

Analisis Sentimen Pada Ulasan Aplikasi Indodax di *Google Play Store* Menggunakan Metode *Support Vector Machine*

Alfio Kusuma¹, Ermatita², Helena Nurramdhani Irmanda³
Informatika

Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Jakarta Selatan,
Daerah Khusus Ibukota Jakarta 12450
alksm699@gmail.com¹, ermatita@upnvj.ac.id², helenairmanda@upnvj.ac.id³

Abstrak. Indodax merupakan salah satu aplikasi yang digunakan sebagai *platform* jual beli aset kripto yang telah terdaftar di Badan Pengawas Perdagangan Berjangka Komoditi (Bappebti) dan dapat diunduh gratis di *google play store*. Ulasan yang diterima aplikasi memungkinkan dapat mempengaruhi pengguna yang akan mengunduh aplikasi Indodax. Oleh karena itu, pada penelitian ini akan dilakukan analisis sentimen terhadap ulasan pada aplikasi Indodax yang diberikan oleh pengguna di *google play store*. Tahapannya dimulai dari pengambilan data ulasan aplikasi Indodax pada *google play store*. Kemudian data diberi label negatif atau positif, lalu masuk ke praproses data untuk pembersihan data, dilanjutkan dengan pembobotan kata *Term Frequency-Invers Document Frequency* (TF-IDF) dan diklasifikasikan menggunakan metode *Support Vector Machine* yang menghasilkan akurasi sebesar 85% pada rasio pembagian data sebesar 80:20.

Kata Kunci: Analisis Sentimen, Aset Kripto, Indodax, *Support Vector Machine* (SVM).

1. Pendahuluan

Aset kripto merupakan bagian dari kelas mata uang digital [1] yang telah ditetapkan menjadi salah satu komoditas yang dapat diperdagangkan di Bursa Berjangka dan telah berkembang luas di masyarakat [2]. Berdasarkan laporan Kajian Stabilitas Keuangan Nomor 37 oleh Bank Indonesia, pada Juni 2021 jumlah investor aset kripto diperkirakan telah melebihi ± 6.5 juta, melewati jumlah investor di pasar saham yang sebesar ± 2.4 juta. Bank Indonesia menyebutkan bahwa Indodax menjadi *platform* jual beli (*marketplace*) aset kripto terbesar di Indonesia [3].

Indodax telah terdaftar di Badan Pengawas Perdagangan Berjangka Komoditi (Bappebti) sebagai calon pedagang fisik aset kripto [4]. Sejak didirikan dari 2014 hingga saat ini, Indodax telah melayani lebih dari 4,3 juta member terdaftar dan terverifikasi. Indodax tersedia dalam bentuk situs dan aplikasi *mobile* yang dapat diunduh gratis di *Apps markets* seperti Google Play Store atau Apple store [5]. *Google Play Store* juga mempunyai fitur lain yaitu fitur *rating* dan ulasan. Pengguna dapat menilai layanan atau aplikasi dari 1 hingga 5 (1 adalah nilai terendah dan 5 adalah yang terbaik), dan memberikan tanggapan atau penilaian aplikasi yang digambarkan dalam bentuk kalimat teks.

Untuk kemudahan pengolahan data ulasan, penelitian ini akan menggunakan analisis sentimen. Terdapat penelitian sebelumnya dengan menggunakan salah satu model *machine learning* yaitu *Support Vector Machine*, analisis sentimen dapat diklasifikasikan dengan baik [6]. Analisis sentimen pada data ulasan lainnya juga telah diteliti sebelumnya seperti pada data ulasan film [7] dan ulasan 13 aplikasi dari *Google Play Store* [8]; [9]; [10]. Dari beberapa penelitian tersebut, model-model *machine learning* yang digunakan menghasilkan akurasi yang cukup baik. Serupa dengan penelitian yang telah dilakukan sebelumnya, penelitian kali ini akan melakukan analisis sentimen dengan menggunakan model *Support Vector Machine* namun dengan data yang berbeda yakni menggunakan data ulasan dari aplikasi Indodax.

2. Landasan Teori

2.1. Analisis Sentimen

Analisis sentimen atau disebut juga *opinion mining* adalah bidang studi yang menganalisis pendapat, sentimen, penilaian, sikap, dan emosi orang terhadap entitas yang diungkapkan dalam teks tertulis. Entitas tersebut dapat berupa produk, layanan, organisasi, individu, peristiwa, masalah atau topik [11].

2.2. Text Mining

Text mining didefinisikan sebagai kumpulan berbagai dokumen untuk mendapatkan informasi atau pengetahuan menggunakan suatu alat analisis [12].

2.3. Praproses Data

Tahap praproses data merupakan tahap awal untuk mengolah data yang akan digunakan sebelum membuat model, dataset yang akan digunakan harus dilakukan pembersihan terlebih dahulu dari kata yang tidak relevan dengan sentimen seperti tanda baca, angka, *link* url dan symbol-simbol. Selain itu, juga dilakukan proses *stemming*, *normalization*, dan *removing stop words* [13].

2.3.1 Case Folding

Case folding yaitu proses dimana seluruh huruf kapital pada teks diubah menjadi huruf kecil [14], hal ini bertujuan agar pada tahap pembuatan model klasifikasi huruf pada text seragam dan tidak terjadi kesalahan saat proses tokenisasi.

2.3.2 Data Cleaning

Data Cleaning dilakukan untuk mengurangi *noise data* yaitu dengan membersihkan dan menghilangkan kata-kata yang tidak perlu, seperti *link* URL, *hashtag* (#), *username* (@username), tanda baca dan email [15].

2.3.3 Normalization

Pada tahap *normalization* dilakukan perubahan semua kata singkatan menjadi kata-kata normal, peneliti membuat kamus yang didalamnya terdapat kata-kata asli dan kata-kata singkatan. Sebagai contoh, “bgt:banget”, “tdk:tidak”, dan kata singkatan lainnya [13].

2.3.4 Stopword Removal

Stopword Removal merupakan tahap dimana kata yang tidak mempunyai arti dihapus, biasanya seperti kata keterangan, kata sambung, dan lain sebagainya. Kata-kata tersebut terdapat dalam *library* sastrawi dan data pendukung yaitu data *stoplist* yang sudah disiapkan sebelumnya. Data *stoplist* tersebut berisi kata-kata yang berkemungkinan besar tidak berpengaruh dalam tahap prediksi. Pada tahap selanjutnya, kata yang tidak berguna kemudian dibuang dan tidak digunakan kembali [16].

2.3.5 Stemming

Pada proses *stemming*, masing-masing kata yang memiliki kata imbuhan dihilangkan sehingga kata tersebut berubah menjadi kata dasar, pada tahap ini kata yang memiliki ejaan yang kurang tepat juga dibersihkan [17]

2.3.6 Tokenization

Tokenization atau tokenisasi yaitu tahap untuk pemotongan pada string input sesuai kata penyusunnya [18]. Proses tokenisasi ini akan mempermudah dalam membedakan karakter yang dapat dijadikan sebagai pemisah kata atau

bukan, karena secara garis besar pada proses ini karakter-karakter yang ada dalam teks akan dipecahkan ke dalam satuan kata.

2.4. Pembobotan Kata

Pemberian bobot pada kata dalam *text mining* atau disebut juga *Term Frequency-Inverse Document Frequency* (TF-IDF) yaitu teknik ekstraksi fitur dengan memberi nilai atau bobot untuk setiap kata (*term*) pada data yang berhasil di ekstrak [19]. Hal ini bertujuan agar pentingnya sebuah kata dalam mewakili sebuah kalimat dapat diketahui, sehingga perlu dilakukan perhitungan atau pembobotan pada kata [20]. Pembobotan kata dilakukan menggunakan data ulasan yang sudah dibersihkan sebelumnya melalui praproses data.

Perhitungan pembobotan kata dalam TF-IDF dapat menggunakan rumus yaitu:

1. *Term Frequency – Inverse Document Frequency* (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) digunakan untuk pemberian bobot pada setiap *term* dengan melakukan perkalian antara *term frequency* dan *inverse document frequency*.

$$W_{t,d} = TF_{t,d} \times IDF_t \quad (1)$$

Dimana $W_{t,d}$ adalah bobot kata t pada suatu dokumen, $TF_{t,d}$ yaitu frekuensi munculnya kata t pada suatu dokumen d , dan IDF_t adalah kemunculan suatu *term* pada koleksi dokumen seluruhnya, dan *Term Frequency* (TF) merupakan jumlah kemunculan kata (t) dalam suatu dokumen (d).

2. *Invers Document Frequency* (IDF)

Inverse Document Frequency (IDF) digunakan untuk menghitung kemunculan suatu *term* t pada koleksi dokumen seluruhnya.

$$IDF_t = \log \left(\frac{N}{df_t} \right) \quad (2)$$

N adalah total semua dokumen pada koleksi, df_t adalah total dokumen yang mengandung *term* (kata) t .

2.5. Support Vector Machine

Support Vector Machine adalah metode untuk melakukan pemisahan kumpulan data ke dalam beberapa kelompok dengan membangun garis yang memisahkan satu kelompok dan kelompok lainnya menggunakan pemisah yang tegas [21]. *Support Vector Machine* mempunyai kemampuan untuk mempelajari pola klasifikasi data dengan akurasi dan reproduktifitas yang seimbang dan fleksibilitas yang tinggi [22]. SVM termasuk pada metode klasifikasi kategori *supervised learning*, yang sebenarnya menggunakan konsep dasar yang merupakan kominasi harmonis dari kumpulan teori komputasi yang telah ada, seperti yang diperkenalkan pada tahun 1950 oleh Aronszajn yang bernama margin *hyperplane kernel*, dan demikian juga dengan dukungan dari konsep-konsep yang lainnya [23].

2.6 Evaluasi

Pada tahap evaluasi akan dilakukan pengujian seberapa baik model klasifikasi yang telah dibentuk dengan menghitung *accuracy*, *precision*, *recall* dan *f1-score*. Pengujian dilakukan dengan menggunakan *confusion matrix* yang terdiri dari 4 bentuk kombinasi yang berbeda dari nilai aktual dan nilai prediksinya dalam menentukan matrix untuk mengukur klasifikasi [24].

Tabel 1. *Confusion Matrix*

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

Sumber: Ting,2017

Perhitungan *accuracy*, *precision*, *recall*, dan *F1 score*, dilakukan berdasarkan hasil TP, TN, FP, dan FN diatas. Perhitungannya dilakukan dengan rumusan berikut ini:

1. *Accuracy* merupakan hasil rasio prediksi yang benar dari total data yang ada, dengan rumus menghitungnya:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. *Precision* merupakan rasio prediksi benar positif dibanding dengan seluruh hasil yang diprediksi positif, dengan rumus perhitungannya:

$$Precision (P) = \frac{TP}{TP+FP} \quad (4)$$

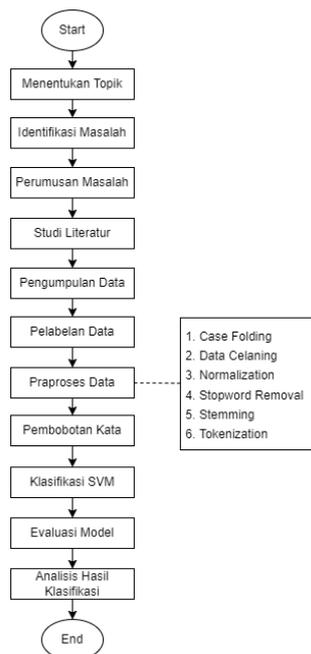
3. *Recall* merupakan rasio prediksi benar positif dibanding dengan seluruh data aktual positif, dapat dihitung dengan rumusan:

$$Recall (R) = \frac{TP}{TP+FN} \quad (5)$$

4. *F1 Score* merupakan keseimbangan dari *precision* dan *recall*, persamaan rumus untuk menghitungnya yaitu:

$$F1\ Score = \frac{2.P.R}{P+R} \quad (6)$$

3. Metodologi Penelitian



Gambar 1. merupakan alur penelitian yang terdiri dari beberapa tahapan, dengan penjelasannya yaitu sebagai berikut:

1. **Menentukan Topik**, topik yang dipilih yaitu analisis sentiment pada ulasan pengguna di Google Play Store yaitu aplikasi Indodax yang merupakan salah satu aplikasi yang bergerak dibidang asset kripto dan *blockchain*.
2. **Identifikasi Masalah**, pengamatan dilakukan terhadap kondisi aplikasi Indodax di *Google Play Store* dan situs lain yang berhubungan.
3. **Perumusan Masalah**, setelah identifikasi masalah dilakukan, perumusan masalah diperlukan agar dapat menjadi pedoman yang digunakan dalam melakukan penelitian.
4. **Studi Literatur**, Studi literatur diperlukan sebagai bahan pendukung untuk memecahkan masalah penelitian yang dijadikan sebagai sumber pustaka, diantaranya dengan mencari dan mengumpulkan berbagai macam buku, penelitian/jurnal terdahulu dan situs-situs yang berkaitan dengan analisis sentimen dan algoritma *Support Vector Machine*.

5. **Pengumpulan Data**, Data yang dikumpulkan berupa data ulasan pengguna aplikasi Indodax yang berbahasa Indonesia dengan periode ulasan pada bulan Oktober 2021 di *Google Play Store* dengan menggunakan *google-play-scraper*.
6. **Pelabelan Data**, Pada tahap ini, sebelum masuk ke proses analisis sentimen, data yang digunakan akan diberi label sesuai dengan *rating* ulasan. Penulis hanya memilih kelas sentimen negatif dan positif untuk digunakan. Pada ulasan yang menerima *rating* 1,2, atau 3 akan diberi label negatif (nilai 0), sedangkan ulasan yang menerima *rating* 4 atau 5 akan diberi label positif (nilai 1) [25].
7. **Praproses Data**, Praproses disini adalah kegiatan yang melakukan proses sentimen analisis yang terdiri dari tahap *case folding*, *data cleaning*, *normalization*, *stopword removal*, *stemming* dan *tokenization*.
 - a. **Case Folding**, tahap bertujuan untuk mengubah semua data ulasan berupa teks menjadi huruf kecil.
 - b. **Data Cleaning**, *Data Cleaning* bertujuan untuk membersihkan dan menghapus *link URL*, *hashtag*, simbol-simbol, tanda baca dan huruf selain dari huruf alphabet yang ada pada data ulasan.
 - c. **Normalization**, Pada tahap ini dilakukan perubahan dan perbaikan terhadap kata-kata yang disingkat menjadi kata-kata normal.
 - d. **Stopword Removal**, Selanjutnya hasil dari proses *normalization* akan memasuki tahap berikutnya yaitu tahap *stopword removal*. Tahap ini merupakan tahap untuk menghapus kata yang tidak mempunyai makna penting untuk klasifikasi, biasanya berupa kata sambung, kata keterangan, dan sebagainya.
 - e. **Stemming**, *stemming* dilakukan untuk mengubah kata yang memiliki imbuhan dalam bahasa indonesia menjadi kata dasar. Pada proses *stemming*, penghapusan imbuhan akan dilakukan sesuai aturan *library* sastrawi.
 - f. **Tokenization**, Tahap *tokenization* bertujuan untuk memotong atau memecah kalimat pada teks menjadi bentuk satuan kata menggunakan spasi (*whitespace*) untuk pemisahannya.
8. **Pembobotan Kata**, Pada tahap ini, setiap kata yang ada akan diberikan bobot yang bertujuan agar kata yang didapat dari ekstraksi mendapat nilai, yaitu dengan menggunakan metode pemberian bobot *Term Frequency – Inverse Document Frequency* (TF-IDF). Perhitungan dapat dilakukan dengan cara seperti yang ada pada bab 2 persamaan.
9. **Klasifikasi SVM**, Algoritma yang digunakan untuk membuat model klasifikasi pada data ulasan aplikasi Indodax adalah *support vector machine*. Sebelum melakukan klasifikasi, dilakukan pembagian data menjadi data latih dan data uji. Pembagian data latih dan data ujinya akan menggunakan tiga rasio dalam persen dengan perbandingan 60:40, 70:30, dan 80:20. Setiap rasio pembagian data yang digunakan akan mendapat perlakuan yang sama pada pelatihan dan pengujian data.
10. **Evaluasi Model**, Tahap evaluasi akan melakukan pengujian terhadap model klasifikasi yang telah dibuat dengan menggunakan *confusion matrix* yang berguna untuk melihat hasil *accuracy*, *precision*, *recall*, dan *F1-score* dengan persamaan rumus masing-masing yakni pada persamaan
11. **Analisis Hasil Klasifikasi**
 Analisis dari hasil klasifikasi setiap ulasan yang mengandung sentimen negatif atau positif tersebut dapat ditampilkan melalui visualisasi yaitu melalui *wordcloud*, dimana akan dibentuk *wordcloud* yang akan menggambarkan visualisasi kelas sentimen negatif dan positif dan *wordlink* yang dapat menunjukkan kata yang berhubungan pada kelas yang ada.

4. Hasil dan Pembahasan

Data dikumpulkan dengan teknik *web scraping* menggunakan *google-play-scraper*. *Google-play-scraper* menyediakan API untuk mempermudah *scraping* data di *Google Play Store* yang tersedia pada *library* bahasa pemrograman *Python*. Data yang dikumpulkan yaitu data ulasan aplikasi Indodax yang berbahasa Indonesia dengan periode ulasan pada bulan Oktober 2021. Hasil yang didapat dari *scraping* yaitu sebanyak 1138 data ulasan.

Data ulasan pada penelitian ini akan dilabelkan menjadi data ulasan negatif dan positif. Pelabelan data disesuaikan dengan *rating* yang diterima aplikasi di *Google Play Store*, dimana *rating* dengan bintang 1 sampai 3 akan dilabeli sebagai sentimen negatif yang diberi nilai 0, sementara itu *rating* dengan bintang 4 dan 5 akan dilabeli sebagai sentimen positif yang diberi nilai 1. Hasil akhir pelabelan dari seluruh data ulasan yang digunakan yaitu terdapat

565 ulasan berlabel sentimen positif dan 573 ulasan berlabel sentimen negatif dari total 1138 data ulasan yang dilakukan secara manual. Setelah pelabelan data, akan dilanjutkan dengan praproses data agar data dapat diolah dan dimodelkan yang dilakukan melalui enam tahapan sentimen analisis yaitu sebagai berikut.

a. Case Folding

Pada tahap *case folding*, setiap data teks pada ulasan aplikasi indodax yang menggunakan huruf kapital akan diubah menjadi huruf kecil. Berikut pada Tabel 2 hasil perubahan data awal dan sesudah proses *case folding* dilakukan:

Tabel 2. Sebelum dan Sesudah *Case Folding*

Sebelum	Sesudah
Pemula jgn download, fee nya terlalu besar utk kita yg memulai dengan modal kecil!!!	pemula jgn download, fee nya terlalu besar utk kita yg memulai dengan modal kecil!!!
Mudah di pahami dan memberikan motif yg jelas. Withdraw juga cepet.	mudah di pahami dan memberikan motif yg jelas. withdraw juga cepet.
Liat komenan jelek jelek bngt saya santai santai aja ga ada problem dg apk ini,makanya plajari dulu cara kerjanya	liat komenan jelek jelek bngt saya santai santai aja ga ada problem dg apk ini,makanya plajari dulu cara kerjanya

b. Data Cleaning

Data cleaning dilakukan setelah proses *case folding*. Pada tahap ini akan dilakukan pembersihan data ulasan yang mempunyai simbol atau karakter selain huruf alphabet. Berikut pada Tabel 3 hasil perubahan data awal dan sesudah proses *data cleaning* dilakukan:

Tabel 3. Sebelum dan Sesudah *Data Cleaning*

Sebelum	Sesudah
pemula jgn download, fee nya terlalu besar utk kita yg memulai dengan modal kecil!!!	pemula jgn download fee nya terlalu besar utk kita yg memulai dengan modal kecil
mudah di pahami dan memberikan motif yg jelas. withdraw juga cepet.	mudah di pahami dan memberikan motif yg jelas withdraw juga cepet
liat komenan jelek jelek bngt saya santai santai aja ga ada problem dg apk ini, makanya plajari dulu cara kerjanya	liat komenan jelek jelek bngt saya santai santai aja ga ada problem dg apk ini makanya plajari dulu cara kerjanya

c. Normalization

Tahap *normalization* adalah tahap lanjutan setelah proses *data cleaning*, yaitu memperbaiki kata singkatan, kata yang salah eja, atau kata yang bermakna sama namun penulisannya berbeda. Sebelum proses *normalization*, telah disiapkan kamus yang berisi daftar kata yang akan digunakan untuk proses normalisasi kata, yaitu sebanyak 2104 kata. Berikut pada Tabel 4 hasil perubahan data awal dan sesudah proses *data cleaning* dilakukan:

Tabel 4. Sebelum dan Sesudah *Normalization*

Sebelum	Sesudah
pemula jgn download fee nya terlalu besar utk kita yg memulai dengan modal kecil	pemula jangan unduh biaya nya terlalu besar untuk kita yang memulai dengan modal kecil
mudah di pahami dan memberikan motif yg jelas withdraw juga cepet	mudah di pahami dan memberikan motif yang jelas tarik juga cepat
liat komenan jelek jelek bngt saya santai santai aja ga ada problem dg apk ini makanya plajari dulu cara kerjanya	lihat komentar jelek jelek banget saya santai santai saja tidak ada masalah dengan aplikasi ini makanya belajar dulu cara kerjanya

d. **Stopword Removal**

Stopword removal merupakan proses penghapusan kata yang sering muncul dan tidak mempunyai makna penting pada klasifikasi. Penghapusan kata akan dilakukan berdasarkan aturan pada *library* sastraawi dan kamus daftar *stopword* yang telah dibuat sebelumnya. Berikut pada Tabel 5 hasil perubahan data awal dan sesudah proses *stopword removal* dilakukan:

Tabel 5. Sebelum dan Sesudah *Stopword Removal*

Sebelum	Sesudah
pemula jangan unduh biaya nya terlalu besar untuk kita yang memulai dengan modal kecil	unduh biaya modal
mudah di pahami dan memberikan motif yang jelas tarik juga cepat	mudah pahami motif tarik cepat
lihat komentar jelek jelek banget saya santai santai saja tidak ada masalah dengan aplikasi ini makanya belajar dulu cara kerjanya	komentar jelek jelek santai santai aplikasi belajar

e. **Stemming**

Tahap *stemming* ini dilakukan untuk merubah kata pada data ulasan menjadi kata dasar yaitu menghapus kata imbuhan baik di awal, di tengah, maupun di akhir kata dengan menggunakan aturan pada *library* sastraawi. Berikut proses *stemming* pada Tabel 6.

Tabel 6. Sebelum dan Sesudah *Stopword Removal*

Sebelum	Sesudah
unduh biaya modal	unduh biaya modal
mudah pahami motif tarik cepat	mudah paham motif tarik cepat
komentar jelek jelek santai santai aplikasi belajar	komentar jelek jelek santai santai aplikasi ajar

f. **Tokenization**

Tahap *tokenization* dilakukan dengan memotong atau memecah deretan kata di dalam kalimat pada data ulasan hasil dari proses *stemming* sebelumnya menjadi satuan kata tunggal atau kata penyusunnya. Berikut pada Tabel 7 hasil perubahan data awal dan sesudah proses *tokenization* dilakukan:

Tabel 7. Sebelum dan Sesudah *Stopword Removal*

Sebelum	Sesudah
unduh biaya modal	['unduh', 'biaya', 'modal']
mudah paham motif tarik cepat	['mudah', 'paham', 'motif', 'tarik', 'cepat']
komentar jelek jelek santai santai aplikasi ajar	['komentar', 'jelek', 'jelek', 'santai', 'santai', 'aplikasi', 'ajar']

Tahap selanjutnya setelah melewati praproses data yaitu pembobotan kata. Pembobotan pada kata dilakukan dengan metode pemberian bobot *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan TF-IDF ini akan dilakukan ke seluruh dokumen dengan total 1138 dokumen yang telah melewati praproses data, dimana dari perhitungan pembobotan tersebut terdapat jumlah kata sebanyak 910 kata.

Setelah proses pembobotan kata, tahap berikutnya adalah membuat model klasifikasi. Pembuatan model klasifikasi dilakukan dengan membagi data latih dan data uji terlebih dahulu. Pembagian data latih dan data uji akan menggunakan tiga rasio perbandingan yaitu 60:40, 70:30, serta 80:20 dari total sebanyak 1138 data ulasan. Berikut pembagian data dapat dilihat pada Tabel 8.

Tabel 8. Pembagian Data

Rasio Pembagian		Jumlah Data Latih	Jumlah Data Uji
Data Latih	Data Uji		
60%	40%	682	456
70%	30%	796	342
80%	20%	910	228

Pada Tabel 8 menunjukkan pembagian data bahwa pada rasio 60:40, pembagian data latih adalah sebanyak 682 dan data uji sebanyak 456. Kemudian pada rasio 70:30 membagi data latih sebanyak 796 dan data uji sebanyak 342. Selanjutnya pada rasio 80:20 membagi data latih sebanyak 910 data sementara itu data uji sebanyak 228 data. Pembuatan model klasifikasinya menggunakan algoritma *support vector machine*, dengan beberapa kernel yang diujikan yaitu kernel *linear*, *rbf*, *polynomial*, dan *sigmoid*. *Hyperparameter tuning* akan dilakukan dengan *Grid Search CV* yang berguna untuk mencari parameter *cost* (C), *gamma* (γ) dan kernel yang memiliki hasil performa terbaik untuk model SVM yang akan digunakan. Penentuan nilai C dan gamma yang akan digunakan mengacu pada penelitian sebelumnya yakni oleh Prangga menggunakan $C = \{0.5, 0.75, 1, 10, 100\}$, $\gamma = \{0.005, 0.05, 0.1, 0.5, 0.75\}$ [26], kemudian pada penelitian oleh Indraini, Ernawati dan Zaidah menggunakan $C = \{1, 10, 100, 100\}$, $\gamma = \{0.01, 0.1, 1, 10, 100\}$ [27] maka pada penelitian ini peneliti melakukan sedikit penyesuaian dalam memilih nilai yang memungkinkan untuk digunakan. Parameter C yang diujikan bernilai 0.5, 0.75, 0.1, 1, 10, 20, 40, 50, 100 dan parameter *gamma* bernilai 1, 0.1, 0.01, 0.001. Hasil proses *Grid Search CV* ditulis dan dapat dilihat dalam bentuk tabel berikut ini.

Tabel 9. Hasil *Grid Search CV*

Kernel	Parameter			
	Rasio pembagian	C	Gamma	Akurasi
Linear	60:40	0,75	1	83,27
	70:30	0,75	1	81,91
	80:20	0,5	1	82,19
RBF	60:40	10	0,1	83,42
	70:30	1	1	82,29
	80:20	20	0,01	82,52
Polynomial	60:40	10	1	74,91
	70:30	10	1	75,88
	80:20	10	1	76,59
Sigmoid	60:40	10	0,1	83,27
	70:30	10	0,1	81,53
	80:20	40	0,01	82,52

Pada Tabel 9 dapat dilihat hasil akurasi dari proses *Grid Search CV* yang menunjukkan model terbaik dari semua nilai parameter C, *gamma* dan kernel yang diujikan pada setiap rasio pembagian data. Performa terbaik terdapat pada kernel RBF dari ketiga rasio yang digunakan, dimana pada rasio 60:40 nilai $c = 10$, $\gamma = 0.1$, akurasi = 83.42%, dan pada rasio 70:30 nilai $c = 1$, $\gamma = 1$, akurasi = 82.29%, serta pada rasio 80:20 hasil akurasinya sebesar 82,52% dengan parameter $c = 20$ dan $\gamma = 0.01$. Kemudian, model-model tersebut akan dilanjutkan dengan melakukan pengujian model menggunakan data uji yang berjumlah 456 (rasio 60:40), 342 (rasio 70:30) dan 228 data (rasio 80:20). Evaluasi model dilakukan untuk mengetahui hasil akurasi proses klasifikasi dari model-model yang telah dibentuk menggunakan data uji dan akan dievaluasi dengan *confusion matrix* yang terdiri dari 4 kombinasi berbeda dari nilai prediksi dan nilai aktual yaitu TP, TN, FP, dan FN. *Confusion matrix* hasil klasifikasi dapat dilihat pada tabel berikut.

Tabel 10. Hasil *Confusion Matrix*

Rasio pembagian	TP	FP	TN	FN
60:40	188	28	184	56
70:30	146	9	143	44
80:20	96	6	100	26

Tabel 10 menunjukkan hasil *confusion matrix*, pada rasio 60:40 nilai *true positive* (TP) atau ulasan positif yang diklasifikasi dengan benar sebagai ulasan positif berjumlah 188, nilai *false negatif* (FN) atau ulasan positif yang salah diklasifikasikan sebagai ulasan negatif berjumlah 56, nilai *false positif* (FP) atau ulasan negatif yang salah diklasifikasikan sebagai ulasan positif berjumlah 28 dan nilai *true negative* (TN) atau ulasan negatif yang diklasifikasikan dengan benar sebagai ulasan negatif berjumlah 184. Sementara itu, pada rasio 70:30 nilai TP berjumlah 146, nilai FN berjumlah 44, nilai FP berjumlah 9, dan nilai TN berjumlah 143. Kemudian pada rasio 80:20 nilai TP berjumlah 96, nilai FN 26, nilai FP berjumlah 6 dan nilai TN berjumlah 100.

Berdasarkan pada hasil *confusion matrix* yang telah diperoleh, dapat dilakukan perhitungan *accuracy*, *precision*, *recall* dan *F1-score* dengan persamaan (3), (4), (5), (6). Hasil yang diperoleh dari perhitungan evaluasi dapat dilihat pada tabel dibawah ini.

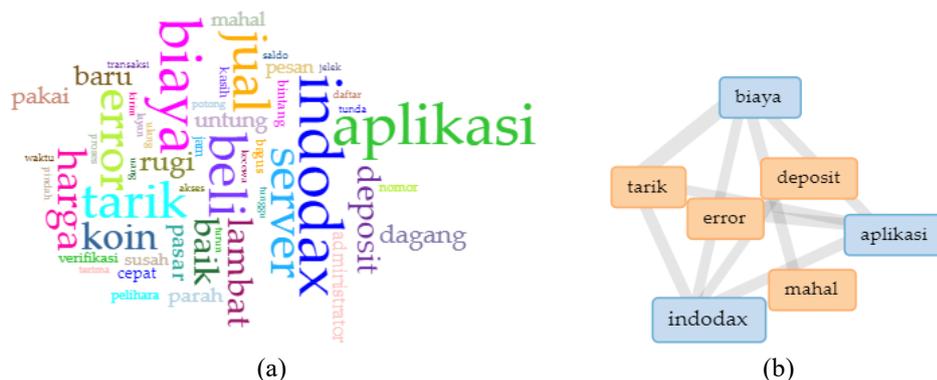
Tabel 11. Hasil Evaluasi

Parameter				
Rasio pembagian	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
60:40	0,815789	0,87037	0,770492	0,817391
70:30	0,845029	0,941935	0,768421	0,846377
80:20	0,859649	0,941176	0,786885	0,857143

Tabel 11 menunjukkan perolehan hasil dari evaluasi data pada setiap rasio pembagian yang digunakan, yang mana akurasi tertinggi terdapat pada rasio pembagian 80:20 dengan nilai *accuracy* sebesar 0.85 atau dengan persentase 85%, *precision* bernilai 0.94 atau dengan persentase 94%, *recall* bernilai 0.78 atau dengan persentase 78% dan *f1-score* bernilai 0.85 atau dengan persentase 85%. Selanjutnya adalah analisis hasil klasifikasi yang dilakukan dalam bentuk visualisasi terhadap hasil klasifikasi yang diperoleh pada setiap kelas sentimen negatif dan kelas sentimen positif yang dapat dilihat pada *wordcloud* dan *wordlink* berikut ini.

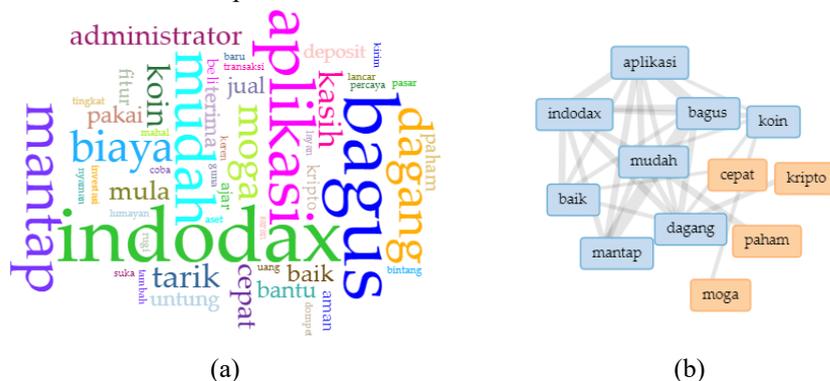
a. *Wordcloud* dan *wordlink* sentimen negatif

Untuk mengetahui hal yang sering dikomentari pada setiap ulasan yang diberikan oleh pengguna dapat menggunakan *wordcloud*. *Wordcloud* dapat menampilkan gambaran kata yang sering paling sering digunakan sehingga hasil dari visualisasi ini dapat berguna bagi Indodax untuk mengetahui bagaimana tanggapan pengguna aplikasi dan upaya apa saja yang harus dilakukan agar kualitas aplikasi dapat ditingkatkan, sementara itu *wordlink* digunakan untuk melihat kata yang berhubungan pada suatu kelas. Visualisasi dari hasil klasifikasi yang telah diperoleh dapat dilihat pada gambar *wordcloud* dan *wordlink* dibawah ini.



Gambar 2. merupakan hasil visualisasi *wordcloud* (a) dan *wordlink* (b) ulasan aplikasi Indodax untuk kelas sentimen negatif. Semakin besar ukuran tulisan yang ditampilkan pada gambar (a), semakin sering kata tersebut digunakan oleh pengguna sebagai topik yang sering dibahas pada ulasan yang diberikan. Hasil visualisasi pada kelas sentimen negatif aplikasi Indodax menunjukkan keluhan pelanggan terkait biaya penarikan yang besar, saat melakukan deposit cepat namun saat melakukan pembelian atau penjualan aset kripto server sering error dan berjalan lambat yang menyebabkan pengguna aplikasi dirugikan. Hasil tersebut dapat dijadikan acuan bagi Indodax untuk memperbaiki dan meningkatkan kualitas aplikasi sesuai dengan keluhan-keluhan yang disampaikan oleh para pengguna saat melakukan perbaruan aplikasi ke versi berikutnya.

b. *Wordcloud* dan *wordlink* sentimen positif



Gambar 3. merupakan hasil visualisasi *wordcloud* (a) dan *wordlink* (b) untuk kelas sentimen positif, gambar (a) menunjukkan hasil visualisasi kepuasan pelanggan terhadap aplikasi Indodax, dimana aplikasi dianggap bagus, baik dan mudah digunakan dalam perdagangan aset kripto. Selain itu, pengguna juga menilai proses *top up*, penarikan dan saat melakukan jual beli aset kripto direspon dengan cepat, namun pengguna juga berharap bahwa biaya penarikan uang dikurangi karena dianggap terlalu mahal, dan harapan agar memperbanyak koin atau aset kripto yang dapat diperjual-belikan di aplikasi Indodax.

5. Kesimpulan dan Saran

Klasifikasi sentimen menggunakan metode SVM pada aplikasi Indodax dengan total 1138 data ulasan yang digunakan, pembagian data latih dan data uji menggunakan tiga rasio dengan perbandingan 60:40, 70:30, dan 80:20 dengan hasil akurasi klasifikasi yang dihasilkan berturut-turut yaitu 81%, 84% dan 85%. Akurasi tertinggi terdapat pada pembagian data dengan rasio perbandingan 80:20. Dimana 910 data sebagai data latih dan 228 data sebagai data uji. Dari data uji tersebut 106 data masuk kelas negatif dan sisanya 122 data masuk pada kelas positif, dengan tingkat akurasi yang dihasilkan sebesar 85% artinya dari 228 ulasan terdapat 196 ulasan terklasifikasikan dengan benar pada kelas sentimennya. Secara umum sentimen negatif pada aplikasi Indodax menunjukkan keluhan terhadap biaya penarikan uang dari aplikasi yang dianggap mahal, sering errornya aplikasi dan lambatnya server berjalan saat jual beli aset kripto namun saat melakukan *top up* server berjalan cepat dan lancar, sehingga pengguna aplikasi merasa dirugikan. Sedangkan sentimen positif pada aplikasi Indodax berkaitan dengan kepuasan pengguna bahwa aplikasinya bagus dan mudah digunakan untuk melakukan jual beli aset kripto dan respons yang cepat dari aplikasi.

Pada penelitian ini data yang digunakan hanya ulasan yang diberikan pengguna pada bulan Oktober 2021, untuk penelitian selanjutnya diharapkan menambah data dengan jangka waktu ulasan yang lebih lama agar hasil klasifikasi yang didapatkan lebih representatif. Selain itu, diharapkan menggunakan sistem pelabelan yang dilakukan oleh tenaga ahli dibidang sentimen sehingga pemberian label pada data yang digunakan lebih akurat, serta menggunakan metode klasifikasi dan pembobotan kata yang lainnya sebagai pembanding dengan performa metode *Support Vector Machine*.

Referensi

- [1] D. L. K. Chuen, Handbook of digital currency, San Diego: Elsevier, 2015.
- [2] Bappebti, "ASET KRIPTO," 2020. [Online]. Available: https://bappebti.go.id/resources/docs/brosur_leaflet_2001_01_09_o26ulbsq.pdf. [Accessed 26 November 2021].
- [3] Bank Indonesia, "BUKU KSK NO.37: STABILITAS SISTEM KEUANGAN SEMESTER I 2021 TERJAGA," 5 Oktober 2021. [Online]. Available: https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_2325321.aspx. [Accessed 24 October 2021].
- [4] Bappebti, "Daftar Perusahaan Pedagang Aset Kripto Yang Terdaftar Di Bappebti (calon Pedagang)," 18 Februari 2021. [Online]. Available: <https://bappebti.go.id/aktualita/detail/7016>. [Accessed 20 October 2021].

- [5] Indodax, "Cerita Kami," 2021. [Online]. Available: <https://blog.indodax.com/newsroom-about-us>. [Accessed 1 October 2021].
- [6] A. S. Manek, P. D. Shenoy, M. C. Mohan and V. K. R., "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web* 20, pp. 135-154, 2017.
- [7] K. Lu and J. Wu, "Sentiment Analysis of Film Review Texts Based on Sentiment Dictionary and SVM," in *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*, Suzhou, China, 2019.
- [8] A. D. A. Putra and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *Jurnal Teknik Informatika dan Sistem Informasi*, pp. 636-646, 2021.
- [9] S. Fransiska, R. Rianto and A. I. Gufroni, "Sentiment Analysis Provider By.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Scientific Journal of Informatics*, pp. 203-212, 2020.
- [10] E. D. P. Daulay and I. Asror, "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes," *eProceedings of Engineering*, pp. 8400-8410, 2020.
- [11] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
- [12] M. M. Sya'bani and R. Umilasari, "Penerapan Metode Cosine Similarity dan Pembobotan TF/IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, pp. 31-42, 2018.
- [13] M. O. Pratama, W. Satyawan, R. Jannati, B. Pamungkas, Raspiani, M. E. Syahputra and I. Neforawati, "The sentiment analysis of Indonesia commuter line using machine learning based on twitter data," in *2018 International Conference of Computer and Informatics Engineering (IC2IE)*, 2019.
- [14] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal of Computer Engineering, System and Science)*, pp. 293-298, 2020.
- [15] S. W. Handani, D. I. S. Saputra, Hasirun, R. M. Arino and G. F. A. Ramadhan, "Sentiment Analysis for Go-Jek on Google Play Store," in *Journal of Physics: Conference Series*, 2019.
- [16] M. I. Ahmadi, F. Apriani, M. Kurniasari, S. Handayani and D. Gustian, "SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)," *Seminar Nasional Informatika (SEMNASIF)*, pp. 196-203, 2020.
- [17] R. Mahendrajaya, G. A. Buntoro and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, pp. 52-63, 2019.
- [18] F. A. Nugraha, N. H. Harani and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*, Kreatif, 2020.
- [19] G. A. Buntoro, "Analisis sentimen hatespeech pada twitter dengan metode naïve bayes classifier dan support vector machine," *Jurnal Dinamika Informatika*, pp. 1-21, 2016.
- [20] A. M. Pravina, I. Cholisoddin and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM).," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 2789-2797, 2019.
- [21] P. D. Kusuma, *Machine Learning Teori, Program, Dan Studi Kasus*, Deepublish, 2020.
- [22] D. A. Pisner and D. M. Schnyer, *Support Vector Machine*, Academic Press, 2020.
- [23] I. Cholissodin, Sutrisno, A. A. Soebroto, U. Hasanah and Y. I. Febiola, *AI, Machine Learning & Deep Learning*, Malang: Fakultas Ilmu Komputer, Universitas Brawijaya, 2020.
- [24] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning and Data Mining*, Springer, Boston, MA, 2017, p. 260.
- [25] H. Nguyen, A. Veluchamy, M. Diop and R. Iqbal, "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches," *SMU Data Science Review*, 2018.
- [26] S. Prangga, "Optimasi Parameter Pada Support Vector Machine Menggunakan Pendekatan Metode Taguchi Untuk Data High-Dimensional," (*Doctoral dissertation, Institut Teknologi Sepuluh Nopember*), 2017.
- [27] A. N. Indraini, I. Ernawati and A. Zaidah, "ANALISIS SENTIMEN TERHADAP PEMBELAJARAN DARING DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)," (*Doctoral dissertation, Universitas Pembangunan Nasional Veteran Jakarta*), 2021.
- [28] D. Y. Praptiwi, "ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA E-COMMERCE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN MAXIMUM ENTROPY," *PROGRAM STUDI STATISTIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS ISLAM INDONESIA*, 2018.
- [29] H. C. Husada dan A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *TEKNIKA*, vol. 10, pp. 18-26, 2021.

- [30] N. K. Wardhani, R. S. Kurniawan, H. Setiawan, G. Gata, S. Tohari, W. Gata dan M. Wahyudi, "SENTIMENT ANALYSIS ARTICLE NEWS COORDINATOR MINISTER OF MARITIME AFFAIRS USING ALGORITHM NAIVE BAYES AND SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION," *Journal of Theoretical and Applied Information Technology*, vol. 96, pp. 8365-8378, 2018.
- [31] V. P. Kour dan S. Arora, "Particle Swarm Optimization Based Support Vector Machine (P-SVM) for the Segmentation and Classification of Plants," *IEEE Access*, vol. 7, pp. 29374-29385, 2019.
- [32] S. T. Simanjuntak, "ANALISIS SENTIMEN PADA LAYANAN GOJEK INDONESIA MENGGUNAKAN XTREME GRADIENT BOOSTING," *PROGRAM STUDI SI TEKNOLOGI INFORMASI FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA*, 2021.
- [33] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
- [34] T. Jo, *Text Mining Concepts, Implementation, and Big Data Challenge*, vol. 45, Cham: Springer, 2019.
- [35] J. Taeho, *Text Mining: Concepts, Implementation, and Big Data Challenge*, Switzerland: Springer International Publishing AG, 2019.
- [36] A. R. Sulthana, A. K. Jaithunbi and L. S. Ramesh, "Sentiment analysis in twitter data using data analytic techniques for predictive modelling," *Journal of Physics: Conference Series*, 2018.
- [37] Y. Nurdiansyah, S. Bukhori and H. Rahmad, "Sentiment Analysis System for Movie Review in Bahasa Indonesia Using Naive Bayes Classifier Method," *Journal of Physics: Conference Series*, 2018.