

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

*by Jane Doe*

---

**Submission date:** 27-Nov-2021 04:11PM (UTC+0700)

**Submission ID:** 1712999338

**File name:** Publikasi\_Ilmiyah.pdf (355.48K)

**Word count:** 1990

**Character count:** 12776

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

4 Author 1  
Program Studi Informatika – Program Sarjana  
Universitas Islam Indonesia  
Daerah Istimewa Yogyakarta  
email Author 1

4 Author 2  
Program Studi Informatika – Program Sarjana  
Universitas Islam Indonesia  
Daerah Istimewa Yogyakarta  
email Author 2

**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, dibuat sistem yang dapat melakukan analisis sentimen ulasan tempat wisata berdasarkan aspek. Tujuan penelitian ini adalah untuk mengetahui akurasi yang didapatkan dari penggunaan metode Multinomial Naïve Bayes dalam melakukan analisis sentimen berdasarkan aspek.

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

## 21 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. 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. Preprocessing

Data-data yang diambil dari internet umumnya tidak lengkap, berantakan, dan tidak konsisten. *Preprocessing* data ini membantu dalam melakukan data *cleaning*, data *reduction*, dan data *discretization*. Tahap *preprocessing* 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 [1].

### C. Aspect-Based Sentiment Analysis

Analisis sentimen berdasarkan aspek merupakan proses untuk mendapatkan informasi suatu sentimen dari sudut pandang tertentu mengenai aspek tersebut [2]. Proses yang dilakukan pada tahap ini adalah menentukan sentimen (positif, negatif, atau netral) berdasarkan aspek dalam suatu kalimat [3].

### 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 [4]. 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 bobot ( $W$ ) pada masing-masing kata di dokumen telah diketahui, selanjutnya dilakukan tahap pengurutan dimana semakin besar nilai bobot, maka akan semakin besar juga tingkat kesamaan dokumen terhadap kata kunci tersebut, demikian pula 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 [5]. Metode Multinomial Naïve Bayes digunakan karena kecepatan dan kesederhanaan dalam melakukan klasifikasi teks [6]. Metode ini juga mengikuti prinsip dari distribusi multinomial dalam melakukan pemrosesan teks [7]. Metode ini tidak hanya menghitung kata yang muncul tetapi juga jumlah kemunculan setiap kata yang ada pada dokumen [8].

#### F. Accuracy Metric

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

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (5)$$

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

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (6)$$

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

### III. 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*. Aspek-aspek ini dipilih didasarkan menurut Cooper dkk (1993: 84-86), berpendapat bahwa terdapat 4 (empat) komponen yang harus dimiliki tempat wisata yaitu *attraction*, *attraction*, *accessible*, *amenities*, dan *ancillary services* [9]. Kemudian, dilakukan *data preparation* yaitu proses menyiapkan data yang

dibutuhkan sebelum melakukan *preprocessing*. Data yang disiapkan adalah data ulasan yang telah diberi label ditransformasikan menjadi *pandas (library Python) dataframe*.

#### B. Preprocessing

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

- a) *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

- b) *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

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

TABLE III. STOPWORDS REMOVAL

Sebelum	Sesudah
mengunjungi tebing ini pada sore hari akan disuguhi temaram senja yang teduh hangat dan pemandangan	mengunjungi tebing sore hari disuguhi 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.

### IV. 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

aspek	positive	neutral	negative
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 Preprocessing

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 *stopwords removal*. Berikut hasil dari tahap *preprocessing*:

TABLE V. HASIL PREPROCESSING

review	clean review
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

review	clean review
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 *preprocessing* 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

Aspek	Akurasi
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*.

### V. 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] P. Chandrasekar and K. Qian, "The Impact of Data Preprocessing on the Performance of a Naïve Bayes Classifier," 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), 2016, pp. 618-619, doi: 10.1109/COMPSAC.2016.205.
- [2] Liesnawan and K. Imam, "Analisis Sentimen Berdasarkan Aspek Menggunakan Bayesian Network", Diss. Universitas Komputer Indonesia, 2019.
- [3] Siregar and O. Pardomuan, "Analisis Sentimen Berdasarkan Aspek Menggunakan Relevance Vector Machine," Diss, Universitas Komputer Indonesia, 2020.

- [4] S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF," *Journal of documentation*, 2004
- [5] 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.
- [6] 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.
- [7] 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.
- [8] I. H. Witten, F. Eibe and M. A. Hall, *Data Mining : Practical Machine Learning Tools and Techniques*, Third Edition USA: Elsevier, 2011.
- [9] Cooper, Chris. et al, *Tourism Principles & Practice*. England: Longman Group, 1993.

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

## ORIGINALITY REPORT

18%

SIMILARITY INDEX

17%

INTERNET SOURCES

7%

PUBLICATIONS

1%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="https://doku.pub">doku.pub</a> Internet Source	2%
2	Aa Zezen Zaenal Abidin, Enung Nurjanah. "SISTEM PERINGKAS TEKS OTOMATIS MULTI DOKUMEN KLIPING ARTIKEL BERITA GEMPA MENGGUNAKAN METODE TF-IDF", Jurnal Teknologi dan Komunikasi STMIK Subang, 2020 Publication	2%
3	<a href="http://www.repository.trisakti.ac.id">www.repository.trisakti.ac.id</a> Internet Source	1%
4	<a href="http://journal.uii.ac.id">journal.uii.ac.id</a> Internet Source	1%
5	<a href="http://elibrary.unikom.ac.id">elibrary.unikom.ac.id</a> Internet Source	1%
6	<a href="http://www.scribd.com">www.scribd.com</a> Internet Source	1%

[www.slideshare.net](http://www.slideshare.net)

7	Internet Source	1 %
8	<a href="http://ejournal.upnvj.ac.id">ejournal.upnvj.ac.id</a> Internet Source	1 %
9	<a href="http://repository.uin-suska.ac.id">repository.uin-suska.ac.id</a> Internet Source	1 %
10	<a href="http://adoc.pub">adoc.pub</a> Internet Source	1 %
11	<a href="http://jpk.ppj.unp.ac.id">jpk.ppj.unp.ac.id</a> Internet Source	1 %
12	<a href="http://repositori.usu.ac.id">repositori.usu.ac.id</a> Internet Source	1 %
13	<a href="http://text-id.123dok.com">text-id.123dok.com</a> Internet Source	1 %
14	<a href="http://inccst.muet.edu.pk">inccst.muet.edu.pk</a> Internet Source	1 %
15	<a href="http://repository.unair.ac.id">repository.unair.ac.id</a> Internet Source	1 %
16	<a href="http://syiarmajelis.blogspot.com">syiarmajelis.blogspot.com</a> Internet Source	1 %
17	Borislava Vrigazova. "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems", Business Systems Research Journal, 2021 Publication	< 1 %

---

18	<a href="http://bluefooz.wordpress.com">bluefooz.wordpress.com</a> Internet Source	<1 %
19	<a href="http://eprints.akakom.ac.id">eprints.akakom.ac.id</a> Internet Source	<1 %
20	<a href="http://eprints.ums.ac.id">eprints.ums.ac.id</a> Internet Source	<1 %
21	<a href="http://id.123dok.com">id.123dok.com</a> Internet Source	<1 %
22	<a href="http://jifosi.upnjatim.ac.id">jifosi.upnjatim.ac.id</a> Internet Source	<1 %

---

Exclude quotes  On

Exclude matches  Off

Exclude bibliography  On