

Analisis Sentimen Berdasarkan Aspek pada Tempat Wisata di Daerah Istimewa Yogyakarta Menggunakan Metode Multinomial Naïve Bayes

Auliya Khanza Qorita
Program Studi Informatika – Program Sarjana
Universitas Islam Indonesia
Daerah Istimewa Yogyakarta
18523214@students.uii.ac.id

Fayruz Rahma
Program Studi Informatika – Program Sarjana
Universitas Islam Indonesia
Daerah Istimewa Yogyakarta
fayruz.rahma@uii.ac.id

Abstrak—Pengunjung tempat wisata umumnya akan melihat ulasan tempat wisata yang akan dikunjungi dari beberapa situs yang menyediakan ulasan tempat wisata. Hal tersebut dilakukan untuk memastikan situasi dan kondisi mengenai tempat wisata tersebut. Oleh karena itu, untuk memudahkan pengunjung wisata dalam melihat ulasan tempat wisata dibuat sebuah model yang dapat melakukan analisis sentimen ulasan tempat wisata berdasarkan aspek. Tujuan penelitian ini untuk mengetahui akurasi yang didapatkan dari penggunaan metode Multinomial Naïve Bayes dalam melakukan analisis sentimen berdasarkan aspek pada ulasan tempat wisata yang berada di Daerah Istimewa Yogyakarta. Berdasarkan hasil pengujian didapatkan bahwa data yang diklasifikasikan benar oleh model sebesar 70.4% pada aspek *attraction*, 89.1% pada aspek *accessible*, 72% pada aspek *amenities*, dan 93.2% pada aspek *ancillary services*.

Kata Kunci—Analisis Sentimen, Ulasan, Tempat Wisata, Multinomial Naïve Bayes

I. PENDAHULUAN

Indonesia memiliki banyak tempat wisata, baik itu wisata alam, budaya, religi, dan lainnya. Berdasarkan Badan Pusat Statistik (BPS), terdapat sebanyak 2958 perusahaan obyek wisata komersial di Indonesia pada tahun 2019. Kota Yogyakarta menjadi salah satu kota dengan destinasi pariwisata terbaik di Indonesia. Yang menjadi daya tarik wisatawan adalah kota ini masih menerapkan sistem kerajaan dan keindahan alam yang terdapat di Yogyakarta.

Dengan berkembangnya internet pada masa ini, berdampak juga dengan kemajuan yang sangat pesat pada sektor pariwisata. Karena internet, wisatawan dapat dengan mudah mencari informasi mengenai tempat wisata yang akan dikunjungi melalui situs yang menyediakan ulasan terkait tempat wisata tersebut. Situs tersebut umumnya menyediakan beragam informasi yang dibutuhkan oleh wisatawan seperti keadaan riil di lokasi, fasilitas yang dimiliki, dan beberapa rekomendasi penginapan terdekat.

Akan tetapi, ulasan tempat wisata yang ditulis oleh wisatawan yang telah mengunjungi tempat tersebut memiliki beberapa aspek dan polaritas yang berbeda. Hal tersebut mengakibatkan kesulitan dalam menentukan polaritas yang ada pada ulasan tempat wisata. Oleh karena itu, dibuat sistem yang dapat melakukan analisis sentimen ulasan tempat wisata berdasarkan aspeknya.

Tujuan penelitian ini dilakukan adalah untuk mengetahui akurasi dari penggunaan metode Multinomial Naïve Bayes dalam melakukan analisis sentimen berdasarkan aspek pada

ulasan tempat wisata yang berada di Yogyakarta. Teknik dan metode yang digunakan untuk mencapai tujuan tersebut adalah dengan menggunakan *Aspect-Based Sentiment Analysis* (ABSA) dan Multinomial Naïve Bayes. Teknik ABSA mengategorikan sentimen berdasarkan aspek menjadi 3 (tiga) kelas yaitu positif, netral, dan negatif. Digunakan 4 (empat) kategori aspek yaitu *attraction*, *accessible*, *amenities*, dan *ancillary services*.

II. STUDI PUSTAKA

Analisis sentimen atau disebut *opinion mining* merupakan ilmu untuk menganalisis pendapat seseorang, perilaku, penilaian, sentimen, dan emosi terhadap suatu entitas (Liesnawan, 2019) [1]. Analisis sentimen terbagi menjadi 3 (tiga) level bagian utama yaitu level dokumen, level kalimat, dan level aspek dan entitas [2].

Sudah banyak penelitian yang melakukan analisis sentimen tetapi masih pada level kalimat. Penelitian mengenai analisis sentimen pada ulasan tempat wisata yang sudah dilakukan diantaranya penelitian yang dilakukan oleh (Rahmi, 2020) [3], pada penelitian tersebut digunakan metode *Modified K-Nearest Neighbor* (MKNN) dan menghasilkan akurasi tertinggi pada nilai $k=1$ yaitu sebesar 43% untuk keseluruhan klasifikasi, 40% dengan $k=1$ untuk aspek harga, 79% dengan $k=5$ untuk aspek pelayanan, dan 46% dengan $k=1$ untuk aspek fasilitas. Metode MKNN dapat digunakan apabila menggunakan *dataset* yang baik dan melakukan *pre-processing* yang tepat.

Terdapat penelitian serupa yang dilakukan oleh (Imron, 2019) [4], peneliti melakukan analisis sentimen pada level kalimat menggunakan metode Naïve Bayes dan menghasilkan akurasi sebesar 82.8% dengan kekurangan data *training* yang digunakan tidak *balance* sehingga performa sistem dalam menemukan kembali suatu informasi pada kelas positif dan negatif hanya sebesar 0.07%.

Penelitian ini melakukan analisis sentimen level aspek karena menghasilkan performa yang lebih baik dibandingkan dengan level dokumen dan kalimat. Hal tersebut dikarenakan pada analisis sentimen level aspek difokuskan untuk menentukan sentimen positif, negatif, dan netral dengan memperhatikan aspek-aspek yang telah ditentukan. Metode yang digunakan adalah Multinomial Naïve Bayes dikarenakan berdasarkan penelitian (Randika SP, 2020) [5], mengenai analisis sentimen terhadap preferensi konsumen pada produk by.U dan MPWR menggunakan metode Multinomial Naïve Bayes menghasilkan nilai akurasi yang cukup bagus yaitu 87.5% untuk *dataset* MPWR. Yang membedakan antara

metode Naïve Bayes dan Multinomial Naïve Bayes adalah Naïve Bayes *classifier* lebih merujuk ke independensi bersyarat dari masing-masing fitur yang ada dalam model, sedangkan Multinomial Naïve Bayes *classifier* merupakan contoh spesifik dari Naïve Bayes *classifier* yang menggunakan *multinomial distribution* untuk masing-masing fitur.

III. LANDASAN TEORI

A. Data Labelling

Data Labelling merupakan proses identifikasi data *raw* lalu ditambahkan satu/lebih label yang memiliki makna dan informatif untuk memberikan konteks sehingga model *machine learning* dapat mempelajarinya. Pelabelan data ini diperlukan untuk beberapa kasus penggunaan contohnya seperti *computer vision*, *speech recognition*, dan pengolahan bahasa alami.

B. Pre-processing

Data-data yang diambil dari internet umumnya tidak lengkap, berantakan, dan tidak konsisten. *Pre-processing* data ini membantu dalam melakukan data *cleaning*, data *reduction*, dan data *discretization*. Tahap *pre-processing* data akan membuat *dataset* menjadi lebih rapi. Beberapa metode yang dilakukan saat *pre-processing* dapat meningkatkan akurasi dari model yang menggunakan Naïve Bayes dan mengurangi waktu pemrosesan [6].

C. Aspect-Based Sentiment Analysis

Analisis sentimen berdasarkan aspek merupakan proses untuk mendapatkan informasi suatu sentimen dari sudut pandang tertentu mengenai aspek tersebut (Liesnawan, 2019). Proses yang dilakukan pada tahap ini adalah menentukan sentimen (positif, negatif, atau netral) berdasarkan aspek dalam suatu kalimat [7]. Pada penelitian ini, sentimen akan dikategorikan menjadi 4 (empat aspek) yaitu, *attraction*, *attraction*, *accessible*, *amenities*, dan *ancillary services*. Aspek-aspek tersebut dipilih didasarkan menurut (Cooper et al., 1994), berpendapat bahwa terdapat 4 (empat) komponen yang harus dimiliki tempat wisata yaitu *attraction*, *attraction*, *accessible*, *amenities*, dan *ancillary services*. [8]

- *Attraction*: segala sesuatu yang dapat menarik minat wisatawan untuk mengunjungi tempat wisata tersebut.
- *Accessible*: menurut (Sugiyama, 2011), aksesibilitas ialah tingkat intensitas suatu daerah tempat wisata dapat dijangkau oleh wisatawan. Fasilitas dalam aksesibilitas contohnya seperti jalan raya, jalan tol, stasiun kereta api, dan lain-lain. Menurut (Stange and Brown, 2015), akses merupakan bagaimana cara seseorang untuk dapat mencapai tujuannya dari tempat asalnya, apa akses yang dilalui mudah atau sulit [9].
- *Amenities*: fasilitas-fasilitas pendukung yang diperlukan oleh wisatawan ketika berada di tempat wisata. Contohnya seperti fasilitas untuk memenuhi kebutuhan akomodasi, tempat hiburan, penyediaan makanan dan minuman, *retailing*, dan layanan lainnta seperti rumah sakit, bank, asuransi, dan keamanan (Cooper et al., 1994).

- *Ancillary services*: layanan pendukung yang disediakan oleh pemerintah daerah tempat wisata tersebut, organisasi, kelompok, atau pengelola tempat wisata untuk menyelenggarakan kegiatan wisata (Cooper et al., 1994).

D. Term Frequency – Inverse Document Frequency

TF-IDF merupakan metode yang paling umum digunakan ketika melakukan pencarian informasi. Metode ini juga sering digunakan karena mudah, efisien, dan mempunyai hasil yang akurat [10]. Metode ini menghitung nilai dari *Term-Frequency* (TF) dengan *Inverse Document Frequency* (IDF) pada setiap kata di setiap dokumen. Jadi, perhitungan bobot setiap token t di dokumen d dirumuskan sebagai berikut:

$$W_{dt} = tf_{dt} * IDF_t \quad (1)$$

Keterangan:

W_{dt} : bobot dari t (kata) dalam satu dokumen

tf_{dt} : frekuensi kemunculan t (kata) dalam dokumen d

IDF_t : *Inversed Document Frequency*

Nilai IDF_t dihitung dari

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

Keterangan:

N : jumlah semua dokumen

n_t : jumlah dokumen yang mengandung kata t

Setelah mengetahui bobot (W) pada masing-masing kata yang terdapat pada dokumen, selanjutnya dilakukan tahap *sorting* yaitu dimana semakin besar nilai bobot, maka akan semakin besar juga tingkat kesamaan dokumen pada kata kunci tersebut, ataupun sebaliknya.

E. Multinomial Naïve Bayes

Metode algoritma Naïve Bayes adalah salah satu metode yang efektif digunakan untuk bidang klasifikasi teks, tetapi akurasi model akan lebih akurat jika menggunakan *training sample set* yang banyak [11]. Metode Multinomial Naïve Bayes digunakan karena kecepatan dan kesederhanaan dalam melakukan klasifikasi teks [12]. Metode ini juga mengikuti prinsip dari distribusi multinomial dalam melakukan pemrosesan teks [13]. Metode ini tidak hanya menghitung kata yang muncul tetapi juga jumlah kemunculan setiap kata yang ada pada dokumen [14]. Menurut (Manning et al., 2008), perhitungan probabilitas sebuah dokumen d berada di kelas c menggunakan Multinomial Naïve Bayes dapat dirumuskan sebagai berikut: [15]

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (3)$$

Keterangan:

- $P(t_k|c)$ merupakan *conditional probability* dari kata t_k yang terdapat dalam sebuah dokumen kelas c .
- $P(c)$ merupakan *prior probability* dari sebuah dokumen yang terdapat pada kelas c . Apabila kata dari suatu dokumen memberikan petunjuk yang kurang jelas untuk satu kelas dibandingkan kelas lainnya, maka akan dipilih satu kelas yang memiliki *prior probability* tertinggi.

- $\langle t_1, t_2, \dots, t_{n_d} \rangle$ merupakan kumpulan *token* pada dokumen d yang merupakan bagian dari *vocabulary* yang digunakan saat melakukan klasifikasi dan n_d adalah jumlah *tokens* yang terdapat dalam d .

Nilai *prior probability* P_c dapat dihitung menggunakan persamaan berikut:

$$P(c) = \frac{N_c}{N} \quad (4)$$

Keterangan:

N_c : jumlah dari dokumen *training* dalam kelas c .

N : jumlah keseluruhan dokumen *training* dari seluruh kelas.

Perhitungan nilai dari *conditional probability* $P(t_k|c)$ dapat dilakukan menggunakan persamaan berikut:

$$P(t_k|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (5)$$

Keterangan:

T_{ct} : jumlah kemunculan kata t dalam sebuah dokumen *training* dalam kelas c .

$\sum_{t' \in V} T_{ct'}$: jumlah total dari keseluruhan kata yang terdapat pada dokumen *training* dalam kelas c .

Perhitungan nilai *prior probability* P_c dan *conditional probability* $P(t_k|c)$ menggunakan *maximum likelihood estimate* (MLE). Masalah yang terjadi pada penggunaan MLE adalah terdapat nilai nol dari kombinasi kelas yang tidak terdapat dalam dokumen *training*. Untuk menghilangkan nilai nol pada dokumen tersebut, dapat menggunakan *Laplace smoothing* sebagai proses menambahkan nilai 1 (satu) untuk setiap nilai T_{ct} dari perhitungan *conditional probability*. Sehingga, formula untuk *conditional probability* menjadi persamaan berikut:

$$P(t_k|c) = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'} \quad (6)$$

Keterangan:

B' : jumlah keseluruhan kata untuk dari seluruh kelas.

F. Accuracy Metric

Matriks akurasi merupakan salah satu matriks yang digunakan untuk mengevaluasi performa model klasifikasi. Akurasi dapat didefinisikan sebagai berikut:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (7)$$

Singkatnya, akurasi adalah rasio perbandingan informasi yang diprediksi benar oleh sistem dengan keseluruhan informasi. Umumnya, akurasi diformulakan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

Keterangan:

TP (*true positive*) : informasi positif yang benar diprediksi positif oleh sistem

TN (*true negative*) : informasi negatif yang benar diprediksi negatif oleh sistem

FP (*false positive*) : informasi negatif yang

diprediksi positif oleh sistem
FN (*false negative*) : informasi positif yang diprediksi negatif oleh sistem

IV. METODOLOGI PENELITIAN

A. Data Labelling

Tahap awal yang dilakukan pada penelitian ini adalah dengan memberi keterangan label kelas pada sentimen (positif/netral/negatif) dari data ulasan tempat wisata. Data tersebut didapatkan dari hasil *crawling* situs web Traveloka dan Tiket. Penulis mengategorikan sentimen menjadi beberapa aspek yaitu *attraction*, *attraction*, *accessible*, *amenities*, dan *ancillary services*. Kemudian, dilakukan *data preparation* yaitu proses menyiapkan data yang dibutuhkan sebelum melakukan *pre-processing*. Data yang disiapkan adalah data ulasan yang telah diberi label ditransformasikan menjadi *pandas (library Python) dataframe*.

B. Pre-processing

Tahap ini dilakukan untuk membuat data menjadi lebih mudah untuk diolah model nantinya. Metode-metode *pre-processing* yang dilakukan penulis antara lain adalah:

- Case Folding*: mengubah semua huruf yang terdapat dalam teks menjadi huruf kecil.

TABLE I. CASE FOLDING

Sebelum	Sesudah
Mengunjungi tebing ini pada sore hari akan disuguhi temaram senja yang teduh, hangat dan pemandangan	mengunjungi tebing ini pada sore hari akan disuguhi temaram senja yang teduh, hangat dan pemandangan
Pemandangan alamnya cukup bagus, bisa melihat sebagian Jogja dari tempat yang cukup tinggi.	pemandangan alamnya cukup bagus, bisa melihat sebagian jogja dari tempat yang cukup tinggi

- Filtering*: menghilangkan karakter selain huruf alfabet a-z seperti tanda baca, *hashtag*, dan lain-lain.

TABLE II. FILTERING

Sebelum	Sesudah
mengunjungi tebing ini pada sore hari akan disuguhi temaram senja yang teduh, hangat dan pemandangan	mengunjungi tebing ini pada sore hari akan disuguhi temaram senja yang teduh hangat dan pemandangan
pemandangan alamnya cukup bagus, bisa melihat sebagian jogja dari tempat yang cukup tinggi	pemandangan alamnya cukup bagus bisa melihat sebagian jogja dari tempat yang cukup tinggi

- Stop words removal*: menghilangkan kata umum yang biasanya muncul dalam jumlah yang banyak dan dianggap tidak memiliki makna. Daftar *stop words* yang akan digunakan terdiri dari 2 (dua) bahasa yaitu Bahasa Indonesia dan Inggris.

TABLE III. STOPWORDS REMOVAL

<i>Sebelum</i>	<i>Sesudah</i>
mengunjungi tebing ini pada sore hari akan disugahi temaram senja yang teduh hangat dan pemandangan	mengunjungi tebing sore hari disugahi temaram senja teduh hangat pemandangan
pemandangan alamnya cukup bagus bisa melihat sebagian jogja dari tempat yang cukup tinggi	pemandangan alamnya cukup bagus melihat sebagian jogja tempat cukup tinggi

C. Feature Extraction

Tahapan yang dilakukan selanjutnya adalah *feature extraction*. Tahap ini dilakukan untuk mentransformasi teks ke dalam angka. Terdapat berbagai teknik yang dapat digunakan untuk mengubah teks ke dalam *array*. Penulis menggunakan TF-IDF. Metode ini akan menghitung bobot untuk masing-masing kata pada dokumen.

D. Modelling

Kemudian, ke tahap *modelling*. Metode yang digunakan adalah algoritma *Multinomial Naïve Bayes*. Dalam proses *modelling*, dilakukan pembagian data *training* dan *testing* dengan rasio pembagian 80-20 (*training-testing*).

E. Evaluation

Berikutnya, tahap evaluasi model yaitu menentukan performa dari model yang telah dibuat sudah baik atau belum. Digunakan *metric accuracy* untuk melihat performa dari model yang sudah dibuat.

V. HASIL DAN PEMBAHASAN

A. Hasil Data Labelling

Data ulasan tempat wisata yang telah didapatkan dari hasil *crawling* akan diberi label kelas (positif/negatif/netral) secara manual. Proses ini dilakukan untuk mengategorikan polaritas yang terdapat pada ulasan tempat wisata berdasarkan aspek-aspek yang telah ditentukan. Hasil dari distribusi data setelah dilakukan *labelling* adalah sebagai berikut:

TABLE IV. HASIL DATA LABELLING

<i>aspek</i>	<i>positive</i>	<i>neutral</i>	<i>negative</i>
sentiment_attraction	1313	75	24
sentiment_accessible	161	21	45
sentiment_amenities	334	9	229
sentiment_ancillary services	161	8	30
Jumlah	1929	113	328

Berdasarkan Table IV, aspek yang memiliki sentimen positif paling banyak adalah aspek *attraction* yaitu sebanyak 1313 ulasan. Banyak ulasan yang tidak dapat dikategorikan dikarenakan isi dari ulasan tersebut tidak termasuk dalam aspek-aspek yang telah ditentukan.

B. Hasil Pre-processing

Sebelum dilakukan pemodelan, data ulasan tempat wisata yang telah dilabeli akan dibersihkan terlebih dahulu. Tahap

ini menggunakan bahasa pemrograman *Python*. Metode-metode yang dilakukan diantara lain adalah *case folding*, *filtering*, dan *stop words removal*. Berikut hasil dari tahap *pre-processing*:

TABLE V. HASIL PREPROCESSING

<i>review</i>	<i>clean review</i>
Masuknya murah, dapat sunset yang bagus pula. Cukup ramai dan asyiknya pantainya bersih banget dan airnya biru	masuknya murah sunset bagus cukup ramai asyiknya pantainya bersih banget airnya biru
Tugu jogja adalah icon dari kota jogja sendiri yang menyuguhkan pemandangan yang indah terlebih di malam hari	tugu jogja icon kota jogja sendiri menyuguhkan pemandangan indah terlebih malam hari
Maliboro kotor, Sangat kotor jalannya. Tidak sesuai foto yang ditampilkan. Sayang sekali .dimana mana sampah. Orangnya ramah dan sopan	maliboro kotor sangat kotor jalannya sesuai foto ditampilkan sayang sekali sampah orangnya ramah sopan

Tahap *pre-processing* bertujuan untuk mempermudah data ketika diolah menjadi model. Tahap selanjutnya adalah ekstraksi fitur menggunakan TF-IDF dan dilanjutkan dengan *modelling*.

C. Perhitungan Akurasi

Berikut merupakan hasil perhitungan akurasi yang ditunjukkan pada Tabel VI.

TABLE VI. HASIL AKURASI

<i>Aspek</i>	<i>Akurasi</i>
sentiment_attraction	0.7046632124352331
sentiment_accessible	0.8911917098445595
sentiment_amenities	0.7202072538860104
sentiment_ancillary services	0.9326424870466321

Berdasarkan hasil akurasi pada Tabel VI, menunjukkan bahwa data yang diklasifikasikan benar oleh model sebesar 70.4% pada aspek *attraction*, 89.1% pada aspek *accessible*, 72% pada aspek *amenities*, dan 93.2% pada aspek *ancillary services*. Pada aspek *attraction* dan *amenities* menghasilkan akurasi yang belum cukup bagus dikarenakan data *training* yang digunakan tidak *balance* dan *dataset* yang digunakan bersifat subyektif sehingga dapat terjadi *error* seperti salah klasifikasi ketika melakukan pengujian model.

VI. KESIMPULAN

Data yang digunakan untuk melakukan penelitian ini adalah 1922 ulasan pengguna yang didapatkan dari hasil *crawling website* Tiket dan Traveloka. Telah dilakukan pelabelan data ulasan tempat wisata yang menghasilkan 1929 ulasan dengan sentimen positif, 113 ulasan dengan sentimen netral, dan 328 ulasan dengan sentimen negatif.

Berdasarkan hasil dari penelitian ini, penulis berhasil melakukan analisis sentimen berdasarkan aspek pada tempat wisata di Yogyakarta menggunakan metode Multinomial Naïve Bayes. Akurasi yang didapatkan dengan menggunakan metode Multinomial Naïve Bayes cukup baik yaitu

menghasilkan nilai 70.4% pada aspek *attraction*, 89.1% pada aspek *accessible*, 72% pada aspek *amenities*, dan 93.2% pada aspek *ancillary services*.

REFERENCES

- [1] Liesnawan and K. Imam, "Analisis Sentimen Berdasarkan Aspek Menggunakan Bayesian Network", Diss. Universitas Komputer Indonesia, 2019.
- [2] Saputra. A, Adiwijaya. A, and Mubarok. M, "Klasifikasi Sentimen Pada Level Aspek Terhadap Ulasan Produk Berbahasa Inggris Menggunakan Bayesian Network (case Study: Data Ulasan Produk Amazon)." *eProceedings of Engineering* 4.3, 2017.
- [3] Rahmi. F, "Analisis Sentimen Berdasarkan Aspek Pada Ulasan Tempat Wisata Menggunakan Modified K-Nearest Neighbor," Diss, Universitas Komputer Indonesia, 2020.
- [4] Imron. A, "Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," Diss, Universitas Islam Indonesia, 2019.
- [5] Randika SP, A, "Preferensi Konsumen Terhadap Produk by.U dan MPWR dengan Analisis Sentimen Berbasis Multinomial Naive Bayes" Diss, Universitas Muhammadiyah Jember, 2020.
- [6] P. Chandrasekar and K. Qian, "The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier," 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), 2016, pp. 618-619, doi: 10.1109/COMPSAC.2016.205.
- [7] Siregar and O. Pardomuan, "Analisis Sentimen Berdasarkan Aspek Menggunakan Relevance Vector Machine," Diss, Universitas Komputer Indonesia, 2020.
- [8] Cooper, Chris. et al, *Tourism Principles & Practice*. England: Longman Group, 1993.
- [9] Stange. J and Brown, "Tourism destination management: Achieving sustainable and competitive results", The International Institute for Tourism Studies: Washington, DC: A publication of US Agency for International Development, 2015.
- [10] S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF," *Journal of documentation*, 2004
- [11] Y. Huang and L. Li, "Naive Bayes classification algorithm based on small sample set," 2011 IEEE International Conference on Cloud Computing and Intelligence Systems, 2011, pp. 34-39, doi: 10.1109/CCIS.2011.6045027.
- [12] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 257-261, doi: 10.1109/NGCT.2016.7877424.
- [13] Farisi, A. Abdurrahman, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naive Bayes classifier," *Journal of Physics: Conference Series*. Vol. 1192, No. 1. IOP Publishing, 2019.
- [14] I. H. Witten, F. Eibe and M. A. Hall, *Data Mining : Practical Machine Learning Tools and Techniques*, Third Edition USA: Elsevier, 2011.
- [15] Manning. C, Raghavan. P and Schütze. H, "Introduction to Information Retrieval", Cambridge University Press, 2008.