

# ANALISIS SENTIMEN KUALITAS LAYANAN *ONLINE MARKETPLACE* DI INDONESIA MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Atina Za'ima Rasyida<sup>1</sup>, Indra Dharma Wijaya<sup>2</sup>, Yoppy Yunhasnawa<sup>3</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Jurusan Teknologi Informasi, <sup>3</sup>Politeknik Negeri Malang

<sup>1</sup> atinazr@gmail.com, <sup>2</sup>indra.dharma@polinema.ac.id, <sup>3</sup>yunhasnawa@gmail.com

**Abstrak**—*Online Marketplace* adalah tempat belanja *online* yang sekarang ini banyak digemari oleh masyarakat karena dinilai lebih efektif dan efisien. Respon konsumen terhadap kualitas layanan *online marketplace* digunakan untuk mengetahui tingkat kepuasan masyarakat. Penelitian ini melakukan analisis sentimen terhadap lima *online marketplace* besar di Indonesia. Analisis sentimen adalah cara yang digunakan untuk menentukan kecenderungan opini publik. Langkah pertama yang dilakukan yaitu mencari opini masyarakat dengan cara mengumpulkan data komentar pada Twitter, Play Store, dan portal *online marketplace* melalui data *scraping*. Selanjutnya, menerapkan *pre-processing* meliputi *case folding*, *filtering*, *stemming*, *tokenizing*, dan *stopwords removal*. Setelah itu, melakukan pembobotan menggunakan *tf-idf* dan pelabelan manual dengan tiga kategori yaitu positif, netral, dan negatif. Data yang sudah memiliki label tersebut dibandingkan dengan menggunakan algoritma *Support Vector Machine (SVM)* untuk mendapatkan hasil pengujian serta klasifikasi otomatis. *SVM* merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi dan regresi. Sesudah mendapatkan hasil klasifikasi otomatis, dilakukan perbandingan dengan cara menghitung *score* klasifikasi yang kemudian digambarkan dalam bentuk grafik untuk menjelaskan posisi masing-masing *online marketplace*. Pengujian dilakukan dengan melakukan penghitungan akurasi, presisi, *recall*, dan *f-measure* terhadap 1891 dataset yang menghasilkan nilai tertinggi pada percobaan data latih 90% dengan perolehan akurasi 90%, presisi 86%, *recall* 98%, dan *f-measure* 91%. Hasil penelitian ini memberikan rekomendasi terhadap masyarakat agar tidak salah dalam memilih situs yang baik untuk melakukan transaksi secara *online*.

**Kata kunci**—Analisis Sentimen, kualitas, layanan, *online marketplace*, *Support Vector Machine*.

## I. PENDAHULUAN

Perkembangan teknologi telah mengubah cara manusia dalam mengonsumsi informasi, termasuk salah satu diantaranya adalah berbelanja. Kemajuan teknologi ini membuat proses transaksi menjadi lebih mudah dengan adanya *e-commerce*. *E-commerce* dapat didefinisikan

sebagai arena terjadinya transaksi atau pertukaran informasi antara penjual dan pembeli di dunia maya [1].

Berbeda dengan *e-commerce* yang hanya menjual barang dari satu toko, *online marketplace* adalah sebuah situs yang memfasilitasi proses jual beli dari berbagai toko [2]. *Online marketplace* menjadi pilihan yang paling banyak digemari oleh masyarakat karena mempunyai banyak kelebihan, antara lain tidak perlu datang tatap muka antara penjual dan pembeli, wilayah bukan suatu batasan dalam jual beli, toko dapat diakses 24 jam, banyak diskon dan promo seperti bebas ongkos kirim.

Terlepas dari kelebihan *online marketplace*, terdapat pula beberapa kekurangan, antara lain banyaknya penipuan pada saat transaksi, penjual tidak mengirimkan barang sesuai pesanan, gambar produk yang ditampilkan tidak seperti keadaan asli, serta kemasan yang sudah rusak [3]. Masalah tersebut sangat berpengaruh terhadap kepuasan pelanggan yang seharusnya menjadi tolak ukur kualitas pelayanan dari *online marketplace* apakah sudah berjalan secara efektif dan memenuhi harapan konsumen.

Sumber data pada penelitian ini menggunakan Twitter, portal *marketplace*, dan Play Store. Hasil dari pengolahan data tersebut berupa analisis sentimen. Analisis sentimen (*opinion mining*) adalah sebuah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan [4]. Tugas dasar dalam analisis ini adalah mengelompokkan sentimen pelanggan menjadi positif, netral, dan negatif berdasarkan komentar yang ditulis oleh masyarakat kemudian membandingkan antara kelima *online marketplace* tersebut untuk mengetahui posisi masing-masing perusahaan.

Penelitian ini menggunakan metode *Support Vector Machine (SVM)* karena metode tersebut merupakan salah satu metode unggulan dari *machine learning* yang memiliki hasil lebih baik dalam hal klasifikasi dan prediksi [5]. Diharapkan penelitian ini mampu membantu masyarakat untuk mengetahui *online marketplace* mana yang lebih baik sehingga mereka dapat merasa aman dan nyaman saat melakukan transaksi, sedangkan bagi perusahaan *online*

marketplace dapat memperoleh masukan inti dan mendorong peningkatan kualitas pelayanan mereka.

## II. TINJAUAN PUSTAKA

### A. Penelitian Terdahulu

Nur Azizah Vidya dari Universitas Indonesia melakukan penelitian tesis mengenai sentimen analisis terhadap brand reputation dengan menggunakan studi kasus PT. XL Axiata melalui Twitter dengan membandingkan tiga algoritma klasifikasi: Naïve bayes, Support Vector Machine, dan Decision Tree. Tahap evaluasi performansi menggunakan precision, recall, f-measure, dan kurva ROC (AUC). Hasil menunjukkan bahwa model yang dibentuk oleh SVM memberikan performansi yang lebih baik [6].

Petrix Nomleni dari Institut Teknologi Sepuluh Nopember juga melakukan penelitian tesis mengenai data media center yang diambil dari Twitter dan Facebook sapawarga Pemerintah Kota Surabaya menggunakan metode Support Vector Machine. Penelitian ini ditujukan untuk memberikan informasi yang efisien kepada masyarakat dan pemerintah agar mengetahui bidang mana yang perlu dibenahi dalam pembangunan. Pengujian dilakukan dengan perhitungan precision, recall, f-measure, serta akurasi dengan menghasilkan rata-rata diatas 80% dengan akurasi tertinggi 84,4086%, precision 81%, recall 84%, serta f-measure 80% [7].

Metode Support Vector Machine kembali digunakan oleh Ayundha Wulan Kurniawati dari Politeknik Negeri Malang dengan mengambil judul skripsi “Implementasi Metode Support Vector Machine Untuk Identifikasi Penyakit Daun Tanaman Kubis”. Tingkat keberhasilan identifikasi penyakit daun tanaman kubis dengan menggunakan metode SVM dipengaruhi oleh proses segmentasi dan parameter masukan yang digunakan saat training. Dengan proses segmentasi yang baik akan menghasilkan klasifikasi yang lebih bagus, penggunaan parameter terbaik mampu meningkatkan hasil klasifikasi, dengan menggunakan Histogram Equalization dan nilai parameter terbaik untuk proses perhitungan training hasil akurasi yang didapatkan sebesar 80.55% [8].

### B. Analisis Sentimen

Analisis sentimen adalah studi komputasi mengenai sikap, emosi, pendapat, penilaian, pandangan dari sekumpulan teks yang fokusnya adalah mengekstraksi, mengidentifikasi atau menemukan karakteristik sentimen dalam unit teks menggunakan metode NLP (Natural Language Processing), statistik atau machine learning [7].

Nilai dari analisis sentimen bisa dipecah menjadi 3 yaitu sentimen positif, sentimen negatif, dan sentimen netral atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif.

### C. Text Mining

Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data yang belum terstruktur. Text mining memiliki sifat yang mirip dengan data mining, tetapi dengan fokus pada teks daripada bentuk data yang lebih terstruktur seperti dokumen word, PDF, kutipan teks, sedangkan input untuk data mining adalah data yang terstruktur [9].

### D. Web Scraping

Pengumpulan data dilakukan dengan cara *web scraping* yaitu mengambil data *review* dari *website* target. Tujuan dari *web scraping* yaitu untuk mencari informasi tertentu, mengekstraksi informasi tersebut, kemudian menghimpunnya dalam halaman *web* yang baru. Beberapa *web scraping* memfokuskan dirinya untuk mengubah data yang tidak terstruktur dan menyimpannya ke dalam data yang terstruktur.



Gambar 1. Proses Web Scraping

### E. TF-IDF

TF (*Term Frequency*) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan. IDF (*Inverse Document Frequency*) merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. Semakin besar jumlah kemunculan suatu term (TF tinggi) dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar.

$$W_{ij} = tf_{ij} \times \log \frac{D}{df_j} \quad (1)$$

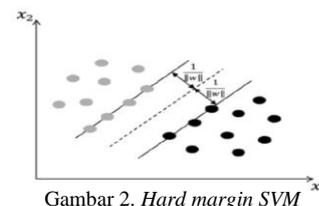
### F. Support Vector Machine (SVM)

SVM dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. SVM merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi (Damila, Mauliza, & Ula, 2019).

Prinsip utama penggunaan SVM adalah mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada ruang input. SVM merupakan salah satu machine learning yang melakukan pelatihan dengan menggunakan training dataset dan melakukan generalisasi dan membuat prediksi dari data baru.

#### a) SVM Linear

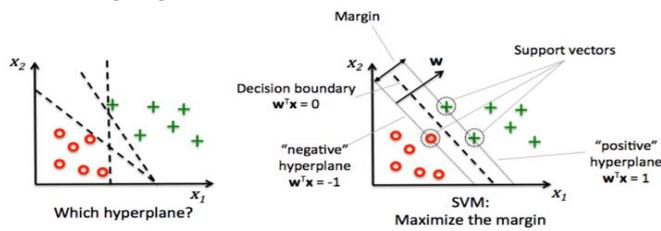
Teknik SVM merupakan *classifier* yang menemukan hyperplane dengan kasus data yang digunakan merupakan data dengan dua kelas yang sudah terpisah secara *linear* seperti pada gambar berikut.



Gambar 2. Hard margin SVM

Berdasarkan pada Gambar 2.2 diatas, terlihat bahwa antara kelas positif dan kelas negatif sudah terpisah secara total terlihat dari lingkaran abu – abu yang berada dekat

dengan garis  $x_2$  sedangkan untuk lingkaran hitam terletak dekat dengan garis  $x_1$ .

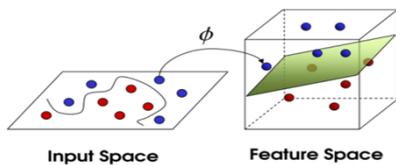


Gambar 3. Hyperplane terbaik yang memisahkan antar dua kelas positif (+1) dan negatif (-1)

Berdasarkan gambar di atas terlihat beberapa pola yang merupakan anggota dari dua buah kelas yaitu positif (+1) dan negatif (-1). *Hyperplane* terbaik dapat ditemukan dengan mengukur *margin hyperplane* dan mencari titik maksimalnya. *Margin* merupakan jarak antara *hyperplane* dengan data terdekat dari masing – masing kelas. *Subset training dataset* yang paling dekat dinamakan sebagai *support vector*. Pada Gambar 3 sebelah kanan menunjukkan *hyperplane* terbaik, yaitu terletak pada garis putus – putus yang berada tepat ditengah – tengah *hyperplane* positif dan *hyperplane* negatif. Sedangkan tanda “positif” dan “bulat” yang berada dalam lingkaran hitam merupakan *support vector*. Berikut merupakan persamaan dari *SVM linear*.

b) *Kernel Support Vector Machine*

Ketika terdapat permasalahan data yang tidak terpisah secara linear dalam ruang input, soft margin SVM tidak dapat menemukan *hyperplane* pemisah yang kuat yang meminimalkan misklasifikasi dari data points serta menggeneralisasi dengan baik. Untuk itu, kernel dapat digunakan untuk mentransformasi data ke ruang berdimensi lebih tinggi yang disebut sebagai ruang kernel, dimana akan menjadikan data terpisah secara linear.



Gambar 4. Kernel SVM untuk memisahkan data secara linear

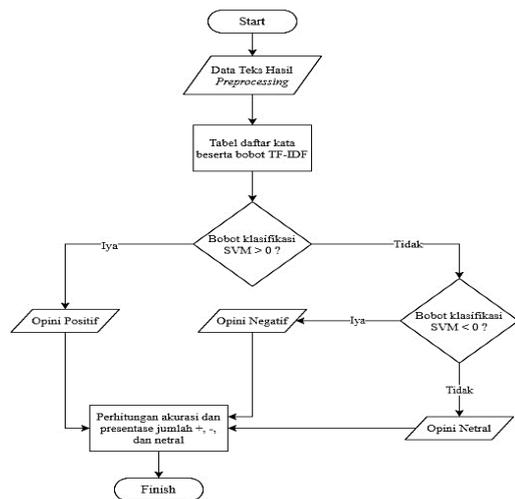
Data disimpan dalam bentuk kernel yang mengukur kesamaan atau ketidaksamaan objek data. Kernel dapat dibangun untuk berbagai objek data 18 mulai dari data kontinu dan data diskrit melalui urutan data dan grafik. Konsep substitusi kernel berlaku bagi metode lain dalam analisis data.

TABEL I. KERNEL YANG UMUM DIGUNAKAN

Jenis Kernel	Rumus
Polynomial	$K(x_i, x_j) = (x_i \cdot x_j + 1)^p$ (2)
Gaussian RBF	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$ (3)
Sigmoid	$K(x_i, x_j) = \tanh(\alpha x_i \cdot x_j + \beta)$ (4)

c) *Proses Klasifikasi Menggunakan Support Vector Machine*

Training pada klasifikasi *SVM* akan menghasilkan sebuah nilai atau pola yang akan digunakan pada proses testing yang bertujuan untuk pemberian label sentimen. Penilaian kemudian dibuat dengan menilai score yang merepresentasikan di sisi mana dokumen itu berada. Proses pengambilan keputusan dengan *SVM* beserta analisis berupa tingkat akurasi dan jumlah dokumen di setiap kelas positif, negatif, dan netral digambarkan seperti pada gambar berikut.

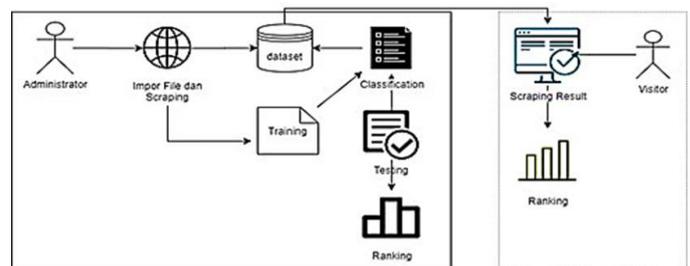


Gambar 5. Flowchart klasifikasi dengan SVM dan Analisis

III. PERANCANGAN DAN IMPLEMENTASI

A. *Deskripsi Sistem*

Sistem terbagi menjadi dua sisi yaitu admin dan pengunjung. Dibawah ini adalah gambaran dari deskripsi sistem.

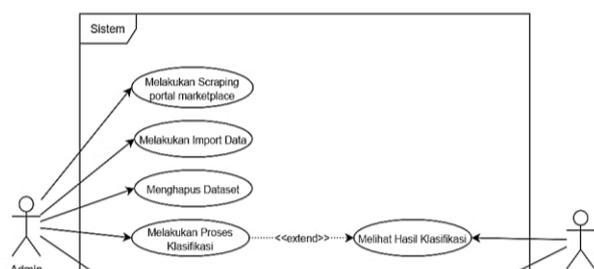


Gambar 7. Proses General Sistem

Pada sisi admin dapat melakukan pengolahan aplikasi seperti melakukan scraping dan impor data, pengelolaan data, pengklasifikasian, pengujian, dan perankingan. Sedangkan pada sisi pengunjung untuk menampilkan hasil ranking dan klasifikasi data hasil scraping.

B. *Use Case Diagram*

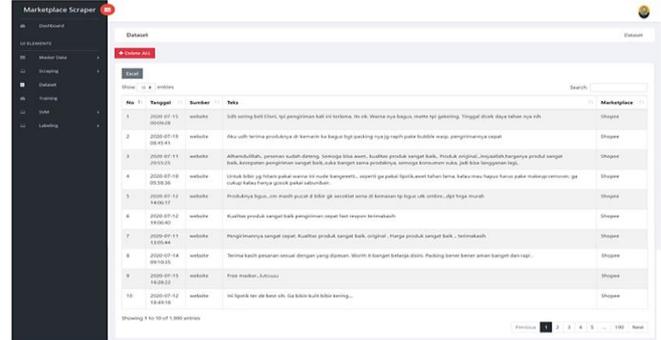
*Use case* pada sistem ini terdiri dari dua aktor, dimana aktor pertama adalah *administrator* dan aktor kedua adalah pengunjung.



kesamaan antara label dan klasifikasi akan meningkatkan tingkat akurasi dari sistem yang dihasilkan.

#### D. Implementasi Database

Implementasi *database* dengan nama *svm\_marketplace* memiliki tujuh tabel dan telah sesuai dengan analisis dan



Gambar 8. Use Case Diagram

#### C. Analisis Sistem

Sistem yang akan dikembangkan dapat melakukan pengklasifikasian data *review* ke dalam positif, negatif, dan netral. Sistem tersebut merupakan satu kesatuan proses yang dapat menghasilkan hasil klasifikasi sesuai dengan data yang dimasukkan.

##### a. Tahap Awal

Tahap awal merupakan proses awal untuk mengolah dataset sebelum dapat digunakan untuk proses training dan testing. Data diambil melalui proses *scraping*. Proses selanjutnya yaitu pemberian label secara manual. Data *review* akan dibagi ke dalam positif, negatif, dan netral. Pemberian label dibantu oleh 3 orang. Pelabelan tersebut bertujuan untuk memberikan klasifikasi secara manual terhadap tweet yang telah diperoleh untuk proses training dan testing. Proses yang selanjutnya adalah preprocessing. Tahap ini memiliki manfaat untuk menghasilkan data masukan yang baik untuk proses training dan testing. Tahap preprocessing adalah tahap untuk menyeleksi kata-kata yang ada pada data *review* sehingga menghasilkan kata-kata yang berisi sentimen dengan membuang kata-kata yang tidak diperlukan.

##### b. Training Dataset

Setelah preprocessing, tweet akan dihitung nilai *TF IDF* yang dihasilkan. *TF IDF* merupakan proses untuk memecah tweet menjadi kata kata berdasarkan frekuensi kata yang digunakan dan nilai dari kata yang digunakan tersebut. Nilai dari setiap tweet akan menghasilkan nilai yang dapat digunakan untuk data masukan menuju ke proses training dan testing.

##### c. Testing Dataset

Proses testing adalah proses untuk menghasilkan klasifikasi data berdasarkan model *classifier* yang telah dihasilkan dari proses training dataset. Dataset yang digunakan sebagai data masukan berupa nilai TF IDF yang telah dihitung pada tahap awal dan label yang dimiliki oleh data *review*. Data akan diambil secara acak dari seluruh data *review* yang ada. Hasil dari testing dataset adalah klasifikasi *review* ke dalam positif, negatif, atau netral.

##### d. Pengujian

Pengujian dari sistem yang dihasilkan dengan cara menghitung tingkat akurasi sistem. Tingkat akurasi didapatkan dengan membandingkan label dan hasil klasifikasi dari tweet pada proses testing dataset. Banyaknya

#### E. Implementasi Tampilan

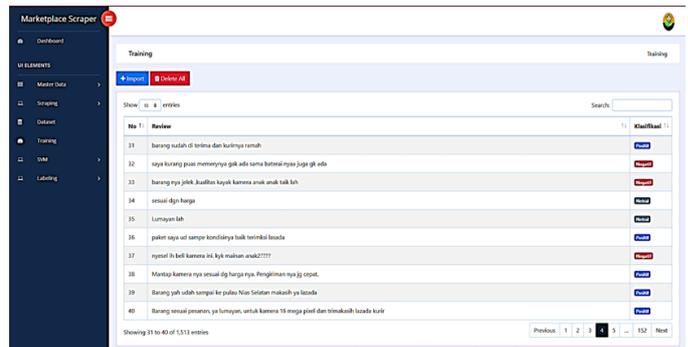
##### a) Dataset

Pada tampilan menu dataset terdapat tabel yang berisi review hasil impor dan scraping, tombol delete all berfungsi untuk menghapus keseluruhan dataset, tombol excel digunakan untuk mengunduh keseluruhan dataset.

Gambar 10. Halaman dataset

##### b) Training

Pada tampilan halaman *training* berisi informasi dari dataset yang digunakan untuk proses data latih. Data ditampilkan dalam bentuk tabel dengan tiga kolom yaitu nomor, review, dan klasifikasi.

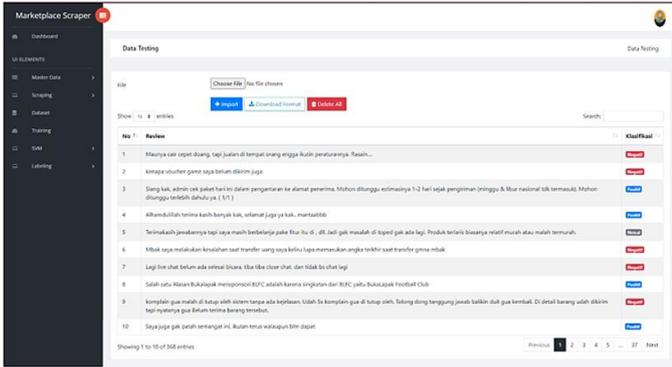


Gambar 11. Halaman training

##### c) Testing

Pada tampilan halaman *testing* berisi informasi dari dataset yang digunakan untuk proses data uji. Data

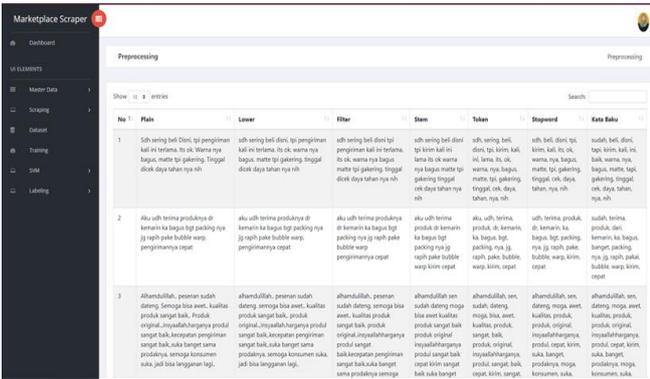
ditampilkan dalam bentuk tabel dengan tiga kolom yaitu nomor, review, dan klasifikasi.



Gambar 12. Halaman testing

d) Preprocessing

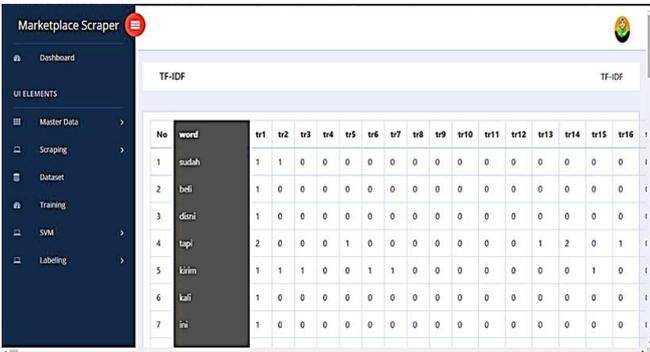
Pada tampilan preprocessing berisi informasi hasil dari tweet yang telah mengalami preprocessing sesuai dengan ketentuan. Terdapat tabel dengan beberapa kolom yang berisi plain, lower, filter, steam, token, stopword, dan kata baku.



Gambar 13. Halaman pre-processing

e) TF-IDF

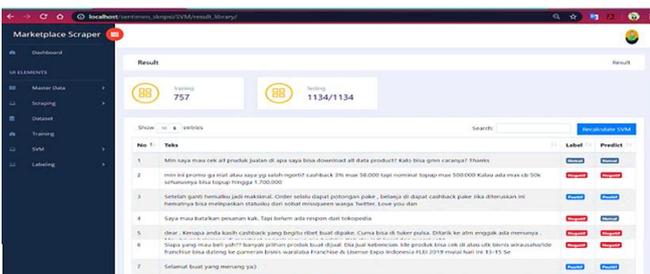
Pada menu TF IDF berisi informasi tentang kata-kata yang telah dipecah dari kalimat setiap data review dan telah memiliki nilai TF IDF sendiri.



Gambar 14. Halaman tf-idf

f) Pengujian (Result)

Pada tampilan pengujian berisi tabel perbandingan antara klasifikasi manual dan otomatis. Terdapat informasi hasil akurasi, presisi, recall, dan f-measure.



Gambar 15. Halaman pengujian

IV. PENGUJIAN DAN PEMBAHASAN

A. Pengujian Fungsionalitas

Uji coba fungsional terhadap layanan yang disediakan oleh sistem memiliki tujuan untuk mengetahui hasil dari setiap proses dalam klasifikasi telah sesuai dengan analisis dan perancangan yang dilakukan.

B. Pengujian Akurasi

Pada pengujian ini digunakan 1.891 data yang akan digunakan untuk training dan testing.

TABEL II. PERBANDINGAN DATA LATIH DAN DATA UJI

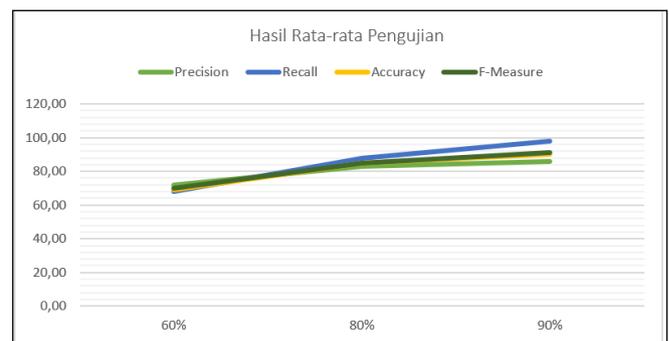
Pengujian	Perbandingan	Data Latih	Data Uji
1	60%: 40 %	1.135	756
2	80%:20%	1.513	378
3	90%:10%	1.702	189

Dari masing-masing percobaan tersebut digunakan perhitungan dengan confusion matrix dan didapatkan hasil seperti berikut.

TABEL III. HASIL PENGUJIAN AKURASI SISTEM

Data Training 60%			
Presisi	Presisi	Presisi	f-measure
72%	68%	69%	70%
Data Training 80%			
Presisi	Presisi	Presisi	f-measure
83%	83%	83%	83%
Data Training 90%			
Presisi	Recall	Akurasi	f-measure
86%	98%	90%	91%

Berdasarkan pengujian yang telah dilakukan, diambil rata-rata dari hasil setiap percobaan tersebut. Berikut adalah hasil representasi nilai presisi, recall, akurasi, dan f-measure dalam gambar grafik.



Gambar 16. Grafik perbandingan hasil pengujian

Sesuai dengan tujuan penelitian ini dibuat, yaitu untuk memberikan informasi berupa rekomendasi marketplace kualitas pelayanan terbaik. Nilai rekomendasi didapatkan dari perolehan jumlah hasil klasifikasi seperti berikut.

TABEL IV. HASIL PEROLEHAN SCORE TIAP MARKETPLACE

Dari perhitungan yang telah dilakukan, score nilai untuk marketplace S lebih unggul bila dibandingkan dengan marketplace lain. Untuk selanjutnya disusul oleh Bi, BL, L, dan terakhir adalah T.

Grafik hasil perhitungan score ditunjukkan seperti gambar di bawah ini.



Gambar 17. Grafik Ranking Marketplace

## V. SIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dapat ditarik simpulan sebagai berikut :

1. Algoritma *Support Vector Machine* dapat digunakan untuk klasifikasi kualitas pelayanan *online marketplace*.
2. Berdasarkan hasil perhitungan, *online marketplace S* mendapat *score* tertinggi.
3. Dari tiga kali pengujian menggunakan *metode Support Vector Machine (SVM)* diperoleh nilai tertinggi pada percobaan data training 90% dan data uji 10% dengan hasil presisi 86%, *recall* 98%, akurasi 90%, dan *f-measure* 91%.
4. Banyak sedikitnya *term* yang sama antara data latih dan data uji sangat berpengaruh terhadap hasil akurasi.

## DAFTAR PUSTAKA

- [1] R. R. Rerung, *E-Commerce Menciptakan Daya Saing Melalui Teknologi Informasi*, Yogyakarta: Deepublish, 2018.
- [2] I. Mubarak, 18 July 2018. [Online]. Available: <https://www.niagahoster.co.id/blog/marketplace-adalah/>.
- [3] Muljono, P. D. Artanti, A. Syukur, A. Prihandono and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," *Konferensi Nasional Sistem Informasi 2018*, pp. 8-9, 2018.
- [4] D. Gore and B. S. D., "Sentiment Analysis on Twitter Data Using Support Vector Machine," *International Journal of Computer Science Trends and Technology (IJCTST)*, vol. 4, no. 3, pp. 365-370, 2016.

- [5] S. Prangga, "Optimasi Parameter Pada Support Vector Machine Menggunakan Pendekatan Metode Taguchi Untuk Data High-Dimensional," 2017.
- [6] N. A. Vidya, "Twitter Sentiment Analysis Terhadap Brand Reputation : Studi Kasus PT XL Axiata Tbk," *Jurnal Fasilkom UI*, 2015.

Marketplace	positif	netral	negatif	Jumlah (positif+netral-negatif)
S	17,5%	4,3%	8%	13,7%
T	5,5%	3,6%	7,9%	1,2%
Bi	8,1%	1,4%	6%	3,5%
L	8,2%	1,4%	8%	1,6%
BL	9%	1%	8%	2%

- [7] P. Nomleni, "Sentimen Analisis Menggunakan Support Vector Machine (SVM)," pp. 5-6, 2015.
- [8] A. R. T. H. Ririd, A. W. Kurniawati and Y. Yunhasnawa, "Implementasi Metode Support Vector Machine Untuk Identifikasi Penyakit Daun Tanaman Kubis," *Jurnal Informatika Polinema*, pp. 181-189, 2018.
- [9] H. Nindito, "Teori Text Mining dan Web Mining," 17 July 2018. [Online]. Available: <https://sis.binus.ac.id/2016/12/15/teori-text-mining-dan-web-mining/>.