

ANALISIS SENTIMEN PENGGUNA APLIKASI DANA BERDASARKAN ULASAN PADA GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Abitdavy Athallah Muhammad¹, Ermatita², Desta Sandya Prasvita³

Prodi S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450

dayyabit@gmail.com, ermatita@upnvj.ac.id, desta.sandya@upnvj.ac.id

Abstrak. Perkembangan teknologi pada era digital ini sangat pesat, salah satu bidang yang sangat berkembang ialah finansial. Beberapa perusahaan dari aplikasi fintech mulai bermunculan, dan salah satunya adalah DANA. Dari kemudahan yang ditawarkan, banyak orang yang mulai menggunakan aplikasi DANA. Dan sebab semakin banyak pengguna, maka semakin banyak juga ulasan pada Google Play Store. Ulasan dari pengguna dapat menjadi bahan menjadi masukan kepada perusahaan yang berkepentingan sehingga dapat menjadi perbaikan kedepannya. Pada penelitian ini data didapat dari ulasan pengguna aplikasi dompet digital DANA pada Google Play Store. Data ulasan dibagi menjadi 2 kategori yaitu data bersifat positif dan negatif berdasarkan pelabelan manual oleh 3 anatator. Metode support vector machine dan seleksi fitur chi square digunakan dalam pembuatan model. Hasil evaluasi diperoleh akurasi sebesar 87,58%, presisi sebesar 91,20%, dan recall sebesar 90,21% untuk model SVM, sedangkan untuk model SVM – chi square diperoleh akurasi sebesar 89,41%, presisi sebesar 93,29%, dan recall sebesar 90,76%.

Kata Kunci: Klasifikasi, Support Vector Machine, Sentimen

1 Pendahuluan

Pada zaman dimana teknologi berkembang pesat, teknologi menjadi peran yang penting bagi pendukung kebutuhan manusia. Hal apapun yang dilakukan manusia dapat dilakukan dengan mudah dan efisien dengan adanya teknologi. Begitu besarnya dampak teknologi ini memulai membawa kita ke peradaban era digital. Dengan meninggi meninggi permintaan akan informasi dan data memicu manusia untuk mengembangkan teknologi yang lebih baru untuk lebih mudah dan cepatnya pengolahan data dan informasi yang dilakukan. Dan dalam era ini teknologi yang sangatlah digunakan dan dibutuhkan banyak orang merupakan smartphone [5].

Dari hasil inovasi dan fasilitas yang ditawarkan oleh media transaksi tersebut, banyak orang yang telah menggunakan aplikasi dompet digital. Salah satu yang populer dari beberapa aplikasi dompet digital yang terdapat pada Google Play Store adalah DANA. DANA adalah Financial Technology (Fintech) yang menyediakan layanan transaksi dengan smartphone pengguna. Tentunya dengan banyaknya layanan yang DANA sediakan menimbulkan feedback dari pengguna DANA dalam ulasan yang terdapat di Google play store. Seperti layanan untuk topup terjadi kesalahan atau gagal, tidak mendapatkan cashback yang seharusnya dan customer service yang sulit untuk dihubungi [5].

Dalam menemukan informasi mengenai produk yang ditawarkan, ulasan dari pengguna digunakan sebagai acuan yang efektif. Menurut (Praptiwi, 2018), menyatakan penelitian baru - baru ini ditemukan bahwa hampir lebih 50% pengguna internet bergantung pada opini-opini pengguna produk sebelum memakai suatu produk, karena ulasan menyediakan informasi yang terbaru terkait produk yang digunakan berdasarkan perspektif dari pengguna produk tersebut. Tetapi, untuk mengobservasi dan mengelompokkan opini dari masyarakat tidaklah mudah. Opini yang dimuat dalam ulasan terutama dari produk yang sudah banyak digunakan, diperlukan metode dan teknik yang bisa mengklasifikasikan ulasan-ulasan tersebut secara otomatis karena banyaknya ulasan untuk diproses secara manual.

Penelitian yang dilakukan Ernawati (2016) menunjukkan bahwa penggunaan seleksi fitur merupakan salah satu langkah penting dalam klasifikasi dan dapat mempengaruhi secara langsung performa dari model klasifikasi. Dapat diketahui bahwa klasifikasi pada teks memiliki masalah yang berkaitan dengan banyaknya dimensi yang digunakan. Hal ini disebabkan karena pada dokumen dari klasifikasi teks, digambarkan sebagai kumpulan dari kata-kata, dimana setiap kata dalam suatu dokumen tidak bergantung dengan kata yang lain.

Nur Fitriyah, Budi Warsito, dan Diasih I Maruddani (2020) melakukan penelitian berjudul “Analisis Sentimen Gojek pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)” dimana dalam penelitian ini digunakan metode support vector machine dan didapatkan akurasi sebesar 79,19% dengan menggunakan kernel RBF. Sehingga tujuan penelitian ini adalah bagaimana membangun model klasifikasi sentiment terhadap ulasan aplikasi DANA menggunakan metode *support vector machine* dan mengetahui performa dari model yang dibuat serta pengaruh penggunaan seleksi fitur *chi square*.

2 Tinjauan Pustaka

2.1 DANA

DANA merupakan dompet digital serta tempat pembayaran digital secara terbuka yang umum digunakan oleh aplikasi-aplikasi yang sudah terdaftar, maupun gerai-gerai online yang konvensional. Platform dompet digital ini mengembangkan infrastruktur pembayaran digital dengan menyelidiki praktik terbaik yang sudah digunakan dan disediakan oleh jutaan pengguna dan sudah terbukti kelegalan dan keamanannya secara umum (Gideon, 2018). Dengan menggunakan aplikasi DANA, pengguna dapat melakukan bermacam transaksi cashless pada gerai – gerai online seperti membeli pulsa, membayar tagihan (internet, listrik, air, hingga BPJS), membayar cicilan, dan seterusnya yang berbasis online[5].

2.2 Pembobotan Kata TF - IDF

Pada tahap ini, guna mengubah data menjadi bentuk numerik digunakan pembobotan metode Term Frequency-Inverse Document Frequency atau TF-IDF [16]. Metode TF-IDF adalah metode untuk menentukan bagaimana kata (term) berhubungan dengan dokumen yang setiap katanya diberikan bobot. Metode TF-IDF ini menyatukan dua buah konsep yaitu frekuensi bermunculannya kata dalam sebuah dokumen dan frekuensi inverse-nya dokumen tersebut. Perhitungan nilai TF perkata dengan bobot setiap katanya 1 didahulukan saat menghitung bobot menggunakan TF-IDF [2]. Untuk mencari nilai TF dan IDF digunakan persamaan berikut.

$$TF = \begin{cases} 1 + \log_{10}(tf_{t,d}), & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \dots\dots\dots (1)$$

$$IDF(word) = \log \frac{D}{df} \dots\dots\dots (2)$$

Dimana TF menghitung kemunculan term (t) pada dokumen (d), jika tidak terdapat term atau t = 0 maka nilai TF menjadi 0. IDF(word) sendiri merupakan nilai IDF dari kata-kata yang dicari, D merupakan jumlah untuk semua dokumen yang ada, dan df merupakan jumlah kata yang terdapat pada semua dokumen. Perkalian dari perhitungan TF – IDF dapat menghasilkan bobot kata yang disebut TF – IDF dengan persamaan berikut:

$$W_{t,d} = TF \times IDF \dots\dots\dots (3)$$

2.3 Seleksi Fitur Chi Square

Seleksi fitur chi square digunakan dalam kategori dengan terms untuk pengamatan dari kebersesuaian (goodness of fit). Pengujian Chi Square didalam statistika diimplementasi untuk menguji independensi dari dua buah peristiwa. Sedangkan dalam teori statistika, dua peristiwa pada seleksi fitur tersebut diartanya adalah kemunculan dari fitur dan kemunculan dari kategori [8]. Chi square merupakan metode seleksi fitur supervised, dimana mampu mereduksi dan mengurangi fitur yang sudah dilabelkan pada kelas tertentu tanpa mengurangi akurasi[17]. Untuk persamaan dari chi square adalah sebagai berikut.

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i} \dots\dots\dots (4)$$

2.4 Support Vector Machine

Support vector machine (SVM) oleh Vapnik tahun 1992 diperkenalkan sebagai metode klasifikasi yang bagus digunakan dalam pemecahan masalah nonlinear. Digunakan sepasang input dan output data sebagai pelatihan dalam format tujuan yang diinginkan. Konsep SVM dengan mudah dapat dijelaskan sebagai upaya untuk menentukan hyperplane yang teroptimal yang berguna untuk memisah dua kelas data pada input space. SVM mencoba menentukan fungsi hyperplane yang dapat memaksimalkan jarak antara kelas. Dalam membentuk model model yang akan digunakan untuk proses klasifikasinya, support vector machine memilih beberapa titik data yang berkontribusi (support vector) [14]. Support vector machine memiliki persamaan sebagai berikut:

$$f(x) = w \cdot x + b \dots\dots\dots (5)$$

Atau

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \dots\dots\dots (6)$$

Terdapat kelebihan berupa kemudahan dalam kernel trick, karna untuk memilih support vector dalam proses pembelajaran SVM, pengguna cukup memahami fungsi dari kernel trick yang akan digunakan, fungsi non-linear wujudnya tidak perlu diketahui. Dari semua kernel trick yang tersedia, kernel trick yang menyampaikan hasil terbaik adalah kernel trick radial basis fuction. Proses klasifikasi kernel trick RBF mendapatkan hasil terbaik terutama untuk data yang tidak dapat dibedakan dengan cara linear [13].

Adapun kernel yang umum digunakan pada support vector machine adalah:

a. Polynomial

Kernel trick polynomial biasa digunakan dalam penyelesaian masalah yang terdapat pada klasifikasi, yang dimana dataset pelatihan sudah normal [13]. Persamaan dalam kernel trick ini adalah.

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1) \dots\dots\dots (7)$$

b. Radial Basis Fuction (RBF) atau Gaussian

Kernel trick radial basis fuction adalah kernel yang sering dipakai dalam penyelesaian masalah klasifikasi untuk dataset yang dipisah dengan cara linear, karena akurasi training dan prediction sangat baik menggunakan kernel RBF [13]. Kernel RBF memiliki persamaan sebagai berikut.

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma(\|\vec{x}_i - \vec{x}_j\|)^2) \dots\dots\dots (8)$$

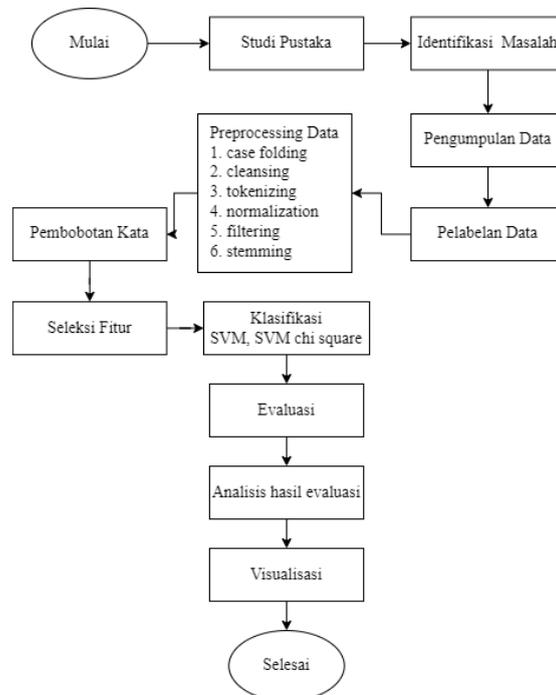
c. Sigmoid Kernel

Sigmoid adalah kernel trick SVM yang merupakan pengembangan dari jaringan saraf tiruan [13]. Kernel trick dinyatakan dalam persamaan.

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i \cdot \vec{x}_j + \beta) \dots\dots\dots (9)$$

3 Metodologi Penelitian

Pada penelitian ini, penulis membuat suatu tahapan penelitian, terdapat beberapa langkah Untuk melakukan proses analisis sentimen dengan menggunakan metode support vector machine untuk mencapai tujuan penelitian. Langkah-langkah dari penelitian ini adalah sebagai berikut:



Gambar 1. Alur Penelitian

Studi literatur. Tahap ini dilakukan dengan mencari informasi-informasi yang memuat tentang scraping, data mining, klasifikasi teks, dan penelitian lainnya yang membuat permasalahan pada analisis sentimen, data preprocessing, seleksi fitur, dan metode klasifikasi metode support vector machine yang berasal dari jurnal, e-book, buku-buku, dan lain-lain

Identifikasi Masalah. Pada tahap ini dilakukan dengan mengamati kondisi aplikasi DANA pada Google Play Store dan lainnya yang masih bersangkutan. Adapun masalah yang dapat diangkat adalah bagaimana membangun model klasifikasi dengan metode support vector machine dalam menganalisis sentimen pengguna aplikasi DANA berdasarkan ulasan.

Pengumpulan Data. Tahapan awal dari penelitian ini adalah pengumpulan data, dalam tahapan ini pengumpulan data dilakukan menggunakan teknik scraping pada Google Play Store. Scraping menggunakan bahasa pemrograman python dan memanfaatkan library google-play-scraper. Proses scraping data dilakukan pada tanggal 21 November 2021 dengan mengambil 1366 ulasan aplikasi DANA terbaru berbahasa indonesia.

Pelabelan Data. Setelah data terkumpul dengan menggunakan teknik scraping, dilanjutkan dengan Tahap pelabelan data untuk membagi data ulasan. Pelabelan dilakukan menggunakan 3 anator dan mengambil suara tertinggi dari ketiga anator tersebut. Data ulasan dibagi ke dalam dua kelas yaitu data yang bersifat positif dan data yang bersifat negatif

Data Preprocessing. Tahap data preprocessing penting dalam data training untuk membantu proses melatih algoritma agar dapat memodifikasi data yang tidak terorganisir menjadi terorganisasi, sehingga dapat mampu menyederhanakan pemrosesan data. Pengumpulan data opini dari ulasan pada google play store terkadang tidak sesuai dengan kata baku, memakai bahasa daerah, kata-kata tidak terdapat dalam kamus, atau disingkat. Untuk meminimalkan noise pada tahap selanjutnya, diperlukan preprocessing atau normalisasi untuk mengatasi masalah ini. Preprocessing terbagi dalam enam tahap, yaitu tahapannya sebagai berikut.

- 1) **Case Folding** : Case folding merupakan tahap dimana proses penyamaan case didalam sebuah dokumen. Tidak semua dokumen teks selalu selaras dalam menggunakan huruf kapital.
- 2) **Cleansing** : tahap cleansing bertujuan untuk menghilangkan noise yang tidak memiliki value atau seperti kata simbol, emoji, dan tanda baca yang tidak digunakan supaya dapat mempermudah dalam pemrosesan data.
- 3) **Tokenizing** : Tahapan Tokenizing berfungsi untuk memecah teks berdasarkan tiap kata yang menyusunnya, yang disebut sebagai term atau token. Tokenizing umumnya dipecah berdasarkan spasi.

- 4) **Normalization** : Tahap Normalization dilakukan agar kata-kata yang diperpanjang dan disingkat dapat menjadi kata kata yang sesuai menurut dengan Kamus Besar Bahasa Indonesia (KBBI).
- 5) **Filtering Stopword Removal** : Pada tahap filtering dokumen diproses agar kata-kata yang tidak penting dan tidak bermakna akan dihilangkan.
- 6) **Stemming** : Pada tahap stemming dilakukan usaha pemrosesan berbagai variasi kata ke dalam satu bentuk yaitu kata baku. Tahapan ini berguna untuk mengubah kata-kata yang terdapat imbuhan dan telah melewati seleksi menjadi kata dasar

Pembobotan Kata. Pembobotan kata bertujuan untuk memberikan bobot nilai pada setiap kata atau term yang dijadikan input pada proses klasifikasi karena model klasifikasi hanya menerima kata dalam bentuk numerik atau desimal. Pada penelitian ini dilakukan pembobotan kata dilakukan dengan menggunakan metode TF – IDF.

Seleksi Fitur. Proses ini dilakukan untuk meningkatkan akurasi dalam klasifikasi. Seleksi fitur sendiri merupakan proses mereduksi atau pengurangan fitur dalam suatu data. Pada penelitian ini seleksi fitur yang digunakan adalah *chi square*.

Klasifikasi. Kemudian dilanjutkan dengan pembagian data menjadi data training dan data testing. Terdapat 2 masukan dalam tahap klasifikasi yaitu, masukan tanpa seleksi fitur dan masukan dengan seleksi fitur *chi square*. Sebelum melakukan proses klasifikasi dilakukan pembagian data dengan pembagian rasio digunakan data training sebesar 80% dan data testing sebesar 20%. Penentuan parameter SVM memanfaatkan library GridSearchCV yang sudah terdapat pada python agar mendapatkan nilai akurasi yang teroptimal.

Evaluasi Model. Evaluasi hasil berguna untuk menganalisis hasil performa model klasifikasi yang selanjutnya akan dilakukan perbandingan terhadap model yang sudah dibuat berupa grafik dari perbandingan akurasi, presisi, dan recall dari model SVM dan SVM-*chi square*.

4 Hasil dan Pembahasan

Dataset yang diambil merupakan data ulasan dari aplikasi DANA yang diambil dari Google Play Store menggunakan teknik *scraping* dari *google-play-scraper*. *Google-play-scraper* ini menggunakan API yang sudah disediakan dimana API akan melakukan parsing HTML dan mengonversikannya kedalam file dengan format CSV yang memudahkan pengambilan data sebanyak 1366 ulasan yang digunakan dalam penelitian. Kemudian, data yang telah di-*scraping*, dilakukan pengonversian data ke dalam format CSV dengan nama file “review-dana-Nov”. Hasil pengonversian dapat dilihat pada Gambar 3 berikut.

	A	B	C	D	E	F	G	H	I	J	K
1	11/21/2021 13:14	5	dompet yg bagus dan anti ribet..								
2	11/21/2021 13:14	5	sangat bagus dan bikin aku semangat								
3	11/21/2021 13:11	5	Keren abis								
4	11/21/2021 13:10	5	orang yang bilang jelek fiks dia bodoh								
5	11/21/2021 13:09	1	saya transfer ke teman saya status pengiriman berhasil, saldo saya berkurang, tapi teman								
6	11/21/2021 13:06	4	Cukuk membantu								
7	11/21/2021 13:06	5	mantap								
8	11/21/2021 13:05	5	Kopet xxxxxxxx								
9	11/21/2021 13:04	4	Saya kasi bintang 4 dulu soal yah baru coba								
10	11/21/2021 13:04	1	Knp akun saya ngga bisa di buka Udh mah ada saldo nya lgi.. Trus pas mau di buka lgi mala								
11	11/21/2021 13:04	5	mantap								
12	11/21/2021 13:03	1	Saya mau tf ke bang lain gak bisa								
13	11/21/2021 13:01	2	Kh bintang 2 dlu, klo bagus tsmbah								
14	11/21/2021 13:01	5	Selalu evaluasi je								
15	11/21/2021 13:00	5	cepat dan aman bertransaksi dengan DANA. Top up pun mudah.								
16	11/21/2021 12:59	1	Tolong hentikan berlangganan Turbo vpn dari google Play.. Yg memotong saldo dana saya								
17	11/21/2021 12:57	1	Kenapa akun dana saya di bekukan apa bisa saya buka kembali akun dana saya?tolong di j								
18	11/21/2021 12:55	5	sangat sangat bergunaV"								
19	11/21/2021 12:54	4	avo download dana bisa berinov emas evus :D								

Gambar 2. Hasil pengonversian data aplikasi DANA

Data ulasan aplikasi DANA yang sudah diperoleh akan diteruskan ke tahap pelabelan data. Pada tabel

berikut ditunjukkan data yang sudah diambil dan akan melewati tahapan pelabelan data.

Tabel 1. Hasil *scraping* data ulasan

Ulasan
saya transfer ke teman saya status pengiriman berhasil, saldo saya berkurang, tapi teman saya tidak menerima transfer dari saya. sebenarnya klo aplikasi ewalet harusnya customer service nya selalu online. toh juga di wa hanya bisa jam kerja, itupun antrian banyak. harusnya ig, twitter, fb juga balas pesan saya.

Setelah melewati tahap pengumpulan data, dilanjutkan dengan pelabelan data. Pelabelan data dilakukan secara manual oleh 3 anatator yang melabelkan data secara masing – masing berdasarkan sudut pandang. Anatator menentukan data yang masuk kedalam kelas positif berupa kalimat yang mendukung atau kepuasan pengguna aplikasi DANA, dan data yang masuk kedalam kelas negatif berupa kalimat yang tidak mendukung, menjatuhkan, ketidak puasan pengguna aplikasi DANA

Case folding. *Case Folding* merupakan proses penyeragaman huruf yang dimana didalam proses tersebut hanya menerima bentuk huruf “a” hingga huruf “z”. penyeragaman ini dilakukan dengan mengubah semua huruf ke dalam huruf kecil upaya mencegah perbedaan arti jika terdeteksi huruf besar dan huruf kecil.

Tabel 2. Hasil Case folding

Input	Output
saya transfer ke teman saya status pengiriman berhasil, saldo saya berkurang, tapi teman saya tidak menerima transfer dari saya. sebenarnya klo aplikasi ewalet harusnya customer service nya selalu online. toh juga di wa hanya bisa jam kerja, itupun antrian banyak. harusnya ig, twitter, fb juga balas pesan saya.	saya transfer ke teman saya status pengiriman berhasil, saldo saya berkurang, tapi teman saya tidak menerima transfer dari saya. sebenarnya klo aplikasi ewalet harusnya customer service nya selalu online. toh juga di wa hanya bisa jam kerja, itupun antrian banyak. harusnya ig, twitter, fb juga balas pesan saya.

Cleansing. Proses yang dilakukan setelah data melalui tahap case folding adalah data cleansing. Pada tahapan ini dilakukan pembersihan dokumen dengan menghilangkan data yang tidak bermakna atau noise seperti tanda baca, simbol, spasi double, angka, dan lainnya. Penerapan data cleansing dapat dilihat pada tabel berikut

Tabel 3. Hasil Cleansing

Input	Output
saya transfer ke teman saya status pengiriman berhasil, saldo saya berkurang, tapi teman saya tidak menerima transfer dari saya. sebenarnya klo aplikasi ewalet harusnya customer	saya transfer ke teman saya status pengiriman berhasil saldo saya berkurang tapi teman saya tidak menerima transfer dari saya sebenarnya klo aplikasi ewalet harusnya customer

service nya selalu online. toh juga di wa hanya bisa jam kerja, itupun antrian banyak. harusnya ig, twitter, fb juga balas pesan saya.	service nya selalu online toh juga di wa hanya bisa jam kerja itupun antrian banyak harusnya ig twitter fb juga balas pesan saya
--	--

Tokenizing. Tahap tokenisasi memudahkan proses perhitungan kata – kata tersebut didalam dokumen ataupun untuk perhitungan frekuensi data tersebut muncul didalam corpus.

Tabel 4. Hasil Tokenizing

Input	Output
saya transfer ke teman saya status pengiriman berhasil saldo saya berkurang tapi teman saya tidak menerima transfer dari saya sebenarnya klo aplikasi ewalet harusnya customer service nya selalu online toh juga di wa hanya bisa jam kerja itupun antrian banyak harusnya ig twitter fb juga balas pesan saya	[saya, transfer, ke, teman, saya, status, pengiriman, berhasil, saldo, saya, berkurang, tapi, teman, saya, tidak, menerima, transfer, dari, saya, sebenarnya, klo, aplikasi, ewalet, harusnya, customer, service, nya, selalu, online, toh, juga, di, wa, hanya, bisa, jam, kerja, itupun, antrian, banyak, harusnya, ig, twitter, fb, juga, balas, pesan, saya]

Normalization. Tahapan selanjutnya adalah Normalization yang merupakan proses dimana dilakukan perubahan atau perbaikan kata-kata yang disingkat, typo, maupun salah ejaan yang akan diubah sesuai kamus yang sudah dibuat.

Tabel 5. Hasil Normalization

Input	Output
[saya, transfer, ke, teman, saya, status, pengiriman, berhasil, saldo, saya, berkurang, tapi, teman, saya, tidak, menerima, transfer, dari, saya, sebenarnya, klo, aplikasi, ewalet, harusnya, customer, service, nya, selalu, online, toh, juga, di, wa, hanya, bisa, jam, kerja, itupun, antrian, banyak, harusnya, ig, twitter, fb, juga, balas, pesan, saya]	[saya, kirim, ke, teman, saya, status, pengiriman, berhasil, saldo, saya, berkurang, tapi, teman, saya, tidak, menerima, kirim, dari, saya, benar, kalau, aplikasi, ewalet, harusnya, customer, service, nya, selalu, online, toh, juga, di, wa, hanya, bisa, jam, kerja, itupun, antrian, banyak, harusnya, ig, twitter, facebook, juga, balas, pesan, saya]

Filtering atau stopword removing. Tahap Filtering adalah proses dimana menghilangkan kata yang terdapat dalam dokumen atau pengurangan dimensi kata didalam corpus yang disebut sebagai stopwords.

Tabel 6. Hasil Filtering

Input	Output
[saya, kirim, ke, teman, saya, status, pengiriman, berhasil, saldo, saya, berkurang, tapi, teman, saya, tidak, menerima, kirim, dari, saya, benar, kalau, aplikasi, ewalet, harusnya, customer, service, nya, selalu,	[kirim, teman, status, pengiriman, berhasil, saldo, berkurang, teman, menerima, kirim, benar, aplikasi, customer, service, online, jam, kerja, antrian, balas, pesan]

online, toh, juga, di, wa, hanya, bisa, jam, kerja, itupun, antrian, banyak, harusnya, ig, twitter, facebook, juga, balas, pesan, saya]	
---	--

Stemming. Didalam proses Stemming dilakukan penghilangan beberapa kata pada ulasan yang terdapat imbuhan seperti contoh kata “mengirim” dikembalikan ke dalam kata dasar. Dalam tahapan ini digunakan library Indonesia yaitu library sastrawi python.

Tabel 7. Hasil Stemming

Input	Output
[kirim, teman, status, pengiriman, berhasil, saldo, berkurang, teman, menerima, kirim, benar, aplikasi, ewalet, customer, service, online, wa, jam, kerja, antrian, ig, twitter, facebook, balas, pesan]	[kirim, teman, status, kirim, hasil, saldo, kurang, teman, terima, kirim, benar, aplikasi, customer, service, online, jam, kerja, antri, balas, pesan]

Pembobotan Kata TF-IDF. Setelah data melewati tahapan preprocessing data, dilanjutkan dengan proses pembobotan kata. Proses ini dilakukan untuk memboboti setiap kata menjadi nilai fitur supaya dapat diproses oleh model machine learning seperti support vector machine yang hanya menerima data berupa angka. Tahapan ini dilakukan pembobotan dengan metode Term Frequency – Inverse Document Frequency (TF-IDF). Metode TF-IDF ini diawali dengan menghitung Term Frequeuncy (TF) terlebih dahulu didalam setiap dokumen dilanjutkan dengan Inverse Document Frequency (IDF) pada setiap dokumen menggunakan persamaan (10). Setelah mendapatkan TF dan IDF maka dilanjutkan pengkalian TF dan IDF supaya menghasilkan persamaan TF-IDF.

Seleksi Fitur Chi Square. Proses penyeleksian dalam chi square dilakukan dengan cara memilih fitur yang akan digunakan dan fitur yang tidak digunakan akan dihilangkan. Proses seleksi fitur chi square dilakukan dengan menguji nilai chi square suatu fitur. Jika nilai chi square suatu fitur tidak melebihi critical value maka fitur tersebut tidak akan digunakan dan hilangkan. Jika nilai chi square fitur tersebut melebihi critical value maka fitur akan digunakan didalam klasifikasi. Pengujian akan menggunakan tarif nyata α 0.90, 0.41, 0.32, 0.10, 0.050, 0.250, dan 0.010 dengan critical value sebesar 0.016, 0.679, 0.989, 2.706, 3.841, 5.204, dan 6.635. Hasil dari seleksi fitur chi square terhadap 1357 fitur yang digunakan dapat dilihat pada tabel berikut.

Tabel 8. Hasil seleksi fitur *chi square*

Critical Value	0.016	0.679	0.989	2.706	3.841	5.024	6.635
Total Fitur	1340	540	319	77	58	45	32

4.1 Klasifikasi SVM tanpa seleksi fitur

Data ulasan akan dibagi menjadi data training sebesar 80% dan data testing sebesar 20% untuk diteruskan ke tahap pemodelan. Pemodelan data menggunakan data training dari hasil pembagian data tanpa menggunakan seleksi fitur. Pemodelan dilakukan dengan memanfaatkan library python yaitu *SVC* dan *GridSearchCV*. Setelah melakukan pemodelan dengan data training, model akan diuji menggunakan data testing untuk mengetahui performa model dengan mengklasifikasikan data test.

Pada pemodelan *GridSearchCV*, parameter kernel dilakukan bertahap. Dengan kata lain pengujian dilakukan satu – persatu terhadap kernel yang digunakan. Pada penelitian ini kernel yang digunakan adalah kernel RBF, polynomial, dan linear. Pemodelan dilakukan dengan menggunakan *GridSearchCV* untuk menentukan parameter yang teroptimal. Pengujian parameter menggunakan parameter berikut :

- Kernel = rbf, polynomial, dan linear
- C = 0.5, 0.01, 0.1, 10, 50, 100
- Gamma = 1, 0.1, 0.01, 0.001

Hasil GridSearchCV untuk pemodelan tanpa seleksi fitur dapat dilihat pada tabel berikut.

Tabel 9. Hasil *GridSearchCV* SVM

Best Parameter			Akurasi
Kernel	C	Gamma	
linear	0.01	0.001	87.58%
	0.1	0.01	
	0.5		
poly	1	0.1	
	10		
rbf	50	1	
	100		

Pada tabel 9 menunjukkan hasil Grid Search CV untuk model SVM tanpa seleksi fitur chi square dengan menggunakan kernel rbf, nilai C = 0.5, dan nilai gamma = 0.01 mendapatkan akurasi sebesar 87.58%. Pada penelitian ini dilakukan pemodelan model SVM tanpa seleksi fitur menggunakan kernel rbf, nilai C = 0.5, dan nilai gamma = 1 terhadap data ulasan aplikasi DANA.

4.2 Klasifikasi SVM dengan seleksi fitur *chi square*

Pada pemodelan SVM menggunakan seleksi fitur chi square, penulis menggunakan fitur yang dimana nilai chi square suatu fitur diatas parameter critical value. Fitur yang diambil adalah fitur dengan nilai χ^2 diatas 0.016 sebanyak 1340 fitur, nilai chi square diatas 0.679 sebanyak 540 fitur, nilai chi square diatas 0.989 sebanyak 319 fitur, nilai chi square diatas 2.706 sebanyak 77 fitur, nilai chi square diatas 3.841 sebanyak 58 fitur, nilai chi square diatas 5.024 sebanyak 45 fitur, dan nilai chi square diatas 6.635 sebanyak 32 fitur. Setelah melakukan pemodelan dengan menggunakan seleksi fitur chi square selanjutnya akan diuji menggunakan data testing untuk mengetahui performa dengan mengklasifikasikan data test. Hasil GridSearchCV untuk pemodelan dengan seleksi fitur chi square dapat dilihat pada tabel berikut.

Tabel 10. Hasil *GridSearchCV* SVM *chi-square*

Critical Value	Jumlah Fitur	Best Parameter			Akurasi
		Kernel	C	Gamma	
0.016	1340	rbf	1	1	85,76%
0.679	540	rbf	1	1	89,41%
0.989	319	rbf	1	1	88,68%
2.706	77	rbf	50	0.1	84,30%
3.841	58	rbf	100	0.1	84,67%
5.024	45	rbf	10	1	83,57%
6.635	32	rbf	100	1	82,11%

Pada tabel 10 menunjukkan hasil Grid Search CV untuk model SVM dengan seleksi fitur chi square untuk critical value sebesar 0.679 menggunakan kernel rbf, nilai C = 1, dan nilai gamma = 1 mendapatkan akurasi sebesar 89.41%. Pada penelitian ini difokuskan kepada pemodelan model SVM dengan seleksi fitur chi square menggunakan critical value sebesar 0.679 terhadap data ulasan aplikasi DANA.

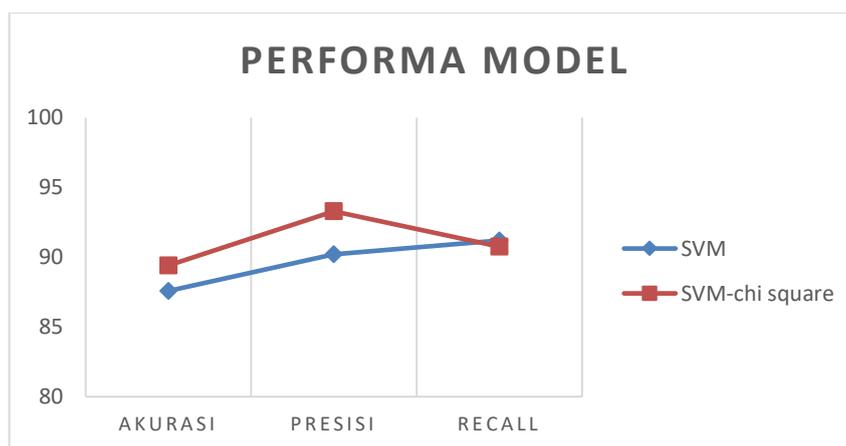
4.3 Evaluasi Model

Tahapan ini dilakukan analisis terhadap model yang sudah dibuat yaitu model support vector machine dan model support vector machine dengan seleksi fitur chi square dimana critical value yang digunakan sebesar 0.679 dengan tarif nyata α 0.41. Hasil pengujian terhadap model yang sudah dibuat dapat dilihat pada tabel berikut.

Tabel 11. Perbandingan hasil evaluasi model

Model		Jumlah Fitur	Performa		
			Akurasi	Presisi	Recall
SVM	-	1357	87,58%	90,21%	91,20%
SVM – chi square	0.679	540	89,41%	93,29%	90,76%

Pada kasus analisis sentimen ulasan aplikasi DANA, hasil performa pada model yang sudah dibuat mendapat peningkatan dan juga penurunan pada performa menggunakan seleksi fitur chi square dengan critical value sebesar 0.679. Dimana model SVM dengan seleksi fitur chi square mendapatkan kenaikan akurasi dari 87,58% menjadi 89,41%, kenaikan presisi dari 90,21% menjadi 93,21%, dan terdapat penurunan pada recall dari 91,20% menjadi 90,76%. Adapun grafik dari performa model yang telah dibuat adalah sebagai berikut.



Gambar 3. Grafik Performa Model

5 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan dari penelitian analisis sentimen aplikasi DANA pada Google Play store dapat diambil beberapa hal yang dapat disimpulkan.

1. Model klasifikasi dibangun setelah pengumpulan data ulasan pengguna aplikasi DANA pada Google Play Store yang diperoleh pada tanggal 21 November 2021 dilakukan pelabelan data secara manual dan didapatkan kelas positif yang lebih mendominasi dibanding kelas negatif dengan kelas positif sebanyak 883 dan kelas negatif sebanyak 483. Kemudian data melalui tahapan preprocessing untuk membersihkan data, data yang sudah bersih dilakukan pembobotan kata pada ulasan menggunakan Term Frequency – Inverse Document Frequency (TF – IDF) untuk dilakukan seleksi fitur menggunakan metode chi square. Pembagian data dibagi menjadi 80% data training dan 20% data testing, fitur yang terseleksi digunakan untuk melakukan data latih pembentukan model klasifikasi yang selanjutnya dilakukan pengujian model menggunakan metode Support Vector Machine..
2. Hasil evaluasi menunjukkan performa model klasifikasi SVM tanpa seleksi fitur yang teroptimal diperoleh dengan kernel rbf, nilai $C = 0.5$, dan nilai $\Gamma = 1$ dengan fitur sebanyak 1357 memperoleh akurasi sebesar 87,58%, presisi sebesar 90,21%, recall sebesar 91,20%. Sedangkan hasil evaluasi performa model klasifikasi

SVM dengan seleksi fitur chi square dengan kernel rbf, nilai $C = 1$, dan nilai $\text{Gamma} = 1$ dengan fitur sebanyak 540 memperoleh akurasi sebesar 89,41%, presisi sebesar 93,29%, dan recall sebesar 90,76%.

3. Terdapat peningkatan menggunakan seleksi fitur chi square dari akurasi model SVM tanpa menggunakan seleksi fitur sebesar 87,58% dan menggunakan seleksi fitur chi square menjadi sebesar 89,41% dengan nilai akurasi tersebut, nilai tarif nyata yang digunakan sebesar 0.41.

6 Daftar Pustaka

- [1] Abah, J. O., 2021. Sentiment Analysis of Amazon Electronic Product Reviews Using Deep Learning, Dublin: Dublin Business School.
- [2] Aprianto, Y., Nurhasanah, & Sanubary, I. (2018). Prediksi Kadar Particulate Matter (PM10) untuk Pemantauan Kualitas Udara Menggunakan Jaringan Syaraf Tiruan Studi Kasus Kota Pontianak. *POSITRON*.
- [3] Deolika, A., K. & Lutfi, E. T., 2019. ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING. *Jurnal Teknologi Informasi*, pp. 179-184.
- [4] Fitriyah, N., Warsito, B. & Maruddani, D. I., 2020. ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM). *Jurnal Gaussian*, 9(3), pp. 376 - 390.
- [5] Gideon, A., 2018. Hal yang Perlu Diketahui Soal DANA, Dompot Digital Besutan Anak Negeri. [Online] Available at: <https://www.liputan6.com/bisnis/read/3802191/hal-yang-perlu-diketahui-soal-dana-dompot-digital-besutan-anak-negeri>
- [6] Chang, Z., Zhang, Y., & Chen, W. (2018). Effective Adam-optimized LSTM Neural Network for Electricity Price Forecasting. *IEEE*, 245.
- [7] Harfian, Y., 2021. KLASIFIKASI SENTIMEN APLIKASI DOMPET DIGITAL DANA PADA KOMENTAR DI INSTAGRAM MENGGUNAKAN NAIVE BAYES CLASSIFIER. *Teknik Informatika*.
- [8] Graves, A. (2012). *Supervised Sequence Labell with RNN*. Springer.
- [9] Huda, H. & Islahudin, N., 2021. MEASURMENT SYSTEM ANALYSIS PADA OPERATOR PENGECEKAN VISUAL MENGGUNAKAN METODE. *Jurnal Integrasi Sistem Industri*, 8(2), pp. 35-40
- [10] Irfani, F. F., Triyanto, M. & K., 2020. Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine. *Jurnal Bisnis, Manajemen, dan Informatika*, pp. 258-266.
- [11] Juen Ling, ., I. P. E. N. K. ., T. B. O., 2014. E-Jurnal Matematika. ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES, 3(3), pp. 92-99.
- [12] Lukmana, D. T., S. S. & Y. S., 2019. Analisis Sentimen Terhadap Calon Presiden 2019. *Seminar Nasional Penelitian Pendidikan Matematika*, pp. 154-160.
- [13] Mahendrajaya, R., Buntoro, G. A. & Setyawan, M. B., 2019. Analisi Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based dan Support Vector Machine. *Jurnal Teknik Universitas Muhammadiyah Ponogoro*, pp. 52-63.
- [14] Mohri, M., Rostamizadeh, A. & A. T., 2012. *Foundations of Machine Learning*. s.l.:MTI Press.
- [15] Nuansa Gumilang, Z. A., 2018. IMPLEMENTASI NAIVE BAYER CLASSIFIER DAN ASOSIASI UNTUK ANALISIS SENTIMEN DATA ULASAN APLIKASI E-COMMERCE SHOPEE PADA SITUS GOOGLE PLAY. *Statistika*.
- [16] Praptiwi, D. Y., 2018. ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA E-COMMERCE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN MAXIMUM ENTROPY. *Statistika*.
- [17] Rofiqoh, U., Perdana, R. S. & Fauzi, M. A., 2017. ANALISIS SENTIMEN TINGKAT KEPUASAN PENGGUNA PENYEDIA LAYANAN TELEKOMUNIKASI SELULER INDONESIA PADA TWITTER DENGAN METODE SUPPORT VECTOR MACHINE DAN LEXICON BASED FEATURES. *Jurnal pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(22), pp. 1725-1732.
- [18] Romadoni, F., Umaidah, Y. & Sari, B. N., 2020. Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal SISFOKOM*, pp. 247-253.
- [19] Sitanayah Que, V. K., Iriani, A. & Purnomo, H. D., 2020. Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, pp. 162-170.
- [20] Somantri, O. & Apriliani, D., 2018. SUPPORT VECTOR MACHINE BERBASIS FEATURE SELECTION UNTUK. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), pp. 537-548.
- [21] Suntoro, J., 2019. DATA MINING ALGORTIME DAN IMPLEMENTASI MENGGUNAKAN BAHASA PEMROGRAMAN PHP. Semarang: s.n.
- [22] Surohman, Aji, S., Rousyati & Wati, F. F., 2020. Analisis Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier dan K-Nearest Neighbor. *Jurnal Sains dan Manajemen*, 8(1), pp. 2657 - 0793.
- [23] W., 2016. *Data Mining Practical Machine Learning Tools and Techniques*. United States of America: s.n..