

Analisis Sentimen Pada *Review* Pengguna *E-Commerce* Bidang Pangan Menggunakan Metode *Support Vector Machine* (Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)

Geyessella Manik¹, Iin Ernawati², Ika Nurlaili³

Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
geyessella.manik@gmail.com¹, iin_ernawati@yahoo.com², nurlailika@upnvj.ac.id³

Abstrak. Sayurbox dan Tanihub merupakan *e-commerce* penyedia hasil pertanian yang dapat diakses dari situs *Google Play*. Pada *Google play* terdapat fitur ulasan yang tentunya akan berpengaruh terhadap calon pengguna karena sebelum mengunduh suatu aplikasi, calon pengguna biasanya akan melihat kolom ulasan sebagai tolak ukur efektivitas suatu produk. Untuk mendapatkan informasi apakah layanan yang diberikan termasuk ke dalam kelas positif atau negatif digunakan tahapan analisis sentimen yaitu pembersihan data dari kata dan simbol yang tidak relevan dengan sentimen lalu melabelinya, dan merubah data yang bernilai kualitatif menjadi data bernilai kuantitatif dengan bantuan *Term Frequency-Invers Document Frequency* (TF IDF) kemudian melakukan klasifikasi dengan metode *Support Vector Machine* dengan kernel linear pada ulasan sayurbox dan tanihub dengan jangka Februari 2020 hingga Januari 2021. Akurasi tertinggi yang didapatkan yaitu pada data sayurbox sebesar 91,4% dengan jumlah sentimen terbanyak adalah sentimen positif sebanyak 738 (70%). Pada data *review* Tanihub, total akurasi sebesar 88,8% dengan sentimen terbanyak yaitu sentimen positif berjumlah 348 (65%).

Kata Kunci: Analisis Sentimen, *E-commerce*, *Support Vector Machine*

1 Pendahuluan

Pemakaian internet dari tahun ke tahun meningkat secara pesat. Pada tahun 2020, peningkatan pengguna internet pun melonjak begitu cepat semenjak kemunculan coronavirus. Coronavirus merupakan penyakit yang dapat berpindah dari satu orang ke orang lainnya. Seseorang yang sudah tertular dapat ditandai dengan gangguan pernafasan, penyakit ini pertama kali ditemukan pada Desember 2019 di Wuhan, Cina. Dalam rangka menghentikan penyebaran, pemerintah Indonesia memberikan beberapa kebijakan *social distancing* seperti bekerja dan belajar dari rumah. Menurut Lembaga Survei Alvara Research Center, kehadiran penyakit coronavirus membuat pengguna internet meningkat cukup signifikan dan berdasarkan data survey yang diselenggarakan, salah satu kegiatan yang sering dilakukan di internet selama pandemi coronavirus adalah belanja *online*. Hasil survei yang dilakukan Alvara Research Center menunjukkan bahwa pengeluaran belanja masyarakat atas kebutuhan internet pada tahun 2020 bertambah 2 persen yaitu 8,1 persen, jika dibandingkan tahun lalu yang hanya sebesar 6,1 persen [1].

Kebijakan pemerintah Indonesia untuk penerapan jaga jarak dan membatasi aktivitas di luar rumah mendorong masyarakat lebih cenderung untuk berbelanja secara online guna memenuhi beberapa kebutuhan seperti kebutuhan pangan yang tentunya menyebabkan kenaikan pemesanan *e-commerce* serta memberi peluang besar bagi *startup digital business* yang bergerak pada bidang pangan untuk menggunakan layanan *e-commerce*. Menurut survei yang dilakukan dailysocial.id mengenai aplikasi belanja *online* terpopuler selama pandemi coronavirus, beberapa diantaranya ialah Sayurbox(31%) dan Tanihub(23%) yang masuk kedalam aplikasi populer yang digunakan untuk memenuhi kebutuhan (bahan makanan atau sebagainya) sehari-hari. Sayurbox dan Tanihub adalah aplikasi *e-commerce* yang bergerak pada bidang pangan dengan memasarkan hasil pertanian seperti sayur dan buah kepada masyarakat yang dapat dibeli secara *online* dengan mengunduhnya lewat situs *google play* [2].

Google Play merupakan layanan konten digital, dimana kita dapat mengunduh aplikasi maupun produk *online* lainnya secara gratis atau berbayar. *Google Play* dikembangkan dan diluncurkan oleh *Google* pada 6 Maret 2012 [3]. Pada *Google play* terdapat fitur rating dan ulasan (*review*), dimana fitur tersebut akan sangat berpengaruh terhadap calon pengguna karena sebelum mengunduh suatu aplikasi, calon pengguna biasanya akan melihat kolom

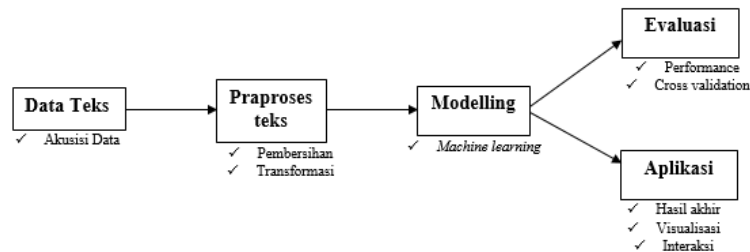
ulasan pengguna aplikasi sebelumnya sebagai tolak ukur efektivitas suatu produk. Opini-opini yang dituangkan tentunya juga akan mempengaruhi citra dari perusahaan.

Dari latar belakang yang diuraikan seiringan dengan peningkatan penggunaan *e-commerce* bidang pangan, peneliti merasa perlu untuk menganalisa lebih lanjut terhadap *review* atau ulasan pengguna *e-commerce* pada dua aplikasi yang dipilih yaitu Sayurbox dan aplikasi Tanihub untuk mengetahui bagaimana opini atau pendapat dari para penggunanya terhadap layanan yang diberikan oleh dua perusahaan tersebut. Salah satu cara untuk melakukan analisa berdasarkan opini pengguna adalah dengan memanfaatkan bidang *text mining* yaitu analisis sentimen yang dapat dipakai untuk melihat pembicaraan apa yang sering dibahas publik. Analisis sentimen biasanya digunakan untuk mengelompokkan tanggapan konsumen yang berupa positif dan negatif, sehingga memudahkan tugas dari pemilik bisnis untuk meninjau beberapa kekurangan produk atau layanan mereka. Dalam melakukan analisis sentimen dibutuhkan juga sebuah metode klasifikasi dengan memanfaatkan algoritma *machine learning* sebagai prosedur klasifikasi. Salah satu algoritma yang berfungsi untuk melakukan pengklasifikasian setiap opini pada data ulasan menjadi suatu sentimen positif dan negatif ialah *Support Vector Machine* (SVM). Pemilihan metode SVM karena algoritma tersebut memiliki akurasi tertinggi dalam hal pengklasifikasian data teks [4]. Untuk membantu pada proses algoritma tersebut diperlukan juga beberapa tahap seperti praproses data teks guna menyiapkan data menjadi data yang lebih bersih dan terstruktur.

2 Landasan Teori

2.1 Text Mining

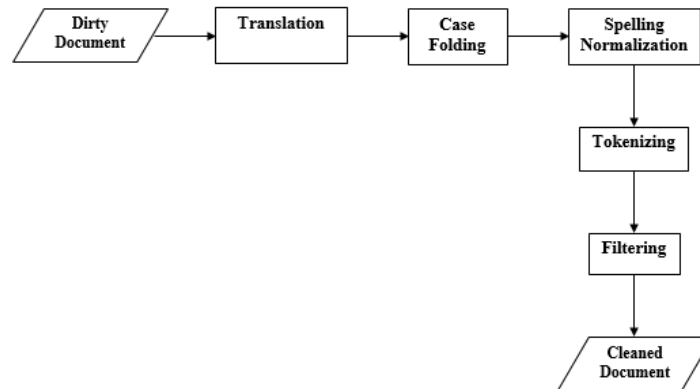
Text mining merupakan penerapan konsep dan teknik data mining atau penggalian data pada dokumen yang berbentuk teks dengan format yang sembarang atau tidak terstruktur. Penggalian teks berguna untuk mencari pola berupa informasi dan pengetahuan yang bermanfaat untuk tujuan tertentu [5]. *Text mining* sendiri telah diterapkan dalam berbagai bidang kebutuhan, salah satunya adalah Analisis sentimen. Analisis sentimen atau *Opinion mining* merupakan cabang dari *Text Mining* yang sering digunakan dalam tahap pengklasifikasian suatu data teks berupa tanggapan berdasarkan sentimen yang diberikan publik. Analisis sentimen adalah proses analisa dalam melakukan identifikasi dan mengelompokkan tanggapan atau opini dari suatu teks yang berguna untuk menentukan apakah opini tersebut berupa opini positif, negatif, atau netral.



Gambar. 1. Proses umum *Text Mining*

2.2 Text Preprocessing

Dalam *text mining*, tahap praproses teks atau disebut juga dengan *text preprocessing* sangat perlu guna membersihkan dan menyiapkan data dari data yang tidak relevan. Data mentah yang digunakan pada proses text mining biasanya tidak dalam kondisi yang ideal untuk diproses, sehingga dibutuhkan praproses teks untuk memperbaiki data dokumen teks yang sembarang atau tidak terstruktur menjadi data yang lebih terstruktur. Adapun tahap yang dilakukan dalam praproses data adalah sebagai berikut :



Gambar. 2. Tahap praproses Teks

2.2.1 Translation (Penerjemahan kata dan simbol)

Translation yaitu tahapan menerjemahkan data *review* yang berbahasa Inggris menjadi Indonesia, dan juga mengubah beberapa simbol dan emoji dengan bantuan kamus emoticon yang telah dilist, seperti emoji ‘☺’ akan diubah menjadi kata ‘senang’.

2.2.2 Case Folding

Case folding berguna untuk mengubah semua huruf yang ada pada dokumen teks menjadi huruf kecil (*lowercase*). Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, tahapan *case folding* digunakan untuk melakukan penyeragaman keseluruhan teks kata yang terdapat pada dokumen.

2.2.3 Spelling Normalization

Spelling normalization atau normalisasi adalah proses memperbaiki kata - kata pada dokumen teks seperti salah eja, disingkat dalam bentuk tertentu, dan kata - kata gaul yang tidak sesuai dengan KBBI dengan bantuan kamus *slangword*. Kamus *slangword* adalah kamus yang berisi daftar kata gaul, salah eja, serta list perbaikan kata dari kata gaul dan salah eja tersebut. Jika ditemukan adanya kata gaul, dan salah eja maka algoritma program akan mengubah kata tersebut sesuai dengan isi kamus *slangword*. Kamus ini didapat dari Github dengan judul Prosentilyzer – Product Review Sentiment Analyzer oleh Christian Wibisono yang dilakukan pengeditan ulang.

2.2.4 Tokenizing

Tokenizing merupakan proses untuk memenggal setiap kata dalam kalimat termasuk karakter atau bisa juga disebut dengan proses potong teks input berdasarkan tiap kata yang menyusunnya. Hasil proses *tokenizing* adalah deretan token (kata) dari proses pemotongan teks input.

2.2.5 Filtering

Filtering atau filtrasi adalah proses pengambilan kata - kata penting dari token hasil dari proses *tokenizing*. Proses pengambilan kata yang penting dapat menggunakan algoritma *stop list* (buang kata yang dianggap kurang penting) atau *word list* (simpan kata yang dianggap penting) dengan bantuan kamus stopword Indonesia.

2.3 Ekstraksi Fitur

Ekstraksi Fitur berguna untuk memberikan nilai bobot kata pada data dokumen, nilai tersebut nantinya akan digunakan dalam tahap klasifikasi. Adapun ekstraksi fitur yang digunakan adalah dengan memanfaatkan

pembobotan *Frequency-Invers Document Frequency* (TF IDF) dengan memberi nilai pada data ulasan. Dalam ekstraksi TF IDF, ada dua hal yang harus diperhatikan dalam perhitungan bobot antara lain TF dan IDF [4].

2.3.1 Term Frequency (TF)

Term Frequency (TF) merupakan frekuensi kemunculan sebuah term yang dapat berupa kata, frase atau unit hasil indexing lainnya dalam dokumen. TF merupakan faktor penentu bobot term, nilai frekuensi kemunculan kata dihitung dengan memberi bobot terhadap suatu kata. Semakin tinggi nilai kemunculan kata, maka semakin besar bobotnya pada dokumen (nilai kesesuaian semakin besar). Adapun perhitungan TF yaitu :

$$tf = 1 + \log \log (tf) \quad (1)$$

2.3.2 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) memperlihatkan hubungan ketersediaan kata dalam dokumen. Semakin rendah jumlah dokumen yang mengandung term atau kata, maka IDF akan semakin besar. Hal ini dikarenakan term dengan kemunculan tertinggi dapat dianggap term umum sehingga tidak penting nilainya. Adapun perhitungan IDF yaitu :

$$IDF = \log \left(\frac{D}{df_t} \right) \quad (2)$$

Dimana :

D : Jumlah semua dokumen

df_t : Total keseluruhan dokumen yang mengandung *term* (t)

Perkalian dari hitungan TF dan IDF akan menghasilkan bobot kata (TF IDF) dengan rumus berikut :

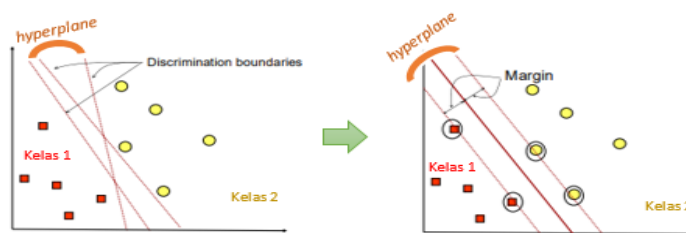
$$W_{t,d} = TF \times IDF \quad (3)$$

Dimana : $W_{t,d}$ adalah bobot dari kata t pada dokumen d.

2.4 Klasifikasi SVM

Klasifikasi SVM (*Support Vector Machine*) adalah salah satu klasifikasi *supervised learning* digunakan untuk analisa data dalam pengenalan pola yang digunakan untuk klasifikasi yang pertama kali diperkenalkan oleh Vapnik pada tahun 1992. SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM). SRM sendiri bertujuan untuk menjamin batas atas dari generalisasi pada data pengujian dengan cara mengontrol "kapasitas" (fleksibilitas) dari hipotesis hasil pembelajaran. Ada tiga hal yang harus diperhatikan dalam menggunakan algoritma SVM yaitu :

1. *Support Vectors*, yaitu kelas kelas yang ada pada saat melakukan klasifikasi.
2. *Hyperplane*, adalah sebuah fungsi untuk memisahkan kelas dengan kelas lainnya.
3. *Margin*, yaitu jarak pada kelas pertama terhadap kelas kedua.



(Sumber : Nugroho, A. S dkk, 2003)

Gambar 3. SVM pada kelas yang linear

2.5 Ketepatan Klasifikasi

Pentingnya melakukan pengukuran ketepatan klasifikasi yaitu untuk melihat seberapa baik model pengklasifikasian yang dilakukan dengan penggunaan *confusion matrix* yang mengukur ketepatan dengan melihat

nilai pada tiap kelas prediksi dan kelas aktual. Tabel matriks adalah bentuk dari *confusion matrix* yang mencatat hasil dari kinerja pemodelan klasifikasi pada data testing yang nilai sebenarnya diketahui [6].

Tabel 1. Model *Confusion Matrix*

Prediksi	Kelas Aktual	
	Negatif	Positif
Negatif	TN (True Negatif)	FN (False Negatif)
Positif	FP (False Positif)	TP (True Positif)

Pengukuran yang umum dilakukan untuk evaluasi yaitu :

1. Akurasi, menggambarkan seberapa tepat model melakukan klasifikasi. Dengan formula :

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. Presisi, menggambarkan tingkat ketepatan dari data yang diminta dengan hasil prediksi yang diberikan oleh model. Dengan formula :

$$presisi = \frac{TP}{TP+FP} \quad (5)$$

3. Recall atau Sensitivity, menggambarkan keberhasilan model dalam menemukan informasi. Dengan formula :

$$recall = \frac{TP}{TP+FN} \quad (6)$$

3 Hasil dan Pembahasan

Data yang digunakan pada penelitian ini berupa ulasan pengguna aplikasi belanja *online* yang diambil melalui dua aplikasi yaitu Sayurbox dan Tanihub. Proses pengumpulan data dilakukan dengan bantuan teknik *scraping* menggunakan *python* dengan dua modul pendukung yaitu *Selenium* dan *BeautifulSoup*. *Selenium* sendiri digunakan untuk otomasi browser dengan melakukan import *webdriver*, sedangkan *BeautifulSoup* adalah *library* yang digunakan untuk mempermudah dalam mengekstraksi informasi dari halaman web yang datanya ingin ditarik. Adapun alamat web yang digunakan pada proses *scraping* berasal dari aplikasi *e-commerce* Sayurbox dengan alamat <https://play.google.com/store/apps/details?id=com.sayurbox&showAllReviews=true> dan aplikasi *e-commerce* Tanihub dengan alamat <https://play.google.com/store/apps/details?id=com.tanihub.vaesdothrak&showAllReviews=true>.

Data yang berhasil ditarik adalah ulasan pengguna dengan rentang waktu Februari 2020 hingga Januari 2021.

Unnamed: 0	rating	tanggal	ulasan
0	0	4 2021-01-31	Kenapa sih akhir2 ini lelet banget aplikasinya...
1	1	5 2021-01-31	Sayur dan buahnya selalu fresh, pembayarannya ...
2	2	1 2021-01-31	Jual alpukat mahal, berulat pula
3	3	5 2021-01-31	Sudah hampir setengah tahun belanja disini, ap...

Gambar. 4. Data mentah

Dari penarikan data menggunakan teknik *scraping* didapatkan hasil ulasan sebanyak 1.051 ulasan pengguna aplikasi Sayurbox dan 533 ulasan dari pengguna aplikasi Tanihub.

Tabel 2. Sampel Data

Ulasan (D)	Sampel data sebelum praproses teks	Sampel data sesudah praproses teks	Sentiment
D1	Sayurnya bagus2, pengemasannya sangat baik, semua barang yang dibeli tidak ada yg kurang."	bagus sangat baik dibeli tidak kurang	Positif
D2	Kualitas bagus dan packing rapi.	kualitas bagus rapi	Positif
D3	Sayur dan buahnya selalu segar, pembayarannya juga sangat mudah."	selalu segar sangat mudah	Positif

Setelah dilakukannya praproses teks dan pelabelan kelas sentiment, maka dilakukan pembagian data. Data latih dipakai untuk melatih suatu algoritma, sedangkan data uji sendiri adalah dataset yang digunakan untuk mengetahui tingkat kinerja dari metode yang telah dilatih sebelumnya ketika mendapatkan data baru yang belum pernah ada.

Adapun pembagian data *training* dan *testing* yang dipakai yaitu menggunakan perbandingan 80:20. Meskipun belum ada penelitian mengenai pemilihan rasio yang optimal, ada beberapa praktik umum dalam memilih ukuran data. Berdasarkan Pareto Principle, rasio yang umum dipakai dalam pembagian data yaitu menggunakan perbandingan 80% : 20% [4].

Tabel 3. Pembagian data *review* sayurbox

Label	Train (latih) 80%	Test (uji) 20%	Jumlah
Positif	581	157	738
Negatif	256	53	309
Jumlah	837	210	1.047

Tabel 4. Pembagian data *review* tanihub

Label	Train (latih) 80%	Test (uji) 20%	Jumlah
Positif	257	73	330
Negatif	150	34	184
Jumlah	407	107	514

Hasil pembagian data akan dilakukan ekstraksi TF IDF untuk mengubah data *review* yang masih mengandung nilai kualitatif menjadi nilai kuantitatif menggunakan metode *Term Frequency-Invers Document Frequency* dengan fungsi `from sklearn.feature_extraction.text import TfidfVectorizer`. Beberapa kata yang didapat dari tiga dokumen hasil pembersihan data pada sampel data yaitu : bagus, sangat , baik, dibeli, tidak, kurang, kualitas, rapi, selalu, segar, mudah.

Perhitungan TF IDF yang dilakukan didapatkan berdasarkan rumus (3) melalui perkalian rumus (1) dan rumus (2). Berikut tabel sampel nilai TF IDF dari tiga kata yang dipilih yaitu bagus, sangat, kualitas :

Tabel 5. Hasil perhitungan TF IDF

Kata	Tf			df	IDF	w		
	D1	D2	D3			D1	D2	D3
Bagus	1	1	0	2	0,176	0,176	0,176	0
Sangat	1	0	1	2	0,176	0,176	0	0,176
Kualitas	0	1	0	1	0,477	0	0,176	0

Adapun pengklasifikasian yang dilakukan menggunakan metode *Support Vector Machine* (SVM) dengan perbandingan pembagian data yaitu 80% data *training* dan 20% data *testing* yang dapat dilihat pada Tabel 3 dan Tabel 4. Data *training* positif dan data *training* negatif akan digunakan oleh algoritma SVM untuk mempelajari pola data sesuai ciri dari data pada tiap kelas. Untuk mengenali data, kelas akan dipisahkan dengan memanfaatkan satu garis linear dalam ruang input menggunakan kernel *linear* dengan bantuan *sklearn library* pada *python* [7].

Tabel 6. *Confusion matrix* klasifikasi SVM pada sayurbox data

Prediksi	Sayurbox		Precision
	<i>Actual Negative</i>	<i>Actual Positive</i>	
Negatif	57	8	0,85
Positif	10	135	0,94
Recall	0,88	0,93	
Akurasi	91,4% (0,91)		

Tabel 7. *Confusion matrix* klasifikasi SVM pada tanihub data

Prediksi	Tanihub		Precision
	<i>Actual Negative</i>	<i>Actual Positive</i>	
Negatif	26	6	0,81
Positif	6	69	0,92
Recall	0,81	0,92	
Akurasi	88,8% (0,89)		

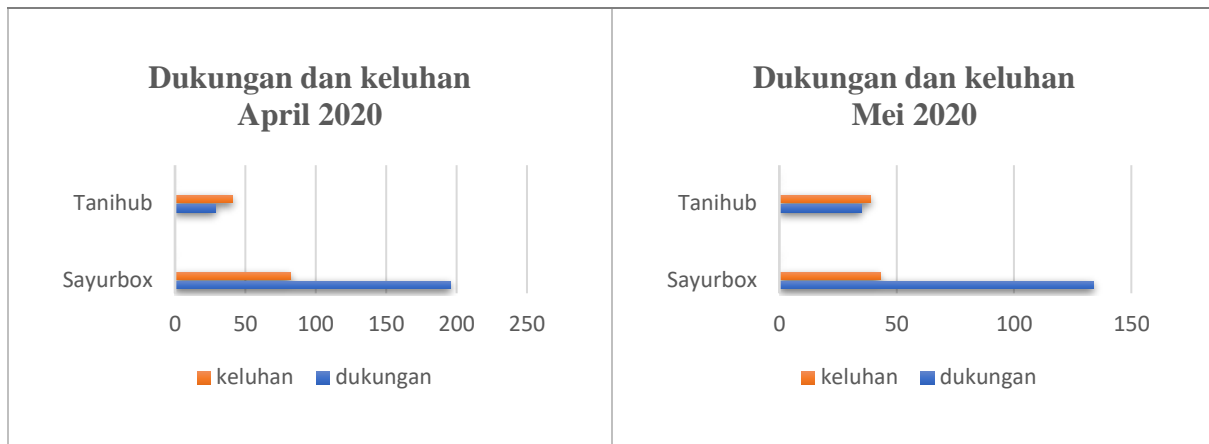
True Negatif (TN), adalah data negatif dengan prediksi yang tepat dengan data sebenarnya. Pada data sayurbox, nilai TN sebanyak 57, dan pada data tanihub yaitu 26. *False Negatif* (FN) atau biasa disebut dengan *Type Error 1*, yaitu data positif dengan prediksi yang tidak sesuai (prediksi negatif). Pada data sayurbox, nilai FN sebanyak 8, dan pada data tanihub yaitu 6. *False Positif* (FP) atau biasa disebut dengan *Type Error 2*, yaitu data negatif dengan prediksi yang tidak sesuai (prediksi positif). Pada data sayurbox, nilai FP sebanyak 10, dan pada data tanihub yaitu 6. *True Positif* (TP), adalah data positif yang diprediksi tepat sesuai data sebenarnya. Pada data sayurbox, nilai TP sebanyak 135, dan pada data tanihub yaitu 69.

Tiga perhitungan yang digunakan dalam *confusion matrix* dengan penggunaan data uji pada Sayurbox = 210 data, dan data uji pada Tanihub = 107 data. Pengukuran yang digunakan pada tabel *confusion matrix* yaitu akurasi, presisi, dan recall yang dapat dilihat pada rumus nomor (4) perhitungan akurasi, rumus nomor (5) perhitungan presisi, dan rumus nomor (6) yaitu perhitungan recall.

Untuk hasil dari data review Sayurbox didapatkan : Akurasi = $(135 + 57) / (135 + 57 + 10 + 8) = 0,914$ = 91,4% Presisi = $(135) / (135 + 10) = 0,931 = 93,1\%$ Recall = $(135) / (135 + 8) = 0,944 = 94,4\%$	Untuk hasil dari data review Tanihub didapatkan : Akurasi = $(69 + 26) / (69 + 26 + 6 + 6)$ = 0,887 = 88,8% Presisi = $(69) / (69 + 6) = 0,92 = 92\%$ Recall = $(69) / (69 + 6) = 0,92 = 92\%$
---	--

Dari data dokumen ulasan diperoleh bahwa ulasan terbanyak jatuh pada bulan April dan bulan Mei yang diperkirakan meningkat karena pada dua bulan tersebut adanya kebijakan pemerintah yaitu PSBB (Pembatasan Sosial Berskala Besar) untuk mencegah penyebaran coronavirus, akibatnya banyak masyarakat yang takut untuk

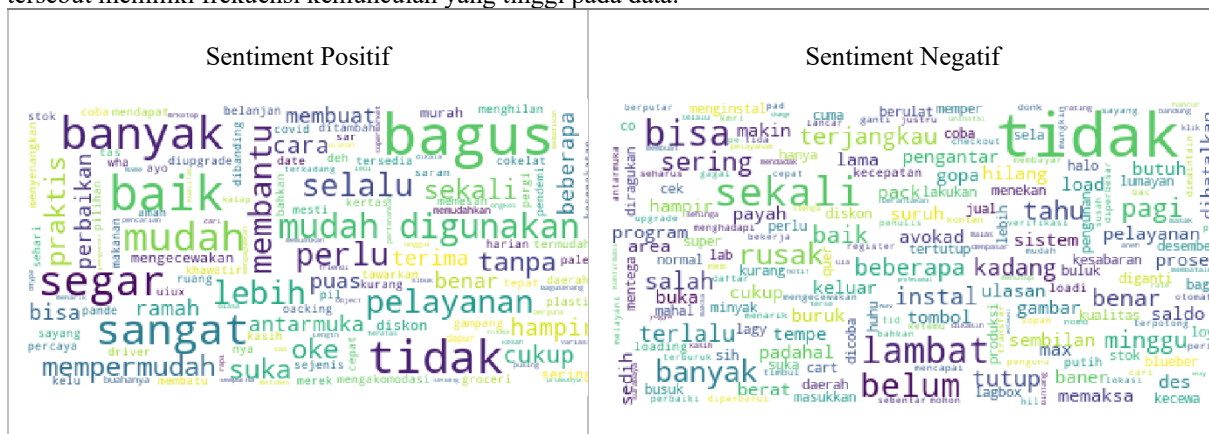
melakukan aktivitas diluar rumah dan lebih memilih membeli kebutuhan pokok secara *online*. Berikut hasil klasifikasi sentiment dukungan dan keluhan pada bulan April dan Mei sebahai bulan penggunaan layanan tertinggi :



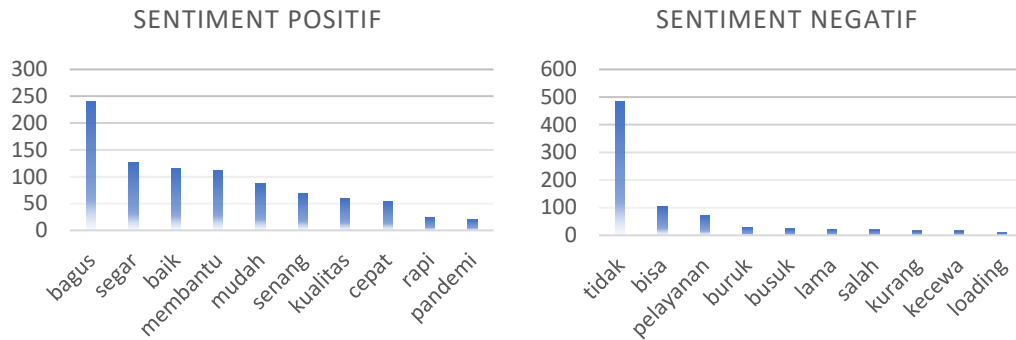
Gambar. 5. Hasil analisa dukungan dan keluhan pada bulan april dan mei

Total ulasan yang diberikan pengguna aplikasi sayurbox lebih banyak dibanding ulasan yang diberikan pengguna aplikasi tanihub. Pada bulan April, jumlah total keseluruhan ulasan data sayurbox sebanyak 278 ulasan, dengan dukungan sebanyak 196 (71%) dan keluhan sebanyak 82 (29%). Sedangkan pada bulan Mei, total ulasan yang didapat pada aplikasi sayurbox sebanyak 177 ulasan, dengan dukungan sebanyak 134 (76%) dan 43 keluhan (24%). Berbeda dengan sayurbox, pada data tanihub di bulan april mendapatkan ulasan sebanyak 70 ulasan yang berisi 29 dukungan (41%) dan 41 keluhan (59%). Selanjutnya pada bulan Mei, tanihub mendapatkan kenaikan ulasan yang berisi 35 dukungan (47%) dan keluhan sebanyak 39 (53%) dengan total keseluruhan yaitu 74 ulasan.

Untuk mengetahui gambaran mengenai topik yang sering dibicarakan dalam kolom ulasan, maka dilakukan visualisasi pada masing-masing klasifikasi kelas sentiment dengan memanfaatkan visualisasi *wordcloud* untuk mengetahui kata yang paling sering muncul, dan visualisasi *barplot* untuk menampilkan frekuensi kata yang sering muncul pada dokumen review. Pada *wordcloud*, ukuran *font* pada kata mempresentasikan seringnya kata yang muncul pada data teks (frekuensi kemunculan kata), semakin besar ukuran kata yang ditampilkan, artinya kata tersebut memiliki frekuensi kemunculan yang tinggi pada data.



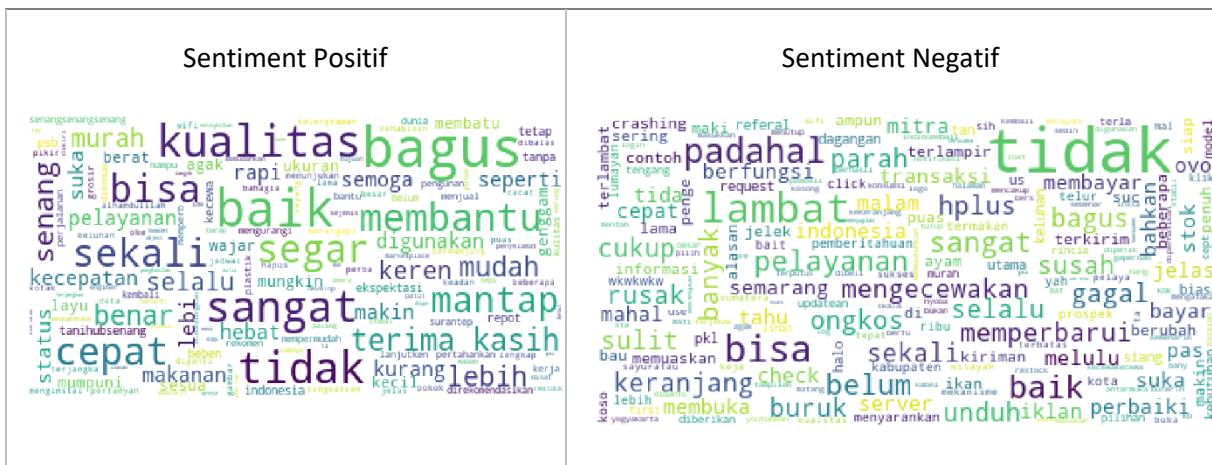
Gambar 6. Visualisasi *wordcloud* pada *review* sayurbox



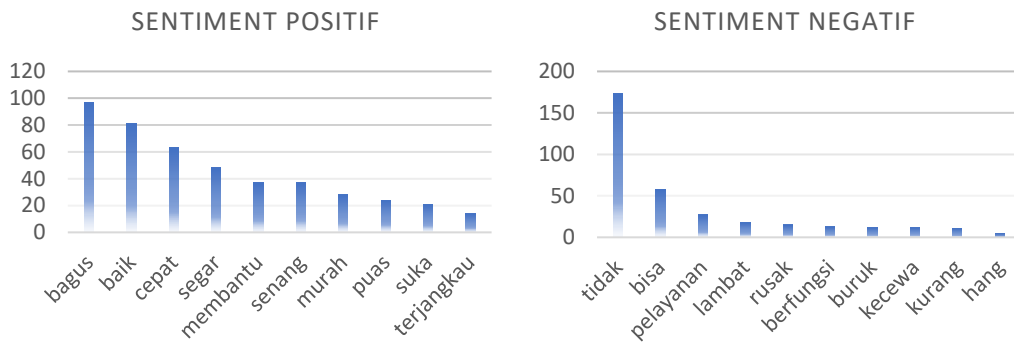
Gambar 7. Frekuensi kata yang paling banyak muncul pada review sayurbox

Pada hasil klasifikasi review aplikasi sayurbox dengan sentiment positif dengan jumlah 738 review, didapatkan beberapa kata yang sering muncul yaitu kata “bagus” dengan frekuensi 240, “segar” sebanyak 127, “baik” sebanyak 115, “membantu” sebanyak 111, “mudah” sebanyak 87, “senang” sebanyak 69, “kualitas” sebanyak 59, “cepat” sebanyak 54, “rapi” sebanyak 24, dan kata “pandemi” sebanyak 20. Dari hasil visualisasi dapat dilihat terdapat kata “pandemi”, kata tersebut mengacu bahwa pengguna merasa terbantu dengan aplikasi sayurbox untuk memenuhi kebutuhan dimasa pandemi.

Sedangkan pada sentiment negatif dengan jumlah 309 review didapatkan kata yang sering muncul yaitu kata “tidak” dengan frekuensi 485, “bisa” sebanyak 105, “pelayanan” sebanyak 71, “buruk” sebanyak 29, “busuk” sebanyak 23, “lama” sebanyak 22, “salah” sebanyak 20, “kurang” sebanyak 17, “kecewa” sebanyak 16, dan kata “loading” sebanyak 10.



Gambar 8. Visualisasi wordcloud pada review tanihub



Gambar 9. Frekuensi kata yang paling banyak muncul pada *review* tanihub

Berdasarkan **Gambar 9** dapat dilihat pada sentiment positif dengan jumlah 348 *review* tanihub, kata yang sering dibicarakan yaitu kata “bagus” dengan frekuensi 97, “baik” sebanyak 81, “cepat” sebanyak 63, “segar” sebanyak 48, “membantu” sebanyak 37, “senang” sebanyak 37, “murah” sebanyak 28, “puas” sebanyak 24, “suka” sebanyak 21, dan kata “terjangkau” sebanyak 14.

Sedangkan pada sentiment negatif *review* tanihub dengan jumlah 184 *review*, kata yang sering muncul yaitu kata “tidak” dengan frekuensi 173, “bisa” sebanyak 58, “pelayanan” sebanyak 27, “lambat” sebanyak 18, “rusak” sebanyak 15, “berfungsi” sebanyak 13, “buruk” sebanyak 12, “kecewa” sebanyak 12, “kurang” sebanyak 11, dan kata “hang” sebanyak 5.

4 Kesimpulan

1. Dengan kurun waktu Februari 2020 hingga Januari 2021, Aplikasi sayurbox lebih banyak dipakai oleh khalayak publik jika dibanding dengan aplikasi tanihub dengan tingkat sentiment positif terbanyak. Didapatkan sentimen negatif yang hanya berjumlah 309 dari total 1.047 *review* pada layanan sayurbox. Pada tanihub, terdapat 184 sentimen *negatif* dari total 532 *review*. Hasil analisa, disimpulkan bahwa pelayanan yang diberikan Sayurbox maupun Tanihub sudah cukup baik, namun, sayurbox lebih unggul dibandingkan dengan tanihub jika dilihat dari persentase jumlah ulasan pengguna dan rating.
2. Dengan analisa deskriptif, secara umum dapat dilihat peningkatan penggunaan layanan sayurbox dan tanihub jatuh pada bulan April dan bulan Mei 2020, hal ini diperkirakan terjadi karena di bulan tersebut terdapat kebijakan PSBB (Pembatasan Sosial Berskala Besar) oleh pemerintah akibat dari dampak penyakit menular coronavirus. Pada bulan April, aplikasi tanihub mendapatkan peningkatan pengguna dengan tingkat keluhan tertinggi dibanding bulan lainnya. Dengan melihat isi *review*, pengguna aplikasi tanihub lebih banyak memberi keluhan mengenai aplikasi pada bulan april dengan persentase sentiment negatif yaitu 59% dengan jumlah 41 ulasan, sedangkan sentiment positif hanya berjumlah 29 ulasan (41%) dengan total keseluruhan ulasan april yaitu 70 ulasan. Banyaknya keluhan diperkirakan karena adanya hang atau eror pada aplikasi, kata hang dalam ulasan tanihub masuk kedalam 10 daftar kata dengan frekuensi tertinggi dalam sentiment negatif. Berbeda dengan tanihub, jumlah ulasan yang didapat *e-commerce* sayurbox pada bulan april jauh lebih tinggi, sentiment negatif pada bulan april berjumlah 82 data (29%) dan sentiment positif yaitu 196 data (71%) dari total 278 data *review*.
3. Dari hasil penelitian, metode SVM pada klasifikasi sentimen berjalan dengan baik. Dengan penerapan pembagian data train dan data test dengan perbandingan 80 : 20 didapatkan hasil akurasi klasifikasi pada data sayurbox dengan metode SVM dengan kernel linear yaitu sebesar 91,4%, presisi 93,1%, dan recall yaitu sebesar 94,4%. Hasil akurasi pada data sayurbox, disimpulkan bahwa dari 210 *review* yang diuji didapatkan 192 *review* yang sesuai dengan pengklasifikasiannya. Sedangkan pada data Tanihub, total akurasi yang diperoleh dengan menggunakan metode yang sama sebesar 88,8%, presisi sebesar 92%, dan recall sebesar 92%. Pada data tanihub, dari 107 *review* yang diuji terdapat 95 *review* yang tepat dengan pengklasifikasiannya.

Referensi

- [1] Alvara, 2020. Catatan akhir tahun alvara [online] Tersedia pada: <http://alvara-strategic.com/wp-content/uploads/2020/12/Catatan-Akhir-Tahun-Alvara-2020.pdf>
- [2] Dailysocial, 2020. *Sederet Aplikasi Belanja Online Terpopuler Selama Pandemi* [online] Tersedia pada: <https://dailysocial.id/post/sederet-aplikasi-belanja-online-terpopuler-selama-pandemi>
- [3] Google Play, 2020. Tersedia pada: https://id.wikipedia.org/wiki/Google_Play
- [4] Praptiwi, D. Y. 2018. ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA E-COMMERCE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN MAXIMUM ENTROPY (Studi Kasus: Review Bukalapak pada Google Play).
- [5] Purbo, O. W. 2019. Text Mining-Analisis Medsos, Kekuatan Brand, & Intelijen di Internet. ANDI (Anggota IKAPI).
- [6] Kuncahyo, 2019. *Confusion Matrix untuk Evaluasi Model pada Supervised Learning*. [online] Tersedia pada: <https://medium.com/@ksnugroho/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>
- [7] Scikit.learn.org, 2020. *1.4. Support Vector Machines*. [online] Tersedia pada: <https://scikit-learn.org/stable/modules/svm.html>
- [8] Ilmawan, L. B., Mude, M. A. 2020. *Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store*. ILKOM Jurnal Ilmiah, 12(2), 154-161.
- [9] Informatikalogi.com, 2021. *Pembobotan Kata atau Term Weighting TF-IDF*. [online] Tersedia pada: <https://informatikalogi.com/term-weighting-tf-idf/>
- [10] Nugroho, A. S., Witarto, A. B., dan Handoko, D. 2003. *Support Vector Machine*. Proceeding Indones. Sci. Meeting Cent. Japan.