

Implementasi Algoritma *Support Vector Machine* Untuk Analisis Sentimen Aplikasi OLX di *Playstore*

Aqila¹, Jeremia Jordan Sihombing², Rizky Irvandi Sitorus³, Arnita⁴

¹⁻⁴Jurusan Ilmu Komputer, FMIPA Universitas Negeri Medan, Sumatera Utara

e-mail: ¹aqila@mhs.unimed.ac.id, ²jeremiasihombing@mhs.unimed.ac.id, ³irvandisitorus@gmail.com, ⁴arnita@unimed.ac.id

ABSTRAK

Analisis sentiment adalah menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Support Vector Machine (SVM) adalah salah satu algoritma klasifikasi yang bisa digunakan untuk sentimen analisis. Penelitian ini bertujuan mengklasifikasikan review OLX Google Play Store menggunakan analisis sentimen yang telah dikumpulkan dan disaring. Hasil dari penelitian ini sangat berguna bagi pemilik aplikasi untuk mengambil keputusan di masa depan. Penelitian ini menggunakan 2500 data ulasan aplikasi OLX dari Google Play Store. Pada tahap pertama penelitian ini menggunakan Case Folding, Filtering, Tokenizing, Slang Word, Stopwords, Stemming, kemudian melakukan Konversi kalimat lalu mentransformasi teks ke vector dengan TfidfVectorizer. Pada tahap kedua melakukan split data menjadi dua bagian dengan perbandingan 80% dan 20%, yaitu 80% bagian untuk data training dan 20% bagian untuk data testing. Pada tahap terakhir membangun model sehingga diperoleh akurasi sebesar 85%, precision negatif sebesar 86% dan positif sebesar 82%, serta recall negatif sebesar 91% serta positif sebesar 73% dan f1-score sebesar 85%. Dapat disimpulkan dari sentimen analisis aplikasi OLX berdasarkan ulasan di Playstore memuat lebih banyak ulasan negatif dibandingkan ulasan positif, hal ini dapat menjadi bahan evaluasi untuk kedepannya bagi pemilik aplikasi.

Kata Kunci: Support Vector Machine, Analisis Sentimen, Ulasan, OLX

I. PENDAHULUAN (Center, Bold, 10 pt)

Media sosial adalah alat atau wadah komunikasi digunakan orang untuk mengekspresikan pandangan atau pendapat mereka tentang berbagai topik. Indonesia memiliki banyak pengguna media sosial, sehingga mendorong munculnya data tekstual tak terbatas. Salah satu kegunaan data ini adalah untuk memahami persepsi masyarakat terhadap pasar. *OLX* adalah alat atau media komunikasi melalui Internet, yang digunakan oleh orang-orang untuk membeli barang bekas atau untuk mempromosikan produk. Oleh karena itu, kehadiran *OLX* sebagai pusat belanja dan penjualan online terbesar di Indonesia telah menjadi gaya hidup sebagian besar masyarakat. Hal ini juga dapat dimanfaatkan oleh masyarakat lain untuk mendapatkan informasi tentang barang yang mereka butuhkan. Masyarakat juga bisa mengurangi barang yang sudah tidak terpakai atau sudah tidak terpakai lagi dengan cara menjualnya kepada orang lain untuk digunakan kembali tanpa kehilangan uang untuk membuangnya.

Google Play Store adalah layanan distribusi digital yang dioperasikan dan dikembangkan oleh *Google*. Tak heran jika banyak orang menyebutnya *Google Play Platform* ini sebelumnya dikenal sebagai *Android Market*. Fitur adalah toko aplikasi resmi untuk semua perangkat dengan sistem operasi *Android*. Untuk menelusuri dan mengunduh aplikasi yang dikembangkan menggunakan *SDK Android* atau kit pengembangan perangkat lunak yang diterbitkan oleh *Platform* ini tidak hanya berfungsi sebagai toko aplikasi, tetapi juga sebagai toko media digital. Tempat di mana pengguna dapat menemukan musik, film, buku, game, dan bahkan acara TV.

Analisis sentimen atau penggalian opini adalah proses memahami, mengekstrak, dan memanipulasi data teks secara otomatis menangkap informasi sentimen yang terkandung dalam kalimat opini. Analisis sentimen dilakukan untuk melihat opini atau kecenderungan opini seseorang terhadap suatu isu atau objek, apakah cenderung memiliki opini yang negatif atau positif. Analisis sentiment penting untuk mengetahui sejauh mana review masyarakat terhadap aplikasi *OLX* di *Google Play Store*.

Support Vector Machine (SVM) adalah salah satu algoritma klasifikasi yang bisa digunakan untuk sentimen analisis. *Support Vector Machine* termasuk kedalam *supervised learning*, yaitu mengenali data dari label tertentu yang telah diberikan sebelumnya dari data. SVM telah banyak digunakan dalam banyak studi klasifikasi text, salah satu kelebihanannya adalah mampu mengatasi permasalahan klasifikasi teks yang bebas linear serta apabila dibandingkan dengan model machine learning lain SVM mempunyai hasil yang lebih baik dalam opinion mining [1]. *Term Frequency Inverse Document (TF-IDF)* digunakan untuk mengubah data teks menjadi vector, dengan memperhatikan informatifnya sebuah teks tersebut.

Konsep dasar SVM sebenarnya merupakan kombinasi harmonis dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya [2]. SVM memiliki kelebihan diantaranya menentukan jarak menggunakan Support Vector Machine sehingga proses komputasi menjadi cepat [3]. SVM juga memiliki kemampuan generalisasi yang tinggi tanpa persyaratan pengetahuan tambahan, bahkan dengan dimensi yang tinggi dari ruang input [4]. Dari beberapa teknik klasifikasi yang paling sering digunakan adalah Support Vector Machine

(SVM). SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda[5]. Kelebihan Metode SVM dibandingkan metode yang lain terletak pada kemampuannya untuk menemukan hyperplane terbaik yang memisahkan dua buah class pada feature space yang ditunjang oleh strategi Structural Risk Minimization (SRM) [6].

Penelitian klasifikasi Support Vector Machine (SVM) untuk menentukan tingkat kemanisan mangga berdasarkan fitur warna yang dilakukan oleh [7] dengan menggunakan metode K-Nearest Neighbor (KNN) sebagai perbandingan, menyatakan kelebihan metode SVM yaitu mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relative sedikit. Dalam penelitian yang membandingkan metode SVM dengan metode Lexicon dalam analisis sentimen bahasa Indonesia yang dilakukan oleh [8] didapatkan bahwa metode SVM lebih diunggulkan dalam mengklasifikasikan sentimen positif dan negatif dengan akurasi sebesar 98.5% dengan parameter ekstraksi fitur unigram dengan rasio dataset 80:20. Pendekatan berbasis lexicon based kurang baik dalam melakukan analisis sentimen dengan akurasi tertinggi sebesar 78.43%. Penelitian yang dilakukan oleh [9] untuk klasifikasi rumah layak huni Desa Kidal Kecamatan Tumpang Kabupaten Malang dengan metode SVM menghasilkan rata-rata akurasi sebesar 98.75% dengan menggunakan metode pengujian K-fold Cross Validation.

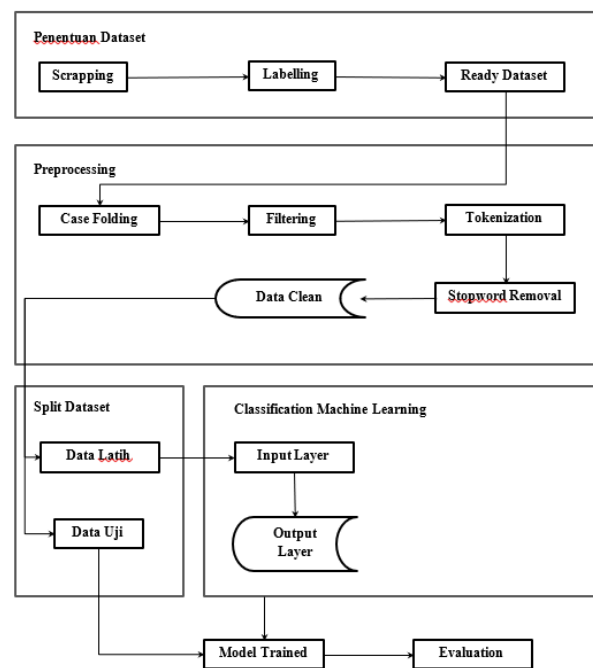
Penelitian ini akan menganalisis sentimen analisis pada review aplikasi *OLX di Playstore* menggunakan metode *Support Vector Machine (SVM)*. Penelitian ini menggunakan 2.500 data review aplikasi *olx* dari *Google Playstore*. Penelitian ini bertujuan mengklasifikasikan review aplikasi *OLX di Google Play Store* menggunakan analisis sentimen yang telah dikumpulkan dan disaring. Hasil dari penelitian ini sangat berguna bagi pemilik aplikasi untuk mengambil keputusan di masa depan.

Penelitian yang dilakukan oleh [10] mengenai klasifikasi Support Vector Machine (SVM) pada data akreditasi sekolah dasar di kabupaten Magelang menggunakan Gaussian Kernel Function RBF diperoleh hasil dengan akurasi 93.902% dari 77 sekolah dasar. Penelitian yang dilakukan oleh [11] mengenai analisis sentimen pengguna twitter terhadap pelayanan Telkom dan Biznet menggunakan metode klasifikasi Support Vector Machine (SVM) memberikan hasil nilai akurasi 79.6%, precision 76.5%, recall 72.8%, dan F1-score 74.6% untuk Telkom, serta akurasi 83.2%, precision 78.8%, recall 71.6%, dan F1-score 75% untuk Biznet. Penelitian mengenai analisis sentimen dengan menggunakan SVM pernah dilakukan oleh [12] pada lebih dari 1.000 review pengguna yang dikumpulkan dari aplikasi Grab Indonesia di Google play store. Hasil dari analisis menggunakan Support Vector Machine menghasilkan akurasi 85,54% dan Hasil Review positif yang paling sering diulas adalah "ovo", sedangkan review negatif yang paling sering diulas adalah "driver".

Selain itu [13] melakukan penelitian mengenai analisis sentimen terhadap ulasan aplikasi pejabat pengelola informasi dan dokumentasi kementerian dalam negeri republic Indonesia di *Google Playstore* menggunakan metode *Support Vector Machine* pada 700 data dengan label 85 positif dan 615 negatif. Dan hasil analisis menggunakan SVM menghasilkan rata-rata k-fold sebesar 88%, precision 94%, recall 100%, f-measure 97%, dan accuracy 97%.

II. METODE PENELITIAN

Metode penelitian ini menggunakan metode kuantitatif. Penelitian ini memiliki beberapa langkah-langkah dalam proses pengerjaan yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Proses pengumpulan data pada penelitian ini yaitu mengambil data primer yang diperoleh secara langsung dari sumber asli (tidak melalui media perantara) yang secara khusus dikumpulkan oleh peneliti untuk menjawab masalah penelitian [14]. Data diambil sejumlah 2500 ulasan pengguna dengan kategori ulasan paling relevan, karena kategori relevan menunjukkan data ulasan yang paling berkaitan dengan aplikasi dan dengan isi dari ulasan pengguna yang jelas mengenai keluhan atau kepuasan dari ulasan pengguna aplikasi *OLX* di situs *Google play* dengan metode pengambilan data scraping. Untuk melakukan scraping pada penelitian ini menggunakan ekstensi di *Google Chrome* yaitu *Scraper* [12]. Setelah data ulasan pengguna aplikasi *OLX* didapat, maka dilakukan *labelling* data dengan rating pada ulasan aplikasi *OLX* di situs *Google play* dari 0 hingga 3 dilabelkan dengan *negative*, sedangkan rating pada ulasan aplikasi *OLX* di situs *Google play* 4 dan 5 dilabelkan dengan *positive*.

Kemudian data ualsan yang sudah dilabeli disimpan kedalam file bertipe .csv.

B. Tahap Preprocessing

Pada tahap *Preprocessing* penelitian ini dilakukan menggunakan Bahasa pemrograman *Python* dan *Microsoft Excel* sebagai tempat penyimpanan data. Tahap *Preprocessing* adalah proses data yang sudah diperoleh akan di olah untuk menemukan data yang diperlukan dan membuang data-data yang tidak berguna untuk tahap analisis. Data Transformasi dengan tahapan yaitu, *Read data*, *Case Folding*, yaitu data-data tersebut akan diubah menjadi sama (*lowercase*). *Data Cleaning* dengan tahapan *Filtering*, yaitu mengeliminasi kata-kata yang tidak memiliki pengaruh atau tidak informatif. *Tokenizing*, yaitu teks pada data-data yang telah diperoleh akan dipisah untuk tiap katanya. *Slang Word* adalah mengubah kata-kata gaul – pada kasus ini, menggunakan Bahasa Indonesia. *Stopwords* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya. *Stemming* adalah mencari makna kata dasar, misalkan dalam bahasa indonesia, Contohnya kata dibuang akan diubah menjadi buang. Kemudian setelah data dibersihkan maka disimpan ulang dengan nama databersih dengan ekstensi csv. Melakukan Konversi kalimat lalu mentransformasi teks ke vector dengan *TfidfVectorizer* [12].

Metode TF-IDF menghitung nilai dari masing-masing kata di dalam dokumen menggunakan frekuensi kata (*term*) tersebut muncul. Semakin besar jumlah *term* yang muncul (*Tf* tinggi) maka semakin besar bobot dokumen [15].

$$W_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \cdot \log_2 \frac{D}{d_i}$$

[14]

- $W_{i,j}$: Pembobotan TF-IDF untuk term ke –i pada dokumen ke –j
- $n_{i,j}$: Banyaknya kemunculan term ke – i pada dokumen ke – j
- $\sum_k n_{k,j}$: Jumlah kemunculan seluruh term pada dokumen ke – j
- D : Banyaknya dokumen yang dibangkitkan
- d_i : Banyaknya dokumen yang dibangkitkan

C. Spilt Data

Berdasarkan penelitian [16], data yang dibagi dengan perbandingan 80:20 memiliki tingkat akurasi yang lebih tinggi, sehingga penelitian ini menggunakan perbandingan 20% dan 80%, yaitu 80% bagian untuk data training dan 20% bagian untuk data testing. Data *Training*, yaitu data yang akan diolah dengan metode yang akan digunakan, yang hasilnya akan digunakan sebagai prediksi untuk data *testing*. Data *Testing*, merupakan data yang akan di uji dan diprediksi pada sebuah penelitian.

D. Tahap Klasifikasi (Machine Learning)

Tahap klasifikasi dengan menggunakan metode *Support Vector Machine (SVM)*, mencari *hyperline* (batas keputusan) terbaik yang memisahkan tiap ulasan kedalam dua kelas yaitu ulasan *positif* dan ulasan *negatif* dengan menggunakan algoritma *Support Vector Machine*. Pada penelitian ini klasifikasi dengan menggunakan metode *Support Vector Machine (SVM)* merupakan model yang berasal dari teori pembelajaran statistika yang akan memberikan hasil yang lebih baik dibandingkan metode lainnya. Dalam SVM Linear, setiap data pelatihan dikenal sebagai (xi, yi), di mana i = 1,2,...,N dan xi={xi1,xi2,...,xiq} T adalah atribut untuk data pelatihan I, yi {-1, +1} adalah Kelas Label. Kemudian bagaimana *Support Vector Machine* dapat menghasilkan akurasi dari hasil klasifikasi pengujian tersebut [17].

Penelitian ini juga menggunakan *confussion matrix*. *Confusion Matrix* Merupakan tabel yang menggambarkan performa dari sebuah model atau algoritma secara spesifik. Setiap baris dari matrix tersebut, merepresentasikan kelas aktual dari data, dan setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya). Matrix tersebut dijelaskan pada Tabel 3 [18].

Tabel 3. Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

[18]

1. True Positive = Berarti seberapa banyak data yang aktual kelasnya positif, dan model juga memprediksi positif.
2. True Negative = Berarti seberapa banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.
3. False Positive = Berarti seberapa banyak data yang aktual kelasnya negatif, namun model memprediksi positif.
4. False Negative = Berarti seberapa banyak data yang aktual kelasnya positif, namun model memprediksi negatif.

Melalui 4 data tersebut, dapat diperoleh data data lain yang sangat berguna untuk mengukur perfoma sebuah model, diantaranya: Accuracy = Total keseluruhan seberapa sering model benar mengklasifikasi. Formula accuracy dapat ditulis menggunakan persamaan 1.

$$\frac{TP + TN}{Tot}$$

1. Precision = Ketika model memprediksi positif, seberapa sering prediksi itu benar. Formula precision dapat ditulis menggunakan persamaan 2.

$$\frac{TP}{FP + TP} \quad (2)$$

2. Recall (Sensitivity / True Positive Rate) = Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Formula recall dapat ditulis menggunakan persamaan 3.

$$\frac{TP}{FN + TP} \quad (3)$$

3. F1-Score = Merupakan rata-rata harmonik dari Precision dan Recall. Formula f1-score dapat ditulis menggunakan persamaan 4.

$$2 * \frac{precision * recall}{precision + recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Data yang diperoleh diambil sejumlah 2500 ulasan pengguna, yang melalui proses *scraping* dan dijadikan dalam bentuk csv. Sample hasil dari pengumpulan data dapat dilihat pada Tabel 1.

Tabel 1. Sampel hasil Pengumpulan Data

No	Ulasan
1	Aplikasi yg sangat membantu dalam "mencari" dan "berbagi" rezeki. Sy udh sering bertransaksi di aplikasi ini. Alhamdulillah, blm prnh dpt data profil yg palsu. Semua asli. Hanya saja, kudu extra hati ² meneliti barang yg akan dibeli. Kadang di foto bagus, no minus. Tapi setelah diterima, malah banyak minusnya. Semua tergantung ketelitian dan kejelian pribadi kita masing ² . Okay, happy shopping here in OLX INDONESIA. Terima kasih OLX
2	Pasang iklan semua stuck dimoderasi, sampe situ doang! Brp iklan pasang cuma nyampe dimoderasi doang. Brp hari nungguin iklannya gk tayang juga tetep dimoderasi. Waduh menuju hapus aplikasi sebelah yg gratis dan gk pake moderasi ampe kiamat kyk olx. Wkwkwk iklan saya kyknya dimoderasi sampai akhir zaman.
3	Ada apa dengan OLX kenapa iklan iklan sy banyak yg DIMODERASI sangat lama Kalau gak bisa di iklankan kasih alasan!, tapi ini hanya di beri tulisan DIMODERASI saat iklan di hapus paket yg di gunakan hangus!!!! Ini penipuan atau permainan uang
4	Kenapa harus banyak sekali yang hightlit, padahal udah difilter, alhasil filternya kayak gak fungsi, nyari 1 tipe mobil masih bnyak yg muncul tipe lain. Tolong tingkatkan
5	Bikin 2 iklan sparepart komputer, barangnya pun berbeda yg satu terdeteksi double iklan.. ga recomend buat jualan.. tidak mendukung umkm seperti ecommerce yg lain. Olx ngutamain yg berbayar.. baru di acc iklannya.. dasar aplikasi t*i

B. Hasil Preprocessing

Setelah diperolehnya data melalui *scraping* dan dijadikan csv, maka dilakukanlah preprocessing dengan beberapa tahapan *preprocessing* yaitu *Case Folding, Filtering, Tokenizing, Slang Word, Stopwords, Stemming*. Tahap *preprocessing* ini akan menghasilkan data bersih, yang didalam penelitian ini di simpan dengan ekstensi csv. Melakukan Konversi kalimat lalu mentransformasi teks ke vector dengan TfidfVectorizer. Hasil dari tahap *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengumpulan data

No	Ulasan
1	['aplikasi', 'bantu', 'cari', 'bagi', 'rezeki', 'transaksi', 'aplikasi', 'alhamdulillah', 'data', 'profil', 'palsu', 'asli', 'kudu', 'extra', 'hati', 'teliti', 'barang', 'beli', 'kadang', 'foto', 'bagus', 'minus', 'terima', 'minus', 'gantung', 'teliti', 'jeli', 'pribadi', 'masing', 'happy', 'shopping', 'indonesia', 'terima', 'kasih']
2	['pasang', 'iklan', 'stuck', 'dimoderasi', 'situ', 'iklan', 'pasang', 'dimoderasi', 'tunggu', 'iklan', 'tayang', 'dimoderasi', 'hapus', 'aplikasi', 'butuh', 'aplikasi', 'belah', 'gratis', 'pakai', 'moderasi', 'kiamat', 'tertawa', 'iklan', 'dimoderasi', 'zaman']
3	['iklan', 'iklan', 'dimoderasi', 'iklan', 'kasih', 'alas', 'tulisi', 'dimoderasi', 'iklan', 'hapus', 'paket', 'hangus', 'tipu', 'main', 'uang']
4	['hightlit', 'filter', 'alhasil', 'filter', 'fungsi', 'cari', 'tipe', 'mobil', 'muncul', 'tipe', 'tolong', 'tingkat']
5	['iklan', 'sparepart', 'komputer', 'barang', 'beda', 'deteksi', 'double', 'iklan', 'rekomendasi', 'jual', 'dukung', 'umkm', 'ecommerce', 'ngutamain', 'bayar', 'iklan', 'dasar', 'aplikasi']
6	['bayar', 'highlight', 'iklan', 'jual', 'harus', 'bayar', 'pasang', 'iklan', 'tunggu', 'pikir', 'tinggal', 'pasang', 'iklan', 'bayar', 'highlight', 'rencana', 'bayar', 'lucu', 'lihat', 'jangka', 'ubah', 'tunggu', 'tertawa', 'main', 'hilang', 'cust', 'highlight']

C. Hasil Spilt Data

Dalam proses penelitian ini data sentimen analisis dibagi menjadi 2 bagian yaitu menjadi 80% data *train* sebanyak 2000 data ulasan dan 20% data *testing* sebanyak 500 data ulasan. Rasio 80:20 digunakan pada penelitian ini dikarenakan rasio tersebut memiliki akurasi yang cukup tinggi untuk metode *Support Vector Machine*.

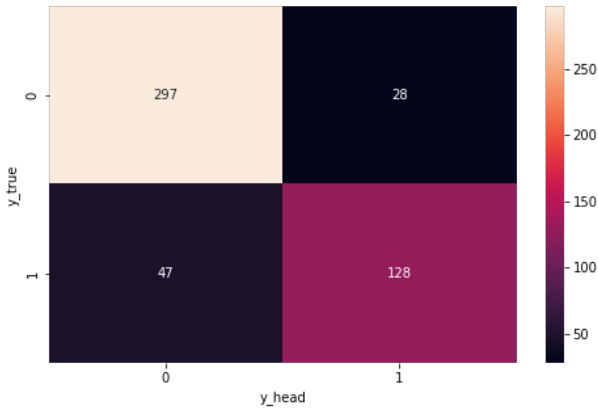
D. Hasil Tahap Klasifikasi (Machine Learning)

Setelah membangun model *machine learning* maka dapat memunculkan *output layer* yang berupa akurasi. Penelitian ini memperoleh akurasi sebesar 85% dan dapat dilihat pada Gambar 3. *Confussion matrix* dari dataset yang berisi 4 dengan nilai prediksi dan nilai actual, *confussion matrix* juga dapat melihat kevalidan dataset dan melihat keseimbangan dataset yang dimiliki. Pada penelitian ini *confussion matrixnya* dapat dilihat pada Tabel 4.

Tabel 4. Output Layer

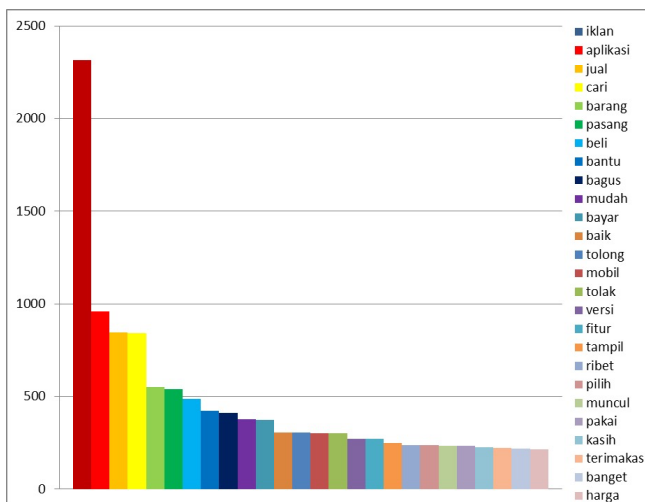
Data	Precision	Recall	f1-	Support
------	-----------	--------	-----	---------

Test			score	
Negative	0.86	0.91	0.89	325
Positive	0.82	0.73	0.77	175
accuracy			0.85	500
macro avg	0.84	0.82	0.83	500
weighted avg	0.85	0.85	0.85	500

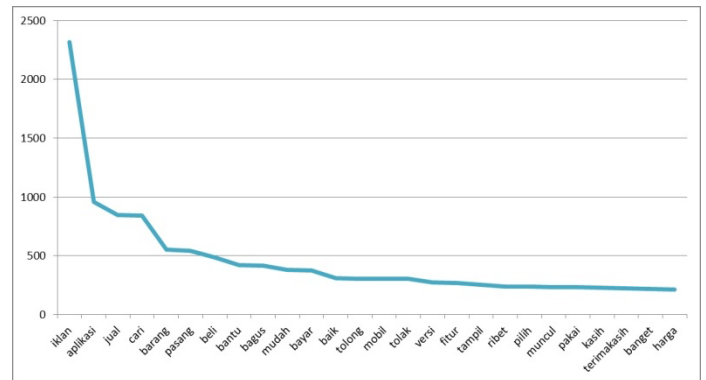


Gambar 3. Confussion Matrix

Dari Confussion Matrix di atas dapat dilihat bahwa terdapat 297 data atau berwarna cream diprediksi negative dan itu benar bahwa ulasan itu memang negative. Untuk yang berwarna hitam atau terdapat 28 data diprediksi negative dan itu salah bahwa ulasan tersebut positive. Sebanyak 47 data atau berwarna ungu kehitaman diprediksi positive dan itu salah bawah ulasan tersebut negative, dan yang terakhir berwarna ungu sebanyak 128 data diprediksi positive dan itu benar bahwa ulasan tersebut positive. Visualisasi ulasan para pengguna dapat dilihat pada Gambar 4 dan 5.



Gambar 4. Diagram Ulasan



Gambar 5. Grafik Ulasan

Pada Diagram Ulasan dan Grafik Ulasan di atas, dapat dilihat bahwa ulasan terkait dengan iklan, aplikasi, jual, dan cari menjadi ulasan yang paling banyak dilakukan oleh pengguna. Baik ulasan negative maupun ulasan positive sangat berkaitan erat dengan kata kunci di atas. Dengan ini penulis menyimpulkan bahwa pengguna mengalami masalah berkaitan dengan iklan dan aplikasi sehingga ada baiknya OLSX perlu melakukan perbaikan terkait fitur iklan dan aplikasinya.

IV. KESIMPULAN

Proses sentiment analisis sentimen pada ulasan aplikasi olx di playstore dapat dilakukan dengan menggunakan algoritma support vector Machine. Data yang digunakan pada penelitian ini didapat dari hasil data scraping sebanyak 2500 data dengan ekstensi scraper. Data yang diambil merupakan data ulasan yang paling berkaitan dengan aplikasi dan dengan isi dari ulasan pengguna yang jelas mengenai keluhan atau kepuasan dari ulasan pengguna aplikasi OLSX di situs Google play. Dalam penelitian ini, juga menunjukkan bahwa SVM menghasilkan hasil yang baik dengan memperoleh akurasi sebesar 85%. Berdasarkan hasil pengujian dan evaluasi yang dilakukan dengan pengujian precision negatif sebesar 86% dan positif sebesar 82% , serta recall negatif sebesar 91% serta positif sebesar 73% dan f1-score sebesar 85%. Penelitian ini mengalami overfitting yang disebabkan tidak seimbangnya data, dan dalam model untuk analisis sentiment pada ulasan negative pada penelitian ini mengalami akurasi yang rendah dibandingkan dengan model untuk analisis sentiment pada ulasan positive.

REFERENSI

[1] S. J. Lewis, "Thumbs up," *Am. J. Orthod. Oral Surg.*, vol. 31, no. 9, pp. 481–482, 1945, doi: 10.1016/0096-6347(45)90048-2.

[2] VAPNIK and V. N., "The Nature of Statistical Learning," *Theory*. p. 334, 1995.

[3] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine

- Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [4] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.
- [5] N. Yunita, “Analisis Sentimen Berita Artis Dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization,” *J. Sist. Inf. STMIK Antar Bangsa*, vol. 5, no. 2, pp. 104–112, 2016, [Online]. Available: www.tribunnews.com
- [6] D. A. Pisner and D. M. Schnyer, “Support vector machine,” *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 101–121, 2019, doi: 10.1016/B978-0-12-815739-8.00006-7.
- [7] M. Ichwan, I. A. Dewi, and Z. M. S., “Klasifikasi Support Vector Machine (SVM) Untuk Menentukan Tingkat Kemanisan Mangga Berdasarkan Fitur Warna,” *MIND J.*, vol. 3, no. 2, pp. 16–23, 2019, doi: 10.26760/mindjournal.v3i2.16-23.
- [8] M. Resa, A. Yudianto, A. Rahim, P. Sukmasetya, and R. A. Hasani, “Perbandingan Metode Support Vector Machine Dengan Metode Lexicon Dalam Analisis Sentimen Bahasa Indonesia,” *J. Teknol. Informasi*, vol. 6, no. 1, 2022, [Online]. Available: <https://github.com/fajri91/InSet>.
- [9] W. Agustina, M. T. Furqon, and B. Rahayudi, “Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. Vol. 2 No., no. 10, pp. 3366–3372, 2018.
- [10] P. A. Octaviani, Yuciana Wilandari, and D. Ispriyanti, “Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang,” *J. Gaussian*, vol. 3, no. 8, pp. 811–820, 2014.
- [11] B. W. Sari and F. F. Haranto, “Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [12] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [13] M. R. Fahlevvi, “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine,” vol. 4, no. 1, pp. 1–13, 2022.
- [14] F. Johnson and S. Kumar Gupta, “Web Content Mining Techniques: A Survey,” *Int. J. Comput. Appl.*, vol. 47, no. 11, pp. 44–50, 2012, doi: 10.5120/7236-0266.
- [15] F. A. Tamardina, H. Yasin, and D. Ispriyanti, “Analisis Sentimen Review Aplikasi Cryptocurrency Menggunakan Algoritma Maximum Entropy Dengan Metode Pembobotan Tf, Tf-Idf Dan Binary,” *J. Gaussian*, vol. 11, no. 1, pp. 1–10, 2022, doi: 10.14710/j.gauss.v11i1.34004.
- [16] S. W. Iriananda *et al.*, “Analisis Sentimen Dan Analisis Data Eksploratif Ulasan Aplikasi Marketplace Google Playstore,” no. Ciastech, pp. 473–482, 2021.
- [17] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Eduatic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [18] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.