengan baik terhadap ulasan masyarakat mengenai aplikasi Identitas Kependudukan Digital pada *Play Store*.

Saran untuk peneliti berikutnya, pada tahap *preprocessing* perlunya penambahan metode *stemming* untuk menghasilkan tingkat akurasi yang lebih baik lagi. Serta perlunya membandingkan beberapa metode klasifikasi untuk mengetahui metode mana yang terbaik.



**REFERENSI**

- Astuti, F. D., & Lenti, F. N. (2021). Implementasi SMOTE untuk Mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN. *Jurnal JUPITER*, 13(1), 89–98. <https://jurnal.polsri.ac.id/index.php/jupiter/article/view/3251>
- Devit, M., & Ridwan, N. (2023). Dampak Framing Hasil Survey Pemilihan Presiden 2024 Mempengaruhi Persepsi Publik. *Jurnal Kajian Ilmiah*, 23(3), 233–242. <https://doi.org/10.31599/jki.v23i3.2805>
- Dwi Antonio, V., Efendi, S., & Mawengkang, H. (2022). Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent. *Int. J. Nonlinear Anal. Appl*, 13(1), 2008–6822. <https://doi.org/10.22075/IJNAA.2021.5735>
- Fauziningrum, E., & Suryaningsih, E. I. (2021). Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni). *Prosiding Kemaritiman 2021*. <http://repository.unimar-amni.ac.id/id/eprint/3837>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Halim, A., & Andri Safuwani. (2023). Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization. *Jurnal Informatika Teknologi Dan Sains*, 5(1), 170–178. <https://doi.org/10.51401/jinteks.v5i1.2229>
- Jaya Hidayat, T. H., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2021). Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. *Procedia Computer Science*, 197(2022), 660–667. <https://doi.org/10.1016/j.procs.2021.12.187>
- Md Saad, N. H., Zainul Abidi, A. S., Yaacod, Z., Mohd Ali, M. M., & Kun, Z. (2023). Twitter Sentiment Analysis on Meta: A Lexicon-Based Analysis using Rapidminer. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(6), 259–270. <https://doi.org/10.17762/ijritcc.v11i6.7561>
- Muhammadin, A., & Sobari, I. A. (2021). Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 2(2), 85–91. <https://doi.org/10.31294/reputasi.v2i2.785>
- Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <https://doi.org/10.30645/j-sakti.v5i2.369>
- Setya Ananto, F., & Hasan, F. N. (2023). Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store. *Jurnal ICT: Information Communication & Technology*, 23(1), 75–80. <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>
- Sulahmi, Wardani, S., & Fairuzabadi, M. (2023). Web Scraping Untuk Pengambilan Data Transaksi Penjualan Pada Markeplace Toko Online Jogjabatik. *JIIFKOM (Jurnal Ilmiah Informatika Dan Komputer)*, 2(2), 14–17. <https://doi.org/10.51901/jiifkom.v2i2.254>
- Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes. *Information*, 12(5), 204. <https://doi.org/10.3390/info12050204>
- Wiranti, N. E., & Frinaldi, A. (2023). Meningkatkan Efisiensi Pelayanan Publik dengan Teknologi di Era Digital. *JIM: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 8(2), 748–754. <https://doi.org/10.24815/jimps.v8i2.24833>
- Yin, J. Y. B., Saad, N. H. M., & Yaacob, Z. (2022). Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee. *TEM Journal*, 11(4), 1508–1519. <https://doi.org/10.18421/TEM114-11>
- Yousef, M., & ALali, A. (2022). Analysis and Evaluation of Two Feature Selection Algorithms in Improving the Performance of the Sentiment Analysis Model of Arabic Tweets. *International Journal of Advanced Computer Science and Applications*, 13(6), 705–711. <https://doi.org/10.14569/IJACSA.2022.0130683>
- Yusrin, Y., & Salpina, S. (2023). Partisipasi Generasi Millennial dalam Mengawasi Tahapan Pemilu 2024. *Journal on Education*, 5(3), 9646–9653. <https://doi.org/10.31004/joe.v5i3.1842>