

ANALISIS SENTIMEN TINGKAT KEPUASAN PENGGUNA PADA SURVEI APLIKASI CEISA DIREKTORAT JENDERAL BEA DAN CUKAI

¹Meta Eri Safitri, S.Kom, MMSI, ²Agnia Dzikri Almadani

¹Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo
mettasafitri.ms@gmail.com

²Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo
agniadzikri32@gmail.com

Abstrak

Direktorat Informasi Kepabeanan dan Cukai merupakan unit IT dibawah Direktur Jenderal Bea dan Cukai yang terus mendukung transformasi DJBC menjadi data driven organization. Dimana data driven organization mendukung paradigma baru yang menggunakan big data dalam menganalisis suatu masalah dan diharapkan hasil analisisnya dapat dijadikan rencana strategis untuk kepentingan organisasi. Setiap tahun di DJBC dilakukan survei layanan kepuasan pengguna aplikasi CEISA. Data survei yang dihasilkan saat ini memiliki output jawaban yang sangat beragam. Sehingga cukup sulit untuk mengklasifikasikan terkait hasil dari survei tersebut. Berdasarkan permasalahan yang telah dijelaskan tersebut, maka perlu adanya analisis sentimen tingkat kepuasan pengguna pada survei kepuasan pengguna aplikasi CEISA DJBC. Metode yang dipakai dalam penelitian adalah Knowledge Discovery in Database (KDD). Adapun tahapan dari metode KDD yaitu Data Selection, Preprocessing, Transformation, Data Mining Interpretation/Evaluasi sehingga dari metode tersebut dapat menghasilkan Knowledge. Algoritma yang digunakan dalam penelitian ini adalah Algoritma Naïve Bayes. Telah dilakukan analisis sentimen tingkat kepuasan pengguna terhadap aplikasi CEISA di Direktorat Jenderal Bea dan Cukai yang terdiri dari sentimen positif dan sentimen negatif. Dari hasil pengujian untuk nilai accuracy 83,97%, untuk nilai precision yaitu 80,99% pada class positif 80,18 % pada class negatif 88,95%, sedangkan nilai recall 90,55 pada class positif dan 77,27% untuk nilai class negatif.

Kata Kunci: Sentimen Analisis, Text Mining, Naïve Bayes, Data Mining, Survei

1. PENDAHULUAN

Direktorat Jenderal Bea dan Cukai merupakan instansi pemerintah yang berada di bawah dan bertanggung jawab kepada Menteri Keuangan dan dipimpin oleh Direktur Jenderal Bea dan Cukai. Direktorat Jenderal Bea dan Cukai mempunyai tugas menyelenggarakan perumusan dan pelaksanaan kebijakan di bidang pengawasan, penegakan hukum, pelayanan dan optimalisasi penerimaan negara di bidang kepabeanan dan cukai sesuai dengan ketentuan peraturan perundang-undangan.

Direktorat Informasi Kepabeanan dan Cukai mengambil peran penting dalam pelaksanaan fungsi pelayanan dan pengawasan di Direktorat Jenderal Bea dan Cukai. Direktorat Informasi Kepabeanan dan Cukai merupakan unit IT dibawah Direktur Jenderal Bea dan Cukai yang terus mendukung transformasi DJBC menjadi data driven organization. Setiap tahun dilakukan survei layanan kepuasan pengguna aplikasi CEISA. Output dari survei tersebut adalah untuk menjadi bahan evaluasi terkait kendala-kendala yang sering muncul terkait

performance dan bug aplikasi. Sehingga dapat dilakukan perbaikan lebih lanjut untuk aplikasi tersebut. Data survei yang dihasilkan saat ini memiliki output jawaban yang sangat beragam. Sehingga cukup sulit untuk mengklasifikasikan terkait hasil dari survei tersebut. Agar data survei dapat dimanfaatkan, perlu adanya metode untuk mengubah data survei menjadi data yang bermanfaat.

Berdasarkan permasalahan yang telah dijelaskan di atas, maka diperlukan suatu metode yang dapat menyelesaikan permasalahan tersebut yaitu text classification. Klasifikasi (Clasification) ialah tugas memilih label kelas yang tepat jika diberikan sebuah input. Salah satu cabang penelitian pada text classification yaitu sentimen analisis yang merupakan metode untuk menganalisis suatu topik tertentu.

2. METODOLOGI

Metode penelitian pada penelitian ini adalah pendekatan deskriptif analitis. Penelitian deskriptif analitis merupakan suatu metode yang berfungsi untuk mendeskripsikan atau memberi

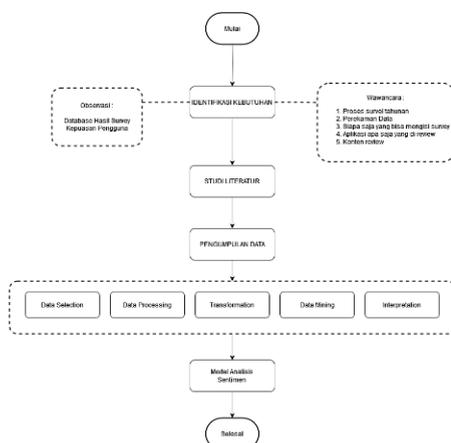
gambaran terhadap objek yang diteliti melalui data sample yang telah terkumpul sebagaimana adanya tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum. Dengan kata lain penelitian deskriptif analitis mengambil masalah atau memusatkan perhatian kepada masalah-masalah sebagaimana adanya saat penelitian dilakukan, hasil penelitian kemudian diolah dan dianalisis untuk diambil kesimpulannya[15].

Langkah-langkah yang perlu dilakukan pada pendekatan deskriptif analitis adalah sebagai berikut :

1. Merumuskan tujuan penelitian
2. Menentukan unit studi dan menghubungkan hal yang akan dikaji berdasarkan sifat-sifat yang dimiliki dan proses yang menjadi landasan penelitian.
3. Menentukan rancangan dalam memilih unit dan teknik pengumpulan data.
4. Mengumpulkan data.
5. Mengorganisasikan informasi, data yang terkumpul, serta melakukan interpretasi dan generalisasi.
6. Menyusun laporan yang diakhiri dengan menyimpulkan hasil penelitian

2.1 Kerangka Pemikiran

Kerangka penelitian penulis lakukan dapat dilihat pada gambar berikut :



Gambar 1 Kerangka Penelitian
Sumber : Penelitian Mandiri

Kerangka pemikiran yang dibuat pada penelitian ini harus memperhatikan data hasil survei kepuasan pengguna pada aplikasi CEISA

secara keseluruhan. Pengumpulan data utama terdapat pada proses wawancara dan observasi

yang dilakukan untuk dapat mengetahui informasi terkait identifikasi kebutuhan untuk membuat *system* tersebut. Setelah dilakukan identifikasi kebutuhan dilakukan *study literature* untuk mengetahui sumber dan referensi untuk membuat analisis sentimen. Selanjutnya pada proses pengumpulan data penulis mengumpulkan seluruh data dari wawancara, observasi dan mengambil data hasil survei pada database transaksional untuk dapat diolah. Selanjutnya dilakukan proses data mining yang menggunakan metode KDD yang terdiri dari proses *data selection, data processing, transformation, data mining* dan *interpretation/evaluation*.

Hasil dari wawancara dan observasi penulis dapat mengumpulkan kebutuhan yang diperlukan oleh user ditemukan bahwa *system existing* tidak terdapat analisis pada hasil survei sehingga organisasi tidak dapat mengetahui *sentiment* aplikasi yang telah dibangun dan telah dipakai oleh pengguna.

2.2 Metode Penelitian

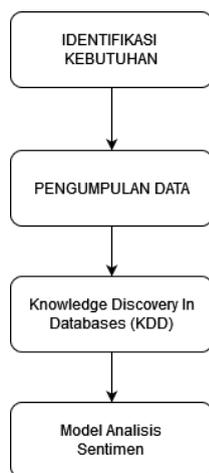
Dalam penelitian ini bahan penelitian yang digunakan untuk kemudian diolah menjadi acuan adalah data hasil survey kepuasan pengguna aplikasi CEISA di DJBC tahun 2021 sebagai objek yang diteliti.

2.2.1 Metode Pengumpulan Data

Tahap investigasi dan analisis dimulai dengan data yang diperoleh melalui beberapa metode. Data merupakan dasar yang kuat dalam menentukan masalah yang akan diselesaikan pada penelitian ini. Pengumpulan data dalam penelitian perlu dipantau agar data yang diperoleh terjaga tingkat validitas dan reliabilitasnya. Untuk mendapatkan data dengan tingkat validitas dan reliabilitas yang tinggi, penulis melakukan pengumpulan data dengan beberapa metode Wawancara, Observasi dan Studi Kepustakaan .

2.2.2 Metode Analisis Data

Tahapan penelitian yang dituangkan dalam diagram alir ini. Menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan :



Gambar 2. Tahapan Penelitian
Sumber : Penelitian Mandiri

Dari table berikut penulis menggambarkan Langkah-langkah penelitian yang terdiri dari 4 langkah sebagai berikut:

1. Identifikasi Kebutuhan
2. Pengumpulan data
3. Pengolahan *data mining*.
4. Model Analisis Sentimen.

Berikut merupakan tahapan yang dituangkan dalam diagram alir ini yang menggambarkan tahap penelitian yang ditempuh :

Tabel 1 Tahapan Penelitian

No	Langkah Penelitian	Proses	Output
1	Identifikasi Kebutuhan	Wawancara dan Observasi	Hasil wawancara terkait kebutuhan implementasi analisis sentimen terhadap data hasil survei aplikasi CEISA
2	Pengumpulan Data	Tarik data dari <i>database</i> transaksional hasil survei ke dalam bentuk <i>dataset</i>	<i>Dataset review</i> pengguna
3	<i>Knowledge Discovery In Databases (KDD)</i>	Terdiri dari proses <i>Data Selection, Preprocessing, Transformation, Data mining, Interpretation/Evaluation</i>	Model Analisis Sentimen
4	Model Analisis Sentimen	Pengujian menggunakan Algoritma Naïve Bayes.	Hasil Analisis Sentimen,

Accuracy,
Precision,
Recall dan
Grafik ROC

Sumber : Penelitian Mandiri

2.2.3 Pengolahan *Data Mining*

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam database terdapat tabel - tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (knowledge base) untuk keperluan pengambilan keputusan[1]. Pengolahan data mining yang dilakukan pada penelitian ini yaitu mengikuti tahapan *Knowledge Discovery in Database (KDD)*. Berikut merupakan tahapan-tahapan dalam KDD :

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data hasil survey kepuasan pelanggan yang perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi akan di gunakan untuk proses data mining, disimpan dalam suatu berkas. Data yang digunakan dalam penelitian ini merupakan data hasil survey kepuasan pengguna aplikasi CEISA. Terdapat tiga atribut utama dalam dataset ini yaitu field Identitas Pegawai (NIP), nama aplikasi yang digunakan, dan review pengguna. Dari semua data yang ada akan diseleksi dan digunakan hanya satu *field* yang digunakan untuk proses *knowledge discovery in database (KDD)*. *Filed* tersebut yaitu pada *field review* pengguna. Karena pada *field review* pengguna berisi *teks* yang dapat diolah untuk kebutuhan sentimen analisis.

2. *Preprocessing*

Preprocessing Data adalah bagian dari *text mining* yang diperlukan untuk menghapus *noise* yang muncul pada kalimat. *Preprocessing* data bertujuan menghindari data tidak sempurna, munculnya gangguan pada data dan *inkonsistent* pada data. Pada umumnya *preprocessing* bertujuan untuk menyelesaikan permasalahan *Bias-Variance Trade Off*.

Tahap ini adalah tahap awal dari proses KDD. Pada tahapan ini data yang tidak *relevan*, *missing value*, dan *radudant* harus dibersihkan. Hal

ini dikarenakan data yang *rrelevan*, tidak *missing*

value, dan tidak *redundant* merupakan syarat awal dalam melakukan *data mining*. Suatu data dikatakan *missing value* jika terdapat *atribut* dalam *dataset* yang tidak berisi nilai atau kosong, sedangkan data dikatakan *radudant* jika dalam satu *dataset* lebih dari satu *record* yang berisi nilai yang sama, setelah melakukan *cleaning* terhadap data yang lebih memenuhi syarat berdasarkan data hasil *survey* layanan pengguna aplikasi CEISA.

3. Transformation

Tahapan *Transformation* merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses transformasi dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Pada tahap ini penulis melakukan pelabelan pembagian data latih dan data uji. Pembagian label pada penelitian ini adalah pemberian label sentimen positif dan sentimen negatif pada *dataset* hasil *survey* layanan pengguna aplikasi CEISA.

4. Data Mining

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode yang digunakan pada penelitian ini adalah metode *Naive Bayes*. *Naive bayes* merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (*independent*). Keuntungan algoritma *naive bayes* adalah tingkat nilai *error* yang didapat lebih rendah ketika *dataset* berjumlah besar, selain itu akurasi *naive bayes* dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam *dataset* yang jumlahnya lebih besar.

5. Interpretation/Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada tahap ini didapatkan hasil sentimen analisis terhadap hasil *survey* aplikasi

CEISA dari proses *data mining* dengan metode *Naive Bayes*, pola atau

informasi yang dihasilkan dari proses data mining adalah berupa *rules* yang didapat dari perhitungan *Naive Bayes*.

2.3 Penentuan Data Training dan Testing

Data latih dan data uji diambil dari hasil *survey* kepuasan pengguna aplikasi CEISA dengan menggunakan metode *Naive Bayes*, dimana setelah dijumlahkan akan di *split* menjadi 70 % data latih dan 30% data uji.

2.4 Analisis Data Mining

2.4.1 Pengolahan Dokumen

Tools yang digunakan dalam eksperimen ini adalah *rapidminer* versi 10.0. Pada proses pengolahan *text rapidminer* membutuhkan *plugin* tambahan yang memiliki fitur pemrosesan teks diantaranya :

1. *Transform Cases* mentranmisikan dari teks ke *lowercase*
2. *Filter Stopword*, yakni menghilangkan teks yang bersesuaian dengan teks yang terdapat pada daftar *stopword*, karena teks tersebut dianggap tidak dapat mewakili konten dokumen.
3. *Filter Token*: melakukan *filter teks* dengan *Min char=5* dan *Max char=25*.
4. *Tokenizer*, merupakan proses yang bertujuan untuk memisah teks menjadi beberapa token berdasarkan pembatas berupa spasi atau tanda baca.

2.4.2 TF-IDF (Term Frequency – Inverse Document Frequency)

TF-IDF merupakan metode pembobotan *term* dengan menggunakan *term frequency* yaitu jumlah *term* yang terdapat pada tiap dokumen serta *inverse document frequency* yaitu *invers* jumlah dokumen yang memuat suatu *term*.

2.5 Instrumen Penelitian

Penelitian ini menggunakan perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai alat pendukung dalam melaksanakan penelitian dan merancang aplikasi. Adapun Perangkat keras yang digunakan dalam pembuatan sistem aplikasi diantaranya meliputi:

3. HASIL DAN PEMBAHASAN

menghasilkan sentimen terhadap kepuasan pengguna pada hasil survei aplikasi CEISA di Direktorat Jenderal Bea dan Cukai dengan menggunakan algoritma *Naive Bayes*.

3.1 Hasil Identifikasi Kebutuhan Sistem

Data yang dibutuhkan untuk penelitian ini diperoleh dengan beberapa cara seperti wawancara dan observasi. Data tersebut merupakan dasar penulis menganalisis kebutuhan yang akan dikembangkan. Hasil dari pengumpulan kebutuhan bisnis yang telah dilakukan menyatakan bahwa Direktorat Jenderal Bea dan Cukai perlu mengetahui level sentimen dari aplikasi yang telah dibuat dan dipakai oleh pengguna. Level sentimen dapat mengetahui respon user yang menggunakan aplikasi CEISA sehingga hasilnya dapat menjadi bahan evaluasi untuk perbaikan kedepan.

Dalam keadaan saat ini pengolahan data survei dinilai tidak dilakukan secara optimal. Saat ini output hasil survei hanya diolah menggunakan data excel. Sehingga tidak akan optimal jika mengolah data dalam jumlah besar. Maka dari itu diusulkan untuk membuat model data mining yang dapat mengetahui level sentimen aplikasi CEISA oleh pengguna.

3.1 Pengumpulan Data

Proses pengumpulan data merupakan proses untuk mengumpulkan seluruh data dukung yang diperlukan untuk penelitian penulis. Data dukung merupakan gabungan dari proses wawancara yang telah dilakukan dengan database analisis, observasi di lapangan serta melakukan penarikan data dari database transaksional ke dalam sebuah dataset sehingga memungkinkan data untuk diolah.

3.2 Proses Knowledge Discovery In Databases (KDD)

Analisis ini berakhir dalam proses *data mining* yang memiliki hasil yang dicapai oleh peneliti adalah untuk mengetahui sentimen terhadap hasil survei kepuasan pengguna di Direktorat Jenderal Bea Cukai. Penulis menggunakan aplikasi Rapidminer yang menghasilkan nilai *accuracy* 83,97%, untuk nilai *precision* yaitu 80,99% pada class *positif* 80,18 % pada class negatif 88,95%, sedangkan nilai *recall* 90,55 pada class *positif* dan 77,27% untuk nilai

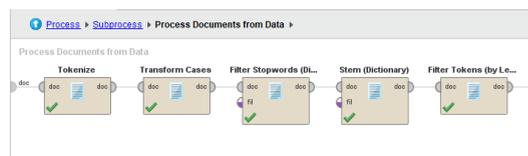
3.2.1 Data Selection

Data diatas merupakan data mentah, untuk data yang akan dilakukan analisis hanya pada atribut *Review Pengguna*. Berikut merupakan data yang siap digunakan dalam proses *preprocessing* dapat dilihat pada tabel.

3.2.2 Preprocessing

Dalam penelitian ini dilakukan tahap *preprocessing*. *Preprocessing* dilakukan untuk menghilangkan *noise* pada *dataset* yang digunakan untuk sentimen analisis. Dimana data yang kita proses akan kita ambil informasi yang terkandung di dalamnya dalam hal sentimen penulisnya yaitu sentimen *positif* dan sentimen *negatif*. Guna memudahkan dalam mengelola data maka data perlu kita berikan analisis sentimen secara manual dengan membaca maksud dari kalimat dari sentimen tersebut. Tahapan *preprocessing* pada penelitian kali ini proses *preprocessing* dibagi menjadi 5 tahap yaitu;

1. *Tokenize* adalah pemecahan kalimat jadi sebuah kata
2. *Transform Cases* adalah mengubah huruf menjadi kecil semua/sebaliknya
3. *Filter Stopwords* adalah penghapusan atau pengecualian kata-kata stopwords dari teks atau kalimat yang sedang dianalisis;
4. *Stemming* adalah mengurangi kata-kata ke dalam bentuk dasar atau menghilangkan kata-kata yang memiliki imbuhan menjadi kata dasar;
5. *Filter Token (by length)* merupakan proses menghilangkan token menjadi lebih pendek ataupun lebih Panjang.



Gambar 3. 1 Tahapan *Preprocessing*

3.2.3 Transform Cases

Dalam fitur *transform case* tersendiri kita dapat otomatis mengubah huruf besar menjadi huruf kecil ataupun merubah dari huruf kecil menjadi huruf kapital. Dalam penelitian ini penulis mengubah dari huruf besar menjadi kecil karena

terdiri dari huruf kecil selain itu cara ini juga untuk meminimalisasi kesalahan pada proses *tokenize*. Berikut merupakan contoh kalimat yang sudah di *transform cases* dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Sebelum *Transform Case*

No	Review
1	Aplikasi tidak dapat diakses Semoga kedepan kendalakendala dapat diatasi sehingga aplikasi CEISA IMPOR dapat digunakan secara maksimal
2	Gagal upload dokumen
3	Terkadang Aplikasi lambat bahkan tidak bisa diakses
4	kadang eror tapi sangat jarang dan biasanya terjadi karena gangguan dari pusat
5	Urutan nomor pos pada menu browse manifest inward berantakan, atas manifest inward kapal A namun pos berisi milik kapal B
6	Pada beberapa kali mengakses terjadi gangguan pada gangguan CEISA Ekspor, Sebaiknya apabila ada gangguan dapat segera diberitahukan pada grup CEISA, selebihnya aplikasi CEISA Ekspor telah sangat bagus
7	diharapkan aplikasi lebih stabil
8	Terkadang terdapat kendala jaringan
9	data penutupan pos manifest masih tidak akurat
10	Aplikasi masih suka error

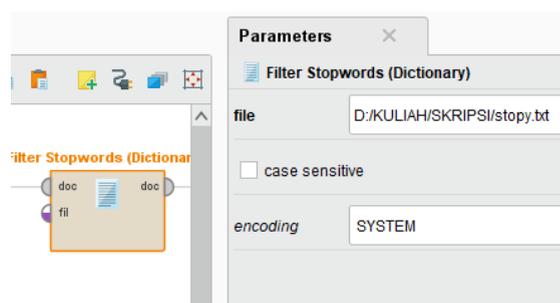
Tabel 3. 2 Sesudah *Transform Case*

No	Review
1	aplikasi akses semoga depan kendala atas aplikasi ceisa impor maksimal
2	gagal upload dokumen
3	kadang aplikasi lambat akses
4	kadang error jarang ganggu pusat
5	urutan nomor menu browse manifest berantakan manifest kapal milik kapal
6	kali akses ganggu ganggu ceisa ekspor ganggu diberitahukan grup ceisa lebih aplikasi ceisa ekspor bagus
7	diharapkan aplikasi stabil
8	kadang kendala jaringan
9	data tutup manifest akurat
10	aplikasi suka error

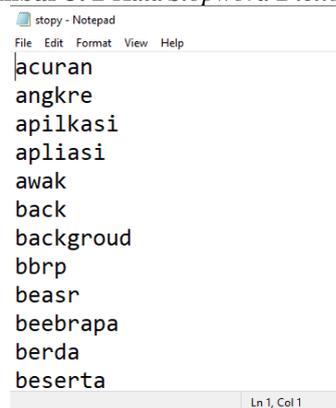
Pada tabel dapat dilihat gambar sebelum dilakukan *transform cases*. Pada tahap tersebut huruf tidak seragam sehingga masih terdapat campuran huruf kapital dan huruf kecil. Hal tersebut merupakan gambar setelah dilakukan *transform cases*. Tabel tersebut terlihat kalimat yang terdiri dari kalimat *lowercase*.

3.2.4 Filter Stopword

Dengan fitur ini maka *teks* sebelum di klasifikasikan di hilangkan dulu teks yang tidak berhubungan dengan analisis sentimen sehingga dimensi teks akan berkurang tanpa mengurangi isi sentimen dari teks tersebut. *Filter stopwords* bahasa indonesia ini penulis ambil dari internet serta dikompilasi dengan menambahkan *stopword* manual yang tidak ada dalam *stopword existing*. Stopword berikut merupakan *stopword* yang berisi kata-kata sambung dan singkatan yang tidak ada hubungannya dalam penelitian ini.



Gambar 3. 2 Kata *Stopword Dictionary*



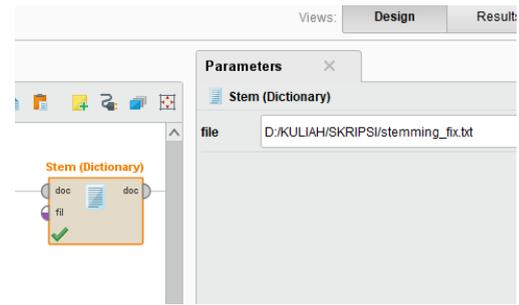
Gambar 3. 3 Kata *Stopword Dictionary* Dalam File txt

Gambar 3.3 merupakan kata yang dijadikan *dictionary* yang dimasukan dalam file txt. Contohnya menghapus kata-kata typo, kata sambung dan singkatan contoh “bbrp”, “awak”, “angkre”.

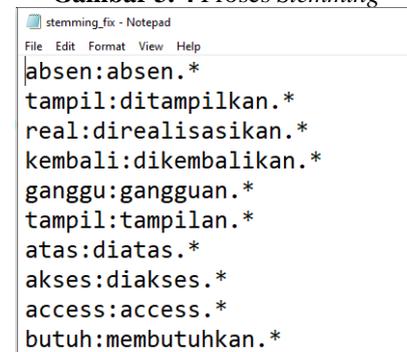
Tabel 3. 3 Proses *Stopword*

No	Sebelum	Sesudah
1	Terkadang Aplikasi lambat bahkan tidak bisa diakses	kadang aplikasi lambat akses
2	kadang eror tapi sangat jarang dan biasanya terjadi karena gangguan dari pusat	kadang error jarang ganggu pusat
3	Urutan nomor pos pada menu browse manifest inward berantakan, atas	urutan nomor menu browse manifest berantakan manifest kapal milik kapal berantakan, atas

	manifest inward kapal A namun pos berisi milik kapal B	
4	Pada beberapa kali mengakses terjadi gangguan pada gangguan CEISA Ekspor, Sebaiknya apabila ada gangguan dapat segera diberitahukan pada grup CEISA, selebihnya aplikasi CEISA Ekspor telah sangat bagus	kali akses ganggu ganggu ceisa ekspor ganggu diberitahukan grup ceisa lebih aplikasi ceisa ekspor bagus
5	diharapkan aplikasi lebih stabil	diharapkan aplikasi stabil
6	Terkadang terdapat kendala jaringan	kadang kendala jaringan
7	data penutupan pos manifest masih tidak akurat	data tutup manifest akurat
8	Aplikasi masih suka error	aplikasi suka error



Gambar 3.4 Proses Stemming



Gambar 3.5 File txt Stemming

Gambar 3.5 merupakan kata yang dijadikan *dictionary* yang dimasukkan dalam file txt. Contohnya merubah kata “gangguan” menjadi “ganggu”. Data Stemming terlampir pada Lampiran 3

3.2.5 Tokenizing

Pada tahap ini setiap kalimat akan dipisahkan menjadi kata berdasarkan spasi yang ditemukan. Berikut merupakan hasil dari proses *preprocessing* yang telah dilakukan *cleanshing*, *filtering*, *stemming* dan *tokenizing*.

Tabel 3.4 Proses Tokenizing

akibat	Akses	akurat	alami	analisis	aplikasi	april
0	0	0	0	0	0	0
0	0.24	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0.49	0	0	0
0	0	0	0	0.32	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0.07	0
0	0.18	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

tokenisasi adalah untuk memisahkan *teks* menjadi bagian-bagian yang lebih kecil sehingga setiap unit bisa dianggap sebagai *entitas*. Nilai yang muncul saat proses tersebut akan bernilai semakin besar

Pada tabel berikut merupakan perbandingan antara data sebelum dan sesudah proses *stopword*. Misal pada kata *input* “Terkadang Aplikasi lambat bahkan tidak bisa diakses” maka akan menghasilkan output kalimat “kadang aplikasi lambat akses”. Kata “bahkan”, “tidak”, dan “bisa” dihilangkan pada proses *stopword*.

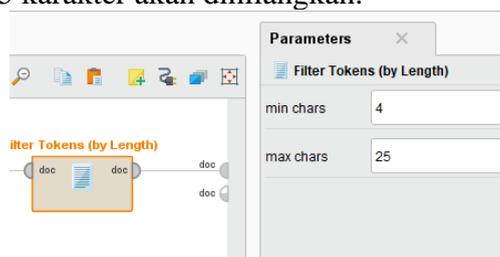
3.2.4 Stemming

Stemming merupakan proses mengubah kata berimbuhan menjadi bentuk kata dasarnya. Algoritma *stemming* tidak dapat digunakan untuk berbagai macam bahasa dikarenakan memiliki morfologi yang berbeda. Pada penelitian ini penulis menggunakan *Stemming (Dictionary)*. *Stemming Dictionary* merupakan proses *stemming* yang melakukan input secara manual pada sebuah file .txt. Penulis menggunakan kombinasi *regex* dalam file *stamming_fix.txt* Pada proses *stemming* ini penulis mencoba melakukan filter untuk kata-kata yang memiliki imbuhan sehingga kata yang ditampilkan hanya menampilkan kata dasar. Hal ini membuat proses pembobotan kata menjadi lebih valid. Berikut merupakan contoh

jika kata nya sering muncul. Jadi semakin besar nilainya maka kata tersebut akan menjadi semakin besar dan mempengaruhi saat analisis sentimen.

3.2.6 Filter Token (by Length)

Ini adalah proses yang ada pada data *preparation* untuk menghilangkan sejumlah kata (setelah proses *tokenize*) dengan panjang karakter tertentu. Pada penelitian ini panjang minimum karakter yang digunakan adalah 4 karakter dan panjang maksimum 25 karakter. Artinya kata yang panjangnya kurang dari 4 karakter dan lebih dari 25 karakter akan dihilangkan.



Gambar 3. 6 Parameter Token (by Length)

Untuk mendapatkan hasil seperti ini maka dilakukan *setting* pada *Parameters* dari operator ini.

Tabel 3. 5 Penggunaan Token By Length

No	Sebelum	Sesudah
1	Terkadang Aplikasi lambat bahkan tidak bisa diakses	kadang aplikasi lambat akses
2	kadang eror tapi sangat jarang dan biasanya terjadi karena gangguan dari pusat	kadang error jarang ganggu pusat
3	Urutan nomor pos pada menu browse manifest inward berantakan, atas manifest inward kapal A namun pos berisi milik kapal B	urutan nomor menu browse manifest berantakan manifest kapal milik kapal
4	Pada beberapa kali mengakses terjadi gangguan pada gangguan CEISA Ekspor, Sebaiknya apabila ada gangguan dapat segera diberitahukan pada grup CEISA, selebihnya aplikasi CEISA Ekspor telah sangat bagus	kali akses ganggu ganggu ceisa ekspor ganggu diberitahukan grup ceisa lebih aplikasi ceisa ekspor bagus
5	diharapkan aplikasi lebih stabil	diharapkan aplikasi stabil

6	Terkadang terdapat kendala jaringan	kadang kendala jaringan
7	data penutupan pos manifest masih tidak akurat	data tutup manifest akurat
8	Aplikasi masih suka error	aplikasi suka error

Pada proses *Filter Token (by length)* panjang minimum karakter yang digunakan adalah 4 karakter dan panjang maksimum 25 karakter. Terdapat *input* pada kata “kadang eror tapi sangat jarang dan biasanya terjadi karena gangguan dari pusat” kemudian menghasilkan *output* “kadang *error* jarang ganggu pusat”. Pada *output* kalimat tersebut sudah dihilangkan seluruh kata penghubung agar terhindar kata-kata noise dan kalimat dibatasi paling sedikit 4 kata saja dan maksimal 25 kata.

3.3 Transformation

Tahap *transformation* ini hasil dari pengelompokan data *preprocessing* kemudian digunakan untuk data training. Pada tahap ini dilakukan proses pemberian label pada data latih. Pemberian label dalam penelitian ini dilakukan secara manual menggunakan *microsoft excel*. Pada tahap pelabelan penulis dibantu dengan pegawai yang ditugaskan untuk melakukan pelabelan pada Direktorat IKC. Data latih yang digunakan sebanyak 526 data latih. Sedangkan data uji yang digunakan merupakan data uji yang sama digunakan untuk data latih. Pada proses pengujiannya akan dilakukan dengan membagi data sentimen menjadi bobot yang sama pada masing-masing label sentimen positif dan sentimen negatif Adapun data training dapat dilihat pada Gambar berikut :



Gambar 3. 7 Proses Pemberian Label Pada Data Latih

Tabel 3. 6 Proses Pelabelan Manual

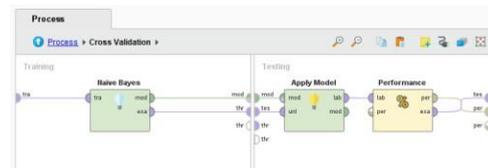
No	Label	Review
1	NEGATIF	beberapa waktu aplikasi tidak bisa

		diakses, agar aplikasi lebih stabil
2	NEGATIF	kalau hari jumat kadang eror
3	POSITIF	Sampai saat ini MyCEISA pada iOS dan Android bermanfaat dan sangat membantu dalam mendukung kelancaran proses pelayanan informasi Namun, kendala yang umum ditemui dalam 3 bulan terakhir adalah tidak bisa terbuka halaman awal, terutama absen pada pukul 700 ? 0800 WIB dan 1700 ? 1900 WIB Hal ini cukup berdampak pada proses pelayanan informasi terutama pada CEISA Impor yang hanya dapat diakses dengan login menggunakan QR Code pada MyCEISA pada shift sore dan shift pagi (pada weekend/ libur nasional), terima kasih
4	NEGATIF	beberapa kali harus menunggu agar aplikasi bisa diakses
5	POSITIF	Mantap
6	POSITIF	Terdapat beberapa menu yang tidak dapat diakses, seperti pada lampiran B
7	NEGATIF	beberapa kali ada keluhan dari pengguna jasa bahwa proses penerbitan NPE lama (biasanya status data di INSW sudah Proses Kepabebean namun saat di cek pada aplikasi Ceisa ekspor belum masuk)
8	NEGATIF	Susah akses

validation. Proses data *training* dan *testing* didalam blok proses (*validation*) dapat ditampilkan dengan melakukan klik 2x menggunakan *mouse* sehingga muncul proses *training* dan *testing*.

3.4.2 Proses Training dan Testing

Proses *training* yaitu melakukan proses pelatihan data pada metode algoritma *Naïve Bayes*. Sedangkan proses *testing* yaitu melakukan pengujian data yang akan menghasilkan grafik atau pola. Pada Gambar 3.9 dijelaskan bahwa proses *training* digunakan untuk blok *model* algoritma *Naïve Bayes*, dihubungkan dengan garis penghubung pada blok *apply model* dan blok *performance* dengan bagian *testing* sebagai penampil informasi hasil dari pengujian data, hasil pengujian ini akan menghasilkan model algoritma *Naïve Bayes*.



Gambar 3. 9 Training dan Testing

3.4.3 Hasil Performance Vector

Proses klasifikasi dengan metode algoritma *Naïve Bayes* yang digunakan untuk mengklasifikasikan komentar positif dan negatif pada penelitian ini sehingga diperoleh nilai *Accuracy*, *precision*, dan *recall*.

1. Accuracy

Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi yaitu 83,97% dari hasil data *testing*

	true NEGATIF	true POSITIF	class precision
jumlah NEGATIF	153	19	88.95%
jumlah POSITIF	45	182	80.18%
class recall	77.07%	88.95%	

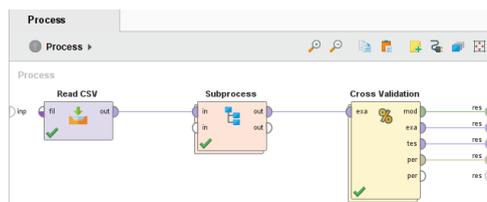
Gambar 3. 10 Accuracy

2. Precision

Precision adalah jumlah data yang *true positive* (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif. Dari hasil pengujian nilai *precision* yaitu 88.95% untuk *class* positif dan 80.18% untuk *class* negatif

3.4 Pemrosesan Data Mining

Pemrosesan data mining dilakukan menggunakan metode dan algoritma tertentu. Algoritma yang digunakan dalam proses ini merupakan algoritma *Naïve Bayes*. Terdapat dua kategori sentimen. Pada tahap awal kalimat-kalimat hasil ulasan yang telah diberikan label pada tahap *transformation* dikategorikan menjadi sentimen positif dan sentimen negatif.



Gambar 3. 8 Proses Cross Validation

3.4.1 Validasi

Proses ini merupakan proses *validation*. Melakukan *validation* yaitu melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik. Pada Gambar berikut merupakan proses validasi, setelah pembacaan *file* data, blok *read csv* dihubungkan dengan blok

precision: 80.99% +/- 7.73% (micro average: 80.99%) (positive class: POSITIF)			
	True NEGATIF	True POSITIF	Class precision
pred NEGATIF	153	19	88.95%
pred POSITIF	45	182	80.18%
class recall	77.27%	90.55%	

Gambar 3. 11 Precision

3. Recall

Recall adalah jumlah data yang *true positive* dibagi dengan jumlah data yang sebenarnya positif (*true positive + true negative*). Untuk nilai *recall* yaitu 90,55% pada *class* positif dan nilai *class* negatif yaitu 77,27% .

recall: 90.55% +/- 8.66% (micro average: 90.55%) (positive class: POSITIF)			
	True NEGATIF	True POSITIF	Class precision
pred NEGATIF	153	19	88.95%
pred POSITIF	45	182	80.18%
class recall	77.27%	90.55%	

Gambar 3. 12 Recall

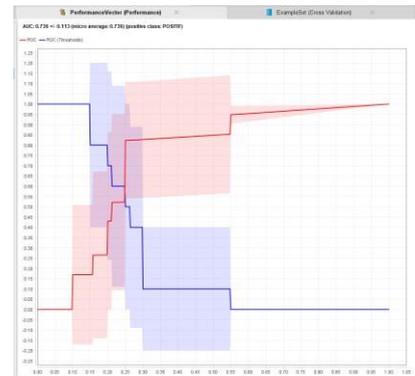
Dari hasil pengujian untuk nilai *accuracy* 83.97%, untuk nilai *precision* yaitu 80.99% pada *class* positif 80.18 % pada *class* negatif 88,95%, sedangkan nilai *recall* 90,55 pada *class* positif dan 77,27% untuk nilai *class* negatif. Hasil *class accuracy*, *precision* dan *recall* untuk memprediksi komentar positif dan negatif pada hasil *survey* layanan kepuasan pengguna pada aplikasi CEISA di Direktorat Jenderal Bea dan Cukai.

Tabel 3. 7 Hasil Class Accuracy, Precision dan Recall

Class	Accuracy (%)	Precision (%)	Recall (%)
Komentar Positif	83.97	80.18	90,55
Komentar Negatif		88,95	77,27

4. AUC (Area Under Curve)

Berikut merupakan kurva *Receiver Oprating Characteristic* (ROC) yang dihasilkan oleh rapid miner 10.0. Kurva ROC digunakan untuk menunjukkan data *confusion matrix*. Garis lurus/ *horizontal* mewakili nilai *False Positive* (FP) dan garis vertikal mewakili nilai *True Positive* (TP). Dari Gambar 3.13 dapat diketahui bahwa nilai *Area Under Curve* (AUC) model algoritma *Naïve Bayes* adalah 0.736. Hal ini menunjukkan bahwa model algoritma *Naïve Bayes* mencapai klasifikasi yang baik.



Gambar 3. 13 Area Under Curve (AUC)

```

PerformanceVector (Performance)
PerformanceVector:
accuracy: 83.97% +/- 6.87% (micro average: 83.96%)
ConfusionMatrix:
True: NEGATIF POSITIF
NEGATIF: 153 19
POSITIF: 45 182
precision: 80.99% +/- 7.73% (micro average: 80.18%) (positive class: POSITIF)
True: NEGATIF POSITIF
NEGATIF: 153 19
POSITIF: 45 182
recall: 90.52% +/- 8.66% (micro average: 90.55%) (positive class: POSITIF)
ConfusionMatrix:
True: NEGATIF POSITIF
NEGATIF: 153 19
POSITIF: 45 182
AUC (optimistic): 0.975 +/- 0.024 (micro average: 0.975) (positive class: POSITIF)
AUC (pessimistic): 0.736 +/- 0.113 (micro average: 0.736) (positive class: POSITIF)
AUC (meanistic): 0.704 +/- 0.109 (micro average: 0.704) (positive class: POSITIF)
    
```

Gambar 3. 14 Simple Distribution Model

Berikut merupakan hasil analisis dari tabel data komentar positif dan komentar negatif menggunakan metode *Naïve Bayes* dapat menghasilkan 2 kelas. Pada hasil *performance vector* tersebut terdapat nilai *accuracy* sebesar 83.97 %. Nilai *accuracy* dikategorikan baik karena ada pada *range* nilai 0.7 – 0.8. Nilai *precision* sebesar 80.99 %. Nilai *accuracy* dikategorikan baik karena ada pada *range* nilai 0.7 – 0.8. Berikut merupakan hasil sentimen analisis berdasarkan pembobotan kata. Pada tabel ini tersedia prediksi yang nilainya berdasarkan pembobotan dari setiap kata yang telah dilakukan pembobotan nilai berdasarkan TF-IDF dan pemodelan *Algoritma Naïve Bayes*.

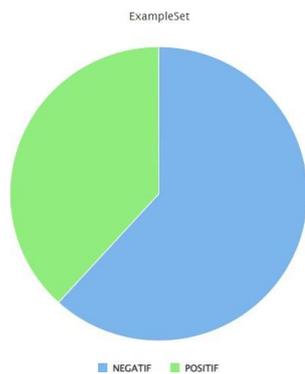
Row No.	LABEL	prediction LABEL	confidence1	confidence2	text	altosen	altobes
23	NEGATIF	POSITIF	1	0	aplikasi baik	0	0
24	NEGATIF	NEGATIF	0	1	tentunya baik	0	0
25	NEGATIF	NEGATIF	0	1	di saat pengp.	0	0
26	NEGATIF	POSITIF	1	0	kegiatan di atas	0	0
27	NEGATIF	NEGATIF	0	1	data kluar ma.	0	0
28	NEGATIF	NEGATIF	0	1	dukumen dtt.	0	0
29	NEGATIF	NEGATIF	0	1	info atampn.	0	0
30	NEGATIF	NEGATIF	0	1	data akan di.	0	0
31	NEGATIF	NEGATIF	0	1	gsk jawa ra.	0	0
32	NEGATIF	POSITIF	1	0	revisi dlm.	0	0
33	NEGATIF	NEGATIF	0	1	kegiatan di atas	0	0
34	NEGATIF	NEGATIF	0	1	aplikasi baik	0	0
35	NEGATIF	NEGATIF	0	1	lamar data dt.	0	0

Gambar 3. 15 Prediksi Sentimen Dengan Pembobotan Kata

Tabel 3. 8 Pengujian

Label Sentimen		Accuracy (%)	Precision(%)	Recall (%)
Positif	Negatif			
50	50	93.29	89.00	92.00
100	100	93.50	93.35	94.00
200	200	83.40	80.54	89.54
201	325	80.79	92.69	75.09

Tabel diatas merupakan tabel pengujian *Confussion Matrix* menggunakan data data *sentiment* yang telah diberi label untuk mengetahui seberapa akurat Analisis Sentimen Pada Hasil Survei Aplikasi CEISA. Tahap pengujian ini memberikan semua label yang seimbang dengan proporsi sentiment positif dan negatif yang sama. Pada tahap pengujian ini dibagi menjadi 4 kali pengujian. Pengujian pertama menggunakan 50 label positif dan 50 label negatif. Pengujian kedua menggunakan 100 label positif dan 100 label negatif. Pengujian ketiga menggunakan 200 label positif dan 200 label negatif. Pengujian terakhir menggunakan 201 label positif dan 325 label negatif.



Gambar 3. 16 Presentasi Sentimen Analisis

Pada *piechart* dibawah ini terdapat informasi jika dalam *dataset survey* layanan pengguna aplikasi CEISA bahwa terdapat informasi jika sentimen negatif lebih dominan daripada sentimen positif. Hasil dapat dilihat pada Gambar 3.16.

3.5 Menguji Sistem

Dilakukan pengujian terhadap sistem baik fungsi dan kesesuaian dengan kebutuhan yang telah dikumpulkan. Pengujian ini merupakan salah satu tolak ukur keberhasilan terhadap pengumpulan kebutuhan, analisis rancangan hingga ke tahap pengembangan.

3.5.1 Skenario Pengujian Sistem

Unit testing ini merupakan tahap dilakukan pengujian hasil preprocessing dan performance

vector serta grafik ROC terhadap suatu model yang dihasilkan. Pengujian dilakukan untuk melihat apakah fungsi dari setiap elemen berjalan sesuai dengan yang diharapkan. Pada pengujian ini, penulis menggunakan teknik *black box testing* dengan meninjau masukan dan keluaran. Pengujian dilakukan oleh pihak yang memiliki kompetensi dalam memahami pembuatan model dan pengujian. Status dikatakan berhasil apabila skenario yang dilakukan terhadap sistem menghasilkan keluaran sesuai dengan yang diharapkan.

Tabel 3. 9 Skenario Pengujian Sistem

No Uji Kasus	Skenario Pengujian	Jenis Pengujian
1 <i>Tokenize</i>	Melakukan pemenggalan menjadi kata	<i>Black Box</i>
2 <i>Transform Case</i>	Mengubah huruf kecil menjadi besar semua	<i>Black Box</i>
3 <i>Filter Stopword</i>	Menghapus mengecualikan kata-kata stopwords dari teks atau kalimat yang sedang dianalisis	<i>Black Box</i>
4 <i>Stemming</i>	Mengurangi kata-kata penghubung menjadi bentuk dasar	<i>Black Box</i>
5 <i>Filter Token (by length)</i>	Menghilangkan token/kata menjadi lebih pendek	<i>Black Box</i>
6 <i>Accuracy</i>	Mengukur dari suatu algoritma yang diterapkan	<i>Black Box</i>

3.5.2 Hasil Pengujian Sistem

Berdasarkan Rancangan skenario pengujian yang telah dibuat sebelumnya, berikut merupakan tabel hasil pengujian dari masing - masing *case* menggunakan sistem pengujian *Black Box*.

No	Uji Kasus	Data Masukan	Hasil yang diharapkan	Hasil yang pengamatan	Kesimpulan
1	Tokenize	Dataset	Sistem mampu melakukan pemenggalan kalimat kedalam suatu kata	Kalimat sudah terpisah menjadi kata	Sesuai
2	Transform Case	Dataset	Sistem mampu melakukan perubahan huruf besar menjadi huruf kecil semua	Huruf besar berubah menjadi huruf kecil semua.	Sesuai
3	Filter Stopword	Dataset	Sistem mampu menghapus atau mengecualikan kata-kata stopwords dari teks atau kalimat yang sedang dianalisis	Sistem mampu menghapus kata-kata hubung dan kata yang tidak diperlukan seperti akronim, singkatan	Sesuai
4	Stemming	Dataset	Sistem mampu mengurangi kata-kata penghubung menjadi bentuk dasar	Sistem mampu menghapus kata-kata hubung seperti "dan", "atau"	Sesuai
5	Filter Token (by length)	Dataset	Sistem Mampu menghilangkan token/kata menjadi lebih pendek	Kata yang tidak penting berhasil dihilangkan dan dibatasi menjadi 4 kata saja	Sesuai
6	Accuracy	Model	Sistem mampu mendapatkan accuracy algoritma minimal rata-rata 0.7.	Accuracy rata-rata yang didapatkan sebesar 0.8079	Sesuai

4. KESIMPULAN DAN SARAN

Bab ini adalah bab terakhir dalam laporan skripsi ini yaitu kesimpulan dan saran. Dalam bab ini, penulis menguraikan kesimpulan dan saran dari hasil penelitian yang telah dilakukan. Kesimpulan dan saran tersebut merupakan jawaban dari pertanyaan penelitian.

4.1 Kesimpulan

Berdasarkan penelitian analisis sentimen tingkat kepuasan pengguna terhadap kepuasan

pengguna aplikasi CEISA di Direktorat Jenderal Bea dan Cukai yang telah dilakukan maka kesimpulan yang dapat diambil adalah:

1. Telah dilakukan analisis sentimen tingkat kepuasan pengguna terhadap aplikasi CEISA di Direktorat Jenderal Bea dan Cukai pada *dataset* yang memiliki *attribute Review* yang berisi ulasan pengguna aplikasi CEISA.
2. Metode yang digunakan pada proses *preprocessing* adalah metode *Knowledge Discovery in Database (KDD)* yang meliputi proses *filter token (by length) filter stopwords, stemming* dan tokenisasi
3. Pengujian menggunakan metode *Confussion Matrix* dengan menggunakan Algoritma *Naïve Bayes* menghasilkan nilai akurasi 80,79 %, presisi 92.69 % *recall* 75.09 % dan nilai AUC 79,5 % yang dapat dikategorikan menjadi nilai akurasi yang baik.

4.2 Saran

Berdasarkan simpulan yang telah diuraikan, maka ada beberapa saran yang dapat diuraikan yaitu:

1. Pada penelitian selanjutnya dapat dikembangkan dengan menambahkan analisis pada sentimen yang bersifat netral.
2. Pada penelitian selanjutnya dapat dikembangkan sentimen analisis menggunakan metode lain yang sehingga nilai akurasi bisa lebih dari 80 % sehingga kemungkinan *error* semakin kecil.

DAFTAR PUSTAKA

- [1] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [2] D. APLIKASINYA Tim Penulis *et al.*, *DATA MINING*. 2021. [Online]. Available: www.penerbitwidina.com
- [3] Z. I. 'Akbar, "Apa itu Text Mining? ," *binus*, Apr. 23, 2021.
- [4] R. dkk Safitri, *ANALISIS SENTIMEN : METODE ALTERNATIF PENELITIAN BIG DATA*, 1st ed., vol. 1. Malang: Tim UB Press, 2021.
- [5] adminlp2m, "Analisis Sentimen (Sentiment

