

Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan *Support Vector Machine*

Yoga Tika Pratama¹, Fitra Abdurrachman Bachtiar², Nanang Yudi Setiawan³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹yogatikap@gmail.com, ²fitra.bachtiar@ub.ac.id, ³nanang@ub.ac.id

Abstrak

Dinas Pariwisata dan Kebudayaan (Disparbud) Kabupaten Malang membutuhkan perspektif pelanggan dalam proses pengawasan dan evaluasi kegiatan pengelolaan dan pengembangan pariwisata pantai Malang Selatan. Akan tetapi Disparbud tidak memiliki data opini pelanggan, maupun penerapan teknologi pengolahan dan analisis data yang mampu memberikan informasi tentang perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan. Perspektif pelanggan bisa didapatkan dari opini pelanggan dalam bentuk ulasan yang ada pada *platform Web*. Metode *Web Scraping* adalah salah satu metode untuk melakukan ekstraksi data opini pelanggan yang bersumber dari *platform Web*. Untuk mendapatkan perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan bisa dengan melakukan Analisis Sentimen pada tingkat aspek. Salah satu metode klasifikasi yaitu *Support Vector Machine* yang dapat digunakan untuk melakukan klasifikasi sentimen dalam proses Analisis Sentimen tersebut. Pada Penelitian ini berhasil mendapatkan 674 data opini pelanggan dalam Bahasa Indonesia dari tahun 2013 hingga tahun 2018 untuk 43 objek pariwisata pantai Malang Selatan yang bersumber dari situs *TripAdvisor*. Pengujian terhadap hasil klasifikasi sentimen terhadap setiap aspek menunjukkan hasil yang baik dengan rata-rata *Accuracy* sebesar 85%, *Precision* sebesar 85%, *Recall* sebesar 87%, dan *F1-Score* sebesar 85%. Luaran dari penelitian ini berupa visualisasi data dalam bentuk *Dashboard* yang terdiri dari 4 visualisasi *Dashboard* yang memuat hasil dari Analisis Sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan.

Kata kunci: *Web Scraping, Analisis Sentimen, Classification, SVM, Visualisasi Data*

Abstract

Department of Tourism and Culture (Disparbud) of Malang Regency requires the customer's perspective in the process of monitoring and evaluation of management and development activities of South Malang beach tourism. But Disparbud neither have the data about the customer's opinion, nor the application of technology on data processing and data analysis that capable of providing information about customer's perspective on aspects of South Malang beach tourism. The customer's perspective can be obtained from customer's opinions in the form of review that available on the web platform. Web Scraping is a method for extraction of customer's opinion data sourced from the web platform. To get the information about customer's perspective on aspects of South Malang beach tourism could be done by performing Aspect Level Sentiment Analysis on the customer's opinion data. One of many classification methods is the Support Vector Machine, which can be used to classify sentiments in sentiment analysis process. This research managed to get 674 customer's opinion data in Bahasa Indonesia ranged from 2013 to 2018 for 43 tourist attractions of South Malang beach tourism sourced from *TripAdvisor*. The tests that have been conducted on the results of the sentiment classification for all aspects resulted good results in an average of 85% on accuracy, 85% on precision, 87% on recall, and 85% on *F1-Score*. The output of this research is in form of data visualization in the form of *Dashboard* consisted of 4 *Dashboards* which contains results from the process of sentiment analysis on the customer opinions to aspects of South Malang beach tourism.

Keywords: *Web Scraping, Sentiment Analysis, Classification, SVM, Data Visualization*

1. PENDAHULUAN

Disparbud Kabupaten Malang melalui Perusahaan Daerah (PD) Jasa Yasa dalam menjalankan fungsinya memiliki strategi dalam 3 perspektif yakni perspektif sumber daya, pelanggan, dan keuangan. Strategi perspektif pelanggan bertujuan untuk memberikan kepuasan dan kenyamanan dengan tujuan meningkatkan dan memaksimalkan pemenuhan kebutuhan pelanggan. Disparbud Kabupaten Malang selaku pembina PD Jasa Yasa, bertanggung jawab dalam melakukan pengawasan dan evaluasi terhadap kegiatan manajemen dan pengembangan pariwisata, termasuk di dalamnya penerapan ketiga strategi tersebut. Berdasarkan hasil wawancara dengan pegawai Disparbud Kabupaten Malang menyatakan bahwa terdapat beberapa masalah dalam pemenuhan strategi perspektif pelanggan. Permasalahan tersebut adalah Disparbud tidak memiliki data mengenai perspektif pelanggan maupun penerapan teknologi pengolahan dan analisis data yang mampu memberikan informasi tentang perspektif pelanggan terhadap objek pariwisata yang ada di Kabupaten Malang. Selain itu, penerapan teknologi komputer sebatas pada menyimpan data, sehingga belum terdapat visualisasi yang mampu menggambarkan perspektif pelanggan terhadap objek pariwisata yang ada di Kabupaten Malang guna mempermudah Disparbud dalam melakukan pengawasan dan evaluasi penerapan strategi perspektif pelanggan.

Dari masalah yang dijelaskan pada Disparbud Kabupaten Malang maka perlu untuk mendapatkan data opini pelanggan. Kemudian melakukan Analisis Sentimen terhadap data yang berisi opini pelanggan tersebut untuk dapat mengetahui perspektif pelanggan terhadap aspek objek pariwisata pantai Malang Selatan. Hasil pengolahan dan analisis data perlu untuk divisualisasikan agar dapat membantu dalam proses pengambilan keputusan strategis.

Pada penelitian terdahulu oleh Marrese-Taylor et al. (2013) telah melakukan penelitian dengan menerapkan Analisis Sentimen (*Opinion Mining*) pada tingkat aspek dan menerapkannya pada domain pariwisata. Hasil dari penelitian memberikan kesimpulan bahwa ulasan mengenai produk pariwisata yang tersedia di *Web* memiliki informasi berharga mengenai preferensi wisatawan terkait produk pariwisata dan bisa dilakukan ekstraksi terhadapnya dengan

menggunakan pendekatan Analisis Sentimen pada tingkatan aspek. Wijayanti dan Arisal (2017) telah dilakukan penelitian terkait Analisis Sentimen terhadap teks berbahasa Indonesia dengan tujuan mengevaluasi algoritma *Machine Learning* dalam menentukan polaritas dari teks. Dalam penelitian ini juga menunjukkan TF-IDF yang dikombinasikan dengan SVM menghasilkan akurasi hingga 91,42%. Fachrina dan Widyantoro (2017) telah dilakukan penelitian terkait Analisis Sentimen terhadap ulasan *Online Marketplace* dengan mengkombinasikan pendekatan *Rule-based* dan *Machine Learning*. Hasil dari penelitian tersebut SVM yang dikombinasikan dengan pendekatan *Rule-based* menunjukkan performa terbaik dengan *F-Measure* berkisar antara 78,9% hingga 92%. Dari tiga penelitian tersebut disimpulkan bahwa Analisis Sentimen pada tingkat aspek bisa diterapkan dalam domain pariwisata untuk mendapatkan preferensi wisatawan terkait produk pariwisata serta penggunaan SVM memberikan hasil yang paling optimal dalam Analisis Sentimen pada tingkat aspek.

Berdasarkan latar belakang yang telah dijelaskan, penulis tertarik untuk melakukan penelitian untuk melakukan Analisis Sentimen terhadap opini pada *platform Web* yang diberikan oleh pelanggan tentang objek wisata pantai Malang Selatan. Akuisisi data opini bersumber dari situs TripAdvisor dengan memanfaatkan teknik *Web Scraping*. Analisis Sentimen dilakukan pada level aspek dengan menggunakan metode SVM dengan menggunakan TF-IDF sebagai *Term Weighting* dan ditampilkan melalui visualisasi *Dashboard*. Hasil visualisasi ini diharapkan dapat memudahkan pemangku kebijakan dalam memahami perspektif pelanggan serta mendapatkan wawasan yang mendukung untuk pengambilan keputusan strategis terkait aspek wisata pantai Malang Selatan.

2. LANDASAN KEPUSTAKAAN

2.1 Web Scraping

Mitchell (2018) mendefinisikan bahwa yang dimaksudkan sebagai *Web Scraping* adalah kegiatan pengumpulan (*gathering*) data yang bersumber dari Internet. Tujuan dari *Web Scraping* adalah mendapatkan data untuk kemudian melakukan ekstraksi informasi yang dimiliki oleh data tersebut. Cara kerja *Web Scraping* adalah dengan mengakses halaman

Web, memilih elemen data yang ada dalam halaman tersebut, melakukan ekstraksi dan transformasi bila diperlukan, dan terakhir menyimpan data tersebut menjadi *dataset* terstruktur (Boeing dan Waddell, 2017).

2.3 Analisis Sentimen pada Text Mining

Text Mining adalah salah area studi yang menggabungkan teknik dari *Data Mining*, *Machine Learning*, *Natural Language Processing* (NLP), *Information Retrieval* serta *Knowledge Management*. Tujuan dari *Text Mining* tidak jauh berbeda dengan *Data Mining* yang tidak lain adalah menemukan informasi/pengetahuan yang berguna dari sumber data yang besar dengan cara mengidentifikasi dan melakukan eksplorasi terhadap pola yang menarik dari sumber data. Adapun perbedaan mendasar antara keduanya adalah terletak pada bentuk dari data yang dijadikan sumber dalam proses pencarian *knowledge*. Dalam *Text Mining* menggunakan data tidak terstruktur dalam bentuk teks (Feldman & Sanger, 2007).

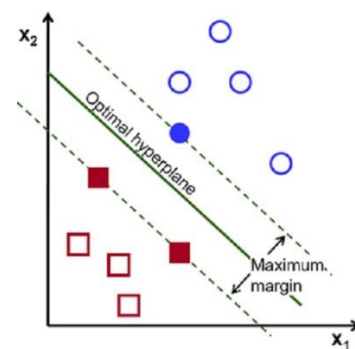
Analisis Sentimen atau sering disebut juga *Opinion Mining* adalah bidang studi yang menganalisis opini, sentimen, evaluasi, perilaku, dan emosi yang tertuju kepada entitas tertentu seperti produk, layanan, organisasi, dan sebagainya (Liu, 2012). Adapun tujuan dari Analisis Sentimen adalah menemukan kuintupel $(e_i, a_{ij}, s_{jkl}, h_k, t_l)$ dalam sebuah opini. Kuintupel tersebut terdiri dari nama dari entitas (e_i), aspek dari entitas e_i (a_{ij}), sentimen terhadap aspek a_{ij} dari entitas e_i (s_{jkl}), pemegang opini (h_k), dan waktu opini tersebut diungkapkan oleh pemegang opini (t_l). Berikut adalah tahapan dalam melakukan Analisis Sentimen:

1. Ekstraksi dan pengkategorian entitas adalah dengan melakukan ekstraksi terhadap semua entitas yang ada dalam sebuah dokumen berisi opini, lalu mengelompokkannya kedalam berbagai kategori yang representatif.
2. Ekstraksi dan pengkategorian aspek adalah dengan melakukan ekstraksi terhadap semua aspek yang dimiliki oleh entitas, lalu mengelompokkannya kedalam beberapa *cluster*.
3. Ekstraksi dan kategorisasi pemegang opini adalah dengan melakukan ekstraksi terhadap semua pemegang opini yang ada dalam dokumen berisi opini atau data terstruktur,

4. Ekstraksi dan penyeragaman waktu adalah dengan melakukan ekstraksi waktu terkait kapan opini tersebut diberikan, dan melakukan standarisasi atau penyeragaman terhadap format waktu.
5. Klasifikasi sentimen terhadap aspek adalah dengan menentukan opini terhadap suatu aspek dari suatu entitas ke dalam kategori positif, negatif atau netral, atau menggunakan angka dalam bentuk *rating* terhadap aspek.
6. Menghasilkan kuintupel opini dengan mendefinisikan semua kuintupel opini $(e_i, a_{ij}, s_{jkl}, h_k, t_l)$ yang ada dalam suatu dokumen berisi opini.

2.4 Support Vector Machine

Han et al. (2012) menjelaskan bahwa *Support Vector Machine* adalah metode klasifikasi untuk mencari *Maximum Marginal Hyperplane* (MMH) atau pemisah terbaik untuk menciptakan pemisahan maksimal untuk semua kelas. *Margin* bisa didefinisikan sebagai jarak terpendek dari sebuah *hyperplane* terhadap satu sisi dari *margin* itu sama dengan jarak *hyperplane* dengan sisi *margin* lainnya, dengan catatan kedua *margin* tersebut dalam posisi paralel dengan *hyperplane*. Seperti yang ditunjukkan dalam Gambar 1. Kemudian dapat disimpulkan bahwa semakin besar *margin* bisa dikatakan memiliki akurasi yang lebih tinggi.



Gambar 1 Optimal Hyperplane

SVM menggunakan *set* data latih dalam bentuk (X_i, y_i) X_i adalah *tuple* dan y_i adalah label kelas dengan $i=1....N$, $X_i \in R^d$ dan $y_i \in \{-1, 1\}$. Tujuan dari SVM adalah untuk dapat membentuk *classifier* seperti persamaan 3 berikut ini.

$$f(x_i) = \begin{cases} \geq 0, & y_i = +1 \\ < 0, & y_i = -1 \end{cases} \quad (3)$$

Untuk membentuk *hyperplane* dijelaskan dalam persamaan 4 berikut.

$$W.X + b = 0 \quad (4)$$

Dimana :

- W : Bobot vektor $\{w_1, w_2, w_3, \dots, w_n\}$ n adalah banyaknya atribut
- b : Skalar, atau biasa disebut sebagai bias
- X : Potongan data latih atau *training tuples*

2.5 Dashboard

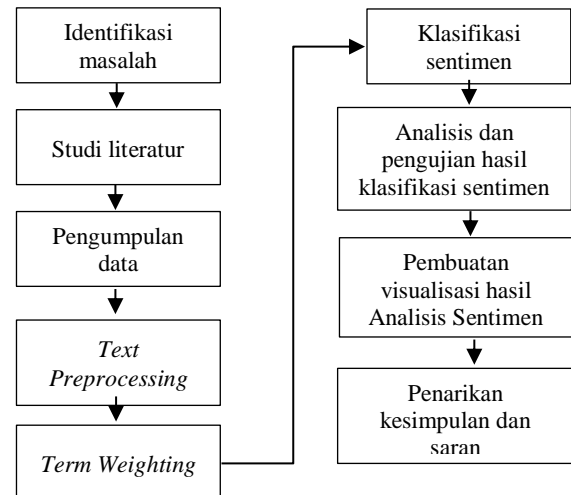
Eckerson (2005) menyatakan bahwa *Dashboard* didesain guna pencapaian tujuan strategis suatu organisasi atau perusahaan. Sehingga dengan penggunaan *Dashboard* diharapkan dapat mengukur, memonitor, dan memprediksi suatu aktivitas baik yang telah berjalan, sedang berjalan, maupun aktivitas selanjutnya. Oleh karena itu, organisasi diharapkan dapat mengukur dan memperbaiki strategi yang ada dalam rangka memaksimalkan kinerjanya.

2.6 Usability Testing

Bangor et al. (2008) mendefinisikan *Usability* sebagai sejauh mana kualitas yang dimiliki sistem dalam segi kemudahan segi kemudahan untuk dipelajari, kemudahan untuk digunakan, dan mendorong pengguna untuk menggunakan sistem sebagai alat positif dalam menyelesaikan tugas. *Usability Testing* dilakukan terhadap hasil visualisasi *Dashboard* dengan tujuan untuk menilai kualitas ketermudahan penggunaan *Dashboard* oleh pengguna.

3. METODE PENELITIAN

Dalam penelitian ini melalui beberapa tahapan yang dilakukan. Dimulai dengan identifikasi masalah, pencarian literatur yang sesuai, pengumpulan data, *Text Preprocessing*, *Term Weighting*, klasifikasi sentimen, analisis dan pengujian hasil klasifikasi sentimen, visualisasi hasil Analisis Sentimen, dan yang terakhir penarikan kesimpulan dan saran dari penelitian yang telah dilakukan. Gambar 2 menunjukkan Metodologi Penelitian untuk melakukan Analisis Sentimen opini pelanggan terhadap aspek Pariwisata Pantai Malang Selatan menggunakan TF-IDF dan *Support Vector Machine*.



Gambar 2 Metodologi Penelitian

3.1. Identifikasi Masalah

Mengidentifikasi permasalahan terkait objek penelitian yakni pada pariwisata pantai Malang selatan. Masalah tersebut terkait dengan perlunya melakukan Analisis Sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan untuk mendapatkan perspektif pelanggan.

3.2. Studi Literatur

Adanya studi literatur adalah untuk mencari teori – teori yang relevan terhadap penelitian sejenis yang pernah dilakukan sebelumnya yang berkaitan dengan penemuan sebelumnya, karena penelitian bertujuan untuk mendapatkan solusi dari permasalahan yang sedang dihadapi. Beberapa teori-teori yang digunakan seperti teori mengenai Opini, Analisis Sentimen, *Data Mining*, *Text Mining*, *Term Weighting*, *Machine Learning*, teori tentang pemilihan dan evaluasi model, serta visualisasi *Dashboard*.

3.3. Pengumpulan data

Tahap ini meliputi pengumpulan data, pemilihan aspek, dan perancangan dataset. Pengumpulan data dilakukan terhadap data opini dalam bentuk ulasan yang diberikan pelanggan terhadap objek wisata pantai Malang Selatan. Data bersumber dari situs TripAdvisor dengan menggunakan teknik *Web Scraping* untuk mendapatkan teks ulasan, pemberi ulasan serta waktu ulasan tersebut diberikan. *Tools* yang digunakan dalam proses ini adalah WebHarvy. Pemilihan aspek dilakukan dengan melakukan observasi terhadap data hasil proses *Web Scraping*, kemudian melakukan konfirmasi

dengan Disparbud Kabupaten Malang. Perancangan *dataset* meliputi tahap pelabelan manual terhadap data opini. *Dataset* dilabelkan ke dalam tiga kelas sentimen yakni positif, negatif, dan netral. Positif untuk ulasan dengan sentimen positif terhadap aspek, negatif untuk ulasan dengan sentimen negatif terhadap aspek, netral untuk ulasan yang tidak memiliki sentimen terhadap aspek.

3.4. Text Preprocessing

Sebelum dilakukan pengolahan, data terlebih dahulu melewati tahap *Text Preprocessing* untuk mengkondisikan data agar sesuai dengan kebutuhan untuk tahapan berikutnya yakni *Term Weighting* dan klasifikasi sentimen. Tahap *Text Preprocessing* meliputi pembuatan indeks kata (*Term indeks*), *Case Folding* untuk menyeragamkan huruf ke dalam bentuk *lowercase*, *Cleansing* untuk menghilangkan tanda baca yang tidak diperlukan, Tokenisasi untuk mengubah struktur teks ke dalam bentuk token. Formalisasi dan Translasi, untuk mengembalikan kata ke bentuk baku serta mengubah teks bahasa asing ke dalam bahasa Indonesia. *Stemming* untuk mengubah kata menjadi kata dasar, serta *Stop-word Removal* untuk menghilangkan kata yang masuk ke dalam kategori *stop-word*. Tahap *Text Preprocessing* menggunakan dalam Bahasa Pemrograman Python dengan memanfaatkan *module* yang sesuai seperti *NLTK*, *Sastrawi*, *Pandas*, dan lainnya.

3.5. Term Weighting

Term Weighting atau pembobotan dilakukan dengan menerapkan *Term Frequency-Inverse Document Frequency* (TF-IDF). Pembobotan dilakukan dengan memanfaatkan *module* dalam Bahasa Pemrograman Python dengan *module Scikit-learn* untuk melakukan pembobotan TF-IDF.

3.6. Klasifikasi Sentimen

Berdasarkan hasil observasi data dan wawancara dengan dinas terkait, aspek yang dimiliki oleh pariwisata pantai Malang Selatan ada 5 aspek. Kelima aspek tersebut yakni akses Umum yang meliputi kondisi objek pariwisata pantai secara keseluruhan, aspek Kebersihan yang merupakan kondisi kebersihan dari objek pariwisata pantai, aspek Keramaian yang menunjukkan tingkat keramaian dari objek pariwisata pantai, aspek Akses Jalan yang

menjelaskan kondisi akses jalan menuju objek pariwisata pantai, dan aspek kondisi Ombak yang menjelaskan kondisi ombak atau perairan yang dimiliki objek pariwisata pantai. Pengujian model dan pemilihan *split data training* dan *testing* dengan menggunakan *Stratified K-Fold Cross Validation* dengan 4 *Fold* dan menghasilkan 75% data latih dan 25% data uji. Klasifikasi sentimen dilakukan dengan menggunakan algoritma *Support Vector Machine* dengan menggunakan *Kernel Linear* dan strategi *Multiclass One-Vs-Rest* untuk menentukan sentimen dari opini terhadap tiap-tiap aspek yang telah ditentukan. Penerapan SVM dan *Stratified KFold Cross Validation* dilakukan dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python dengan menggunakan *LinearSVC* dan *StratifiedKFold*.

3.7. Analisis dan Pengujian Hasil Klasifikasi Sentimen

Selanjutnya dilakukan analisis hasil klasifikasi dan pengujian terhadap hasil dari klasifikasi sentimen pada tahap sebelumnya untuk mendapatkan nilai akurasi dari penggunaan algoritma *Support Vector Machine* yang dikombinasikan dengan *Term Frequency-Inverse Document Frequency*. Jenis pengujian yang dilakukan adalah pengujian *F1-Score*, *Precision*, dan *Recall* serta penghitungan *Accuracy*. Pengujian dilakukan dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python.

3.8. Pembuatan Visualisasi Hasil Analisis Sentimen

Selanjutnya dilakukan visualisasi terhadap hasil Analisis Sentimen. *Dashboard* Analisis Sentimen menampilkan hasil klasifikasi sentimen opini pelanggan terhadap pariwisata pantai Malang Selatan. *Dashboard* trend sentimen memvisualisasikan perubahan sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dari tahun 2013 hingga tahun 2018. *Dashboard* objek wisata adalah *Dashboard* yang menampilkan pemeringkatan objek pariwisata pantai Malang Selatan terhadap 5 aspek yang ada. *Dashboard* aspek memvisualisasikan data sentimen dari tahun 2013 hingga 2018 dari setiap objek wisata pantai Malang Selatan untuk setiap aspek.

3.9. Penarikan Kesimpulan dan Saran

Tahap terakhir yaitu adalah penarikan

kesimpulan berdasarkan rumusan masalah terkait Analisis Sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dengan menggunakan TF-IDF dan *Support Vector Machine*. Kemudian pemberian saran untuk penelitian di masa mendatang.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan data

Pengumpulan atau akuisisi data dilakukan dengan menggunakan metode *Scraping*. *Tool* yang digunakan dalam melakukan *Web Scraping* adalah WebHarvy. Data opini pelanggan pariwisata pantai Malang Selatan bersumber dari situs TripAdvisor. Proses *Scraping* menghasilkan 674 data ulasan berbahasa Indonesia yang diberikan oleh 423 pemberi ulasan terhadap 43 objek pariwisata pantai Malang Selatan. Data ulasan tersebut berjangka waktu dari tahun 2013 hingga tahun 2018. Tabel 1 menunjukkan contoh data ulasan yang didapatkan dari penerapan *Web Scraping* data opini terhadap pariwisata pantai Malang Selatan.

Tabel 1 Contoh data hasil *Scraping*

Judul, Tanggal, Nama, Pantai	Teks
Judul: ombak	Pantai ini terletak di pesisir besar
Tanggal: 19/06/2016	selatan malang. Pantainya bersih dengan air yang cukup bersih juga. Pantai ini berombak cukup
Nama: Tjutjukbs	besar dan tidak disarankan untuk
Pantai: Ngantep	mandi ataupun berenang di pantai ini.

4.2. Pemilihan Aspek dan Pelabelan *dataset*

Berdasarkan observasi terhadap data ulasan yang ada di situs TripAdvisor serta setelah dikonfirmasi dengan proses wawancara dengan pihak Dinas Pariwisata dan Kebudayaan Kabupaten Malang Selatan didapatkan 5 aspek yang dimiliki oleh entitas pariwisata Pantai Malang Selatan yakni aspek Kebersihan, Keramaian, Akses Jalan, Ombak, dan aspek Umum. Tahap perancangan *dataset* meliputi tahap manualisasi pelabelan data yang dilakukan terhadap data teks uraian opini pelanggan untuk setiap aspek yang dimiliki oleh pariwisata pantai Malang Selatan. Label data dikelompokkan menjadi 3 kelas yang mewakili 2 sentimen yakni positif dan negatif terhadap aspek serta kelas netral yang menyatakan bahwa data tersebut tidak memiliki sentimen terhadap aspek. Untuk

dapat menentukan label untuk setiap aspek disesuaikan dengan polaritas sentimen dari masing-masing aspek yang dijelaskan dalam Tabel 2 berikut ini.

Tabel 2 *Polaritas sentimen aspek pariwisata pantai Malang Selatan*

Aspek	Positif	Negatif
Umum	secara umum kualitas objek pariwisata sudah sesuai dengan ekspektasi atau keinginan pelanggan	secara umum kualitas objek pariwisata tidak dapat memenuhi ekspektasi atau keinginan pelanggan
Kebersihan	Kondisi kebersihan atau keasrian objek pariwisata pantai terjaga dengan baik	Kondisi kebersihan tidak lagi terjaga dikarenakan adanya faktor-faktor yang merusak kebersihan
Keramaian	tingkat keramaian pengunjung masih rendah.	tingkat keramaian pengunjung tinggi
Akses Jalan	kondisi sarana akses menuju lokasi pariwisata sudah baik dan memadai	kondisi sarana akses menuju lokasi pariwisata masih buruk dan menyulitkan
Kondisi Ombak	kondisi perairan dan ombak cukup tenang dan tidak berbahaya	kondisi perairan dan ombak membahayakan

4.3 Pemilihan Model Menggunakan *Kfold Cross Validation*

Pemilihan model dilakukan untuk mendapatkan model terbaik dengan akurasi tertinggi untuk setiap aspek. Pemilihan model dilakukan dengan menggunakan metode *K-fold Cross Validation*. *K-fold Cross Validation* adalah proses sistematis untuk melakukan prosedur penyusunan *dataset* menjadi data latih dan data uji secara berulang-ulang dengan tujuan untuk menghindari terjadinya variansi tunggal yang terbentuk saat melakukan hanya dengan satu set data latih dan data uji.

Kfold Cross Validation diterapkan dengan memanfaatkan *module Scikit-learn* pada Python dan memilih *Stratified K-fold* untuk memastikan rasio untuk semua *Class* seimbang untuk semua *fold*. *Dataset* dibagi menjadi 4 *Fold* untuk mendapatkan rasio data latih dan data uji sebesar 75:25. Parameter *shuffle* dalam *StratifiedKfold* bernilai *True* dengan tujuan untuk mendapatkan model dengan variasi lebih banyak. *Parameter tuning* juga dilakukan untuk mendapatkan kombinasi parameter yang menghasilkan rata-rata akurasi tertinggi, *parameter tuning*

dilakukan terhadap parameter dari *random_state* dalam proses *StratifiedKfold*. Setelah mendapatkan parameter terbaik maka dilakukan pemilihan model dengan menjalankan *method cross_val_score* untuk mendapatkan kombinasi data latih dan uji dengan akurasi tertinggi. Setelah mendapatkan kombinasi data latih dan uji yang memiliki akurasi tertinggi maka data tersebut disimpan ke dalam 4 variabel. 4 variabel tersebut adalah *X_train*, *y_train*, *X_test*, dan *y_test*. *X_train* untuk data bobot ulasan yang digunakan sebagai data latih. *y_train* untuk data label yang digunakan sebagai data latih. *X_test* untuk data ulasan yang digunakan sebagai data uji. *y_test* untuk data label yang digunakan sebagai data uji. Semua tahapan tersebut kemudian dilakukan terhadap data untuk masing-masing aspek untuk mendapatkan model terbaik untuk masing-masing aspek.

4.4. Hasil Klasifikasi

Berikut ini adalah hasil klasifikasi sentimen pada data uji untuk masing-masing aspek yang dikelompokkan menjadi 3 kelas sentimen yakni positif, negatif, dan netral.

1. Aspek Umum

Klasifikasi sentimen untuk aspek Umum menghasilkan 14 ulasan dengan sentimen negatif, 141 ulasan dengan sentimen positif, dan 12 ulasan dengan sentimen netral.

2. Aspek Kebersihan

Klasifikasi sentimen untuk aspek Kebersihan menghasilkan 17 ulasan dengan sentimen negatif, 31 ulasan dengan sentimen positif, dan 120 ulasan dengan sentimen netral.

3. Aspek Keramaian

Klasifikasi sentimen untuk aspek Keramaian menghasilkan 11 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 142 ulasan dengan sentimen netral.

4. Aspek Akses Jalan

Klasifikasi sentimen untuk aspek Akses Jalan menghasilkan 34 ulasan dengan sentimen negatif, 15 ulasan dengan sentimen positif, dan 121 ulasan dengan sentimen netral.

5. Aspek Ombak

Klasifikasi sentimen untuk aspek Ombak menghasilkan 29 ulasan dengan sentimen negatif, 14 ulasan dengan sentimen

positif, dan 126 ulasan dengan sentimen netral.

4.5. Hasil pengujian akurasi

Pengujian dilakukan dengan menerapkan persamaan yang dijelaskan pada Persamaan 5, 6, 7, dan 8. Proses pengujian dilakukan dengan memanfaatkan *module Scikit-learn* pada Bahasa Pemrograman Python. Tabel 4 menunjukkan hasil pengujian yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk masing-masing aspek.

Tabel 3 Hasil pengujian

Aspek	Accu.	Prec.	Rec.	F1-Score
Umum	0,85	0,75	0,85	0,79
Kebersihan	0,87	0,87	0,87	0,86
Keramaian	0,92	0,93	0,92	0,91
Akses Jalan	0,85	0,84	0,85	0,83
Ombak	0,87	0,87	0,88	0,86
Rata-rata	0,87	0,85	0,87	0,85

4.4. Hasil visualisasi Dashboard

Gambar 3 menunjukkan salah satu hasil visualisasi *Dashboard* yakni *Dashboard* objek wisata. *Dashboard* objek wisata merupakan *Dashboard* yang menampilkan pemeringkatan objek pariwisata pantai Malang Selatan terhadap 5 aspek yang ada. *Dashboard* menampilkan data dalam bentuk *radar chart* dengan 5 sumbu yang menunjukkan 5 aspek yang ada. Data yang dijadikan *input* adalah sentimen positif dari semua aspek yang dimiliki oleh suatu objek wisata pantai. Dalam visualisasi ini menampilkan 17 objek wisata pantai Malang Selatan. Ketujuh belas objek wisata tersebut dapat ditampilkan maupun dibandingkan dengan objek wisata lainnya dengan cara mengatur penyajian data melalui *button* untuk setiap objek wisata yang terletak di bagian kanan *radar chart*.

Gambar 4 menunjukkan hasil visualisasi *Dashboard trend* sentimen. *Dashboard trend* sentimen memvisualisasikan perubahan sentimen opini pelanggan terhadap aspek pariwisata pantai Malang Selatan dari tahun 2013 hingga tahun 2018. Dalam *Dashboard* tersebut menampilkan data persentase dari 3 kelas sentimen untuk setiap aspek dari tahun 2013 hingga tahun 2018. Data dikelompokkan berdasarkan aspek. Untuk mengatur tampilan data bisa menggunakan kotak abu-abu yang didalamnya terdapat *button* untuk menampilkan data berdasarkan aspek. Kemudian adanya legenda untuk menerangkan penggunaan warna

yang menjadi representasi masing-masing sentimen. Pada bagian atas terdapat 4 *button* untuk dapat berpindah ke *Dashboard* lain.

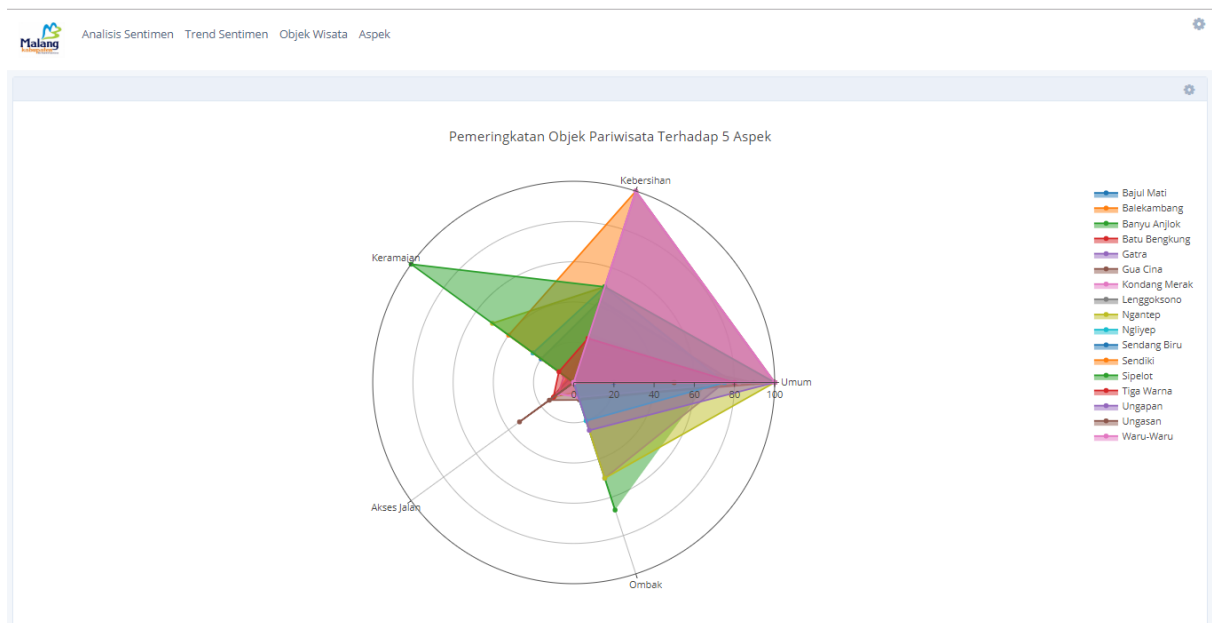
4.5. Pengujian Usability Dashboard

Usability testing dilakukan terhadap hasil visualisasi *Dashboard* dengan tujuan untuk menilai kualitas ketertudahan penggunaan *Dashboard* oleh pengguna. Berdasarkan *SUS score* untuk *Dashboard* mendapatkan skor sebesar 70 maka dapat disimpulkan bahwa *Dashboard* masuk ke dalam kategori *Acceptable* yang artinya *Dashboard* diterima baik oleh Dinas Pariwisata dan Kebudayaan Kabupaten Malang dalam proses pengambilan keputusan

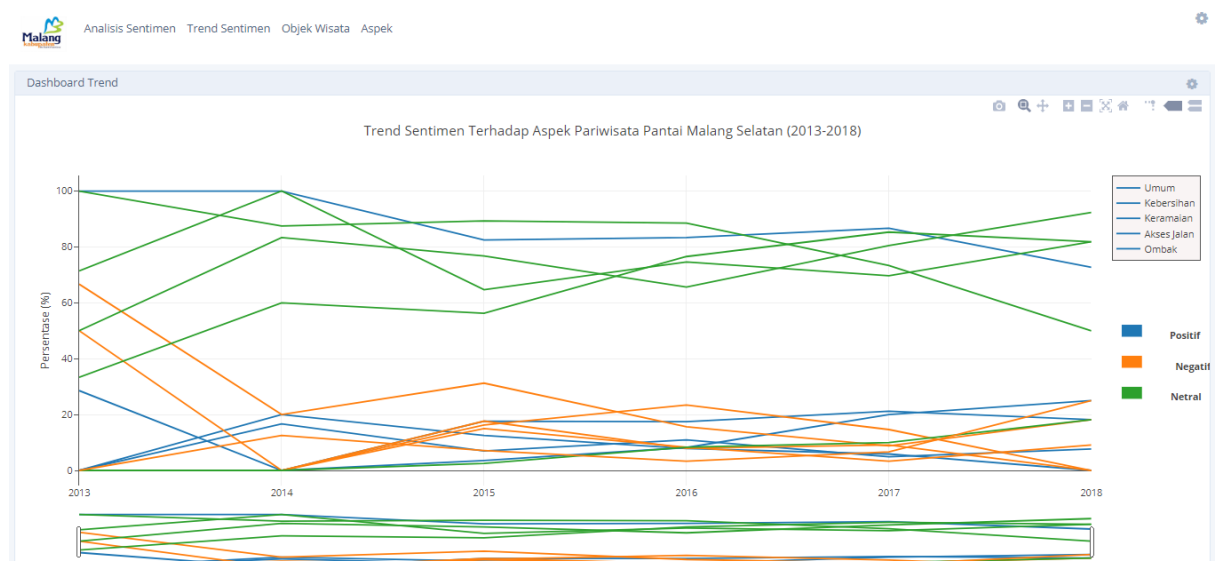
strategis terkait pariwisata pantai Malang Selatan.

5. KESIMPULAN

1. Data yang didapatkan dari penerapan *Web Scraping* berupa data opini berbahasa Indonesia dalam bentuk ulasan terhadap pariwisata pantai Malang Selatan dengan total 674 data ulasan berbahasa Indonesia untuk 43 objek wisata pantai Malang Selatan dari tahun 2013-2018.
2. Klasifikasi sentimen menggunakan *Support Vector Machine* mampu mengklasifikasikan sentimen untuk



Gambar 3 *Dashboard* Objek Wisata



Gambar 4 *Dashboard* Trend Sentimen

aspek Umum dengan tingkat akurasi sebesar 85%. Klasifikasi sentimen untuk aspek Kebersihan dengan tingkat akurasi 87%. Klasifikasi sentimen untuk aspek Keramaian dengan tingkat akurasi sebesar 92%. Klasifikasi sentimen untuk aspek Akses Jalan dengan tingkat akurasi sebesar 85%. Klasifikasi sentimen untuk aspek Ombak dengan tingkat akurasi sebesar 87%.

3. Pengujian terhadap hasil klasifikasi sentimen menghasilkan rata-rata yang baik pada *Accuracy* sebesar 87%, *Precision* sebesar 85%, *Recall* sebesar 87%, dan *F1-Score* sebesar 85%.
4. Proses *Usability testing* terhadap *Dashboard* dengan menggunakan metode *System Usability Testing* (SUS) menghasilkan nilai 70 yang memasukkan *Dashboard* sebagai sistem dengan kategori *Acceptable* dalam rating "Good".

6. SARAN

1. Menambahkan sumber data opini pelanggan dari sumber data lain seperti Media Sosial, Google Reviews, dan platform lain sejenisnya yang memuat data opini pelanggan pariwisata Pantai Malang Selatan.
2. Penggunaan metode untuk memperbaiki penggunaan kata yang tidak baku dan translasi frasa bahasa asing dalam tahap *Text Preprocessing*. Hal ini bertujuan untuk mengoptimalkan jumlah *feature* yang akan digunakan untuk tahap klasifikasi menggunakan algoritme klasifikasi.
3. Pemilihan aspek yang akan dianalisis dalam proses Analisis Sentimen dapat menggunakan pendekatan ekstraksi aspek dengan metode *Part-Of-Speech Tagger* dengan tujuan mendapatkan aspek yang memang benar-benar ada dalam *dataset*.

DAFTAR PUSTAKA

- Bangor, A., Kortum, P.T. & Miller, J.T., 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), hh.574–594.
- Boeing, G. & Waddell, P., 2016. New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings. *Journal of Planning Education and Research*, 37(4), hh.457–476
- Eckerson, W. 2005. *Deploying Dashboard and Scorecards*. TD WI Best Practices Report
- Fachrina, Z. & Widyantoro, D. H. Aspect-sentiment classification in opinion mining using the combination of rule-based and machine learning, 2017 *International Conference on Data and Software Engineering (ICoDSE)*, Palembang, 2017, hh. 1-6.
- Feldman, R. & Sanger, J., 2007. *The Text Mining Handbook: advanced approaches in analyzing unstructured data*, Cambridge: University Press.
- Han, J., Kamber, M. & Pei, J., 2012. *Data mining: concepts and techniques*, Amsterdam: Elsevier/Morgan Kaufmann.
- Liu, B., 2012. *Sentiment analysis and Opinion Mining*, S.I.: Morgan & Claypool.
- Manning, C.D., Raghavan, P. & Schütze Hinrich, 2009. *Introduction to information retrieval*, New York: Cambridge University Press.
- Marrese-Taylor, E. et al., 2013. Identifying Customer Preferences about Tourism Products Using an Aspect-based Opinion Mining Approach. *Procedia Computer Science*, 22, hh.182–191.
- Mitchell, R., 2018. *Web scraping with Python: collecting data from the modern web*, Beijing: O'Reilly.
- Wijayanti, R. & Arisal, A., 2017. Ensemble approach for sentiment polarity analysis in user-generated Indonesian text. 2017 *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*.