

ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER STUDI KASUS JAWA BARAT

Lio Wilianto^{1*}, Tacbir Hendro Pudjiantoro¹, Fajri Rakhmat Umbara¹

¹ Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Jenderal Achmad Yani

Jl. Terusan Jenderal Sudirman, 148 Cimahi, Jawa Barat, Indonesia

*Email: liowilianto@icloud.com

Abstrak

Penelitian ini dilakukan untuk memberikan informasi tentang kualitas sebuah tempat wisata yang ada di Jawa Barat. Dengan menggunakan sentimen dari pengunjung, Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat dapat menentukan langkah apa yang harus diambil untuk memutuskan langkah apa yang harus diambil tanpa harus turun langsung dan melihat satu – persatu. Dengan menggunakan Google Maps sebagai sumber data dalam pembuatan aplikasi, data yang digunakan dalam pembuatan sistem ini yaitu data tempat wisata, sentimen pengunjung dan rating tempat.

Kata Kunci : Google Maps API, Tempat Wisata, Sentimen, Naïve Bayes Classifier.

1. PENDAHULUAN

Travelling atau berwisata merupakan kegiatan yang sangat digemari oleh seluruh masyarakat Indonesia. Jawa Barat merupakan salah satu Provinsi di Indonesia yang memiliki banyak tempat wisata, diantaranya Ekowisata, Wisata Budaya, dan Wisata Petualangan. dari sekian banyak jenis wisata yang ada di Jawa Barat, ekowisata menjadi salah satu wisata yang digemari masyarakat pada umumnya. Ekowisata adalah suatu bentuk perjalanan wisata ke area alami yang dilakukan dengan tujuan mengkonservasi lingkungan, melestarikan kehidupan dan kesejahteraan penduduk setempat.

Analisis sentimen adalah sebuah cara yang dilakukan untuk mendapatkan *Opinion Public*. Dengan Analisis Sentimen kita mendapatkan *feedback* yang cepat dari masyarakat dengan memanfaatkan komentar dari masyarakat yang berkunjung sehingga pemerintah setempat dapat melakukan tindakan cepat dalam laporan masyarakat.

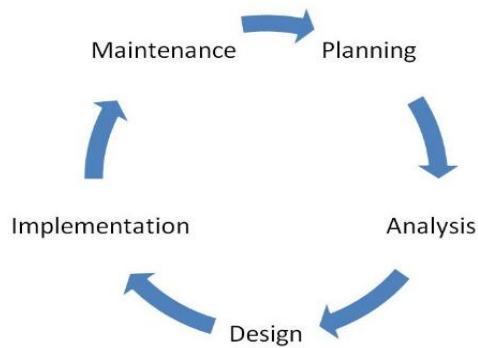
Dengan banyaknya obyek wisata di Jawa Barat cukup menyulitkan Dinas Pariwisata dan Kebudayaan Jawa Barat dalam memantau perkembangan di tiap tempat wisata yang ada, dengan adanya sistem ini diharapkan mempermudah Dinas Pariwisata dan Kebudayaan dalam mendapatkan informasi. Dan dapat cepat menentukan langkah apa yang harus diambil kedepannya.

Google Maps

Google maps API merupakan jasa peta atau *globe virtual* dan *Online* yang disediakan oleh perusahaan Google, Google Maps mempunyai fitur untuk merencanakan sebuah rute dan mencari sebuah alamat, baru – baru ini Google menambahkan fitur rating dalam Google Maps sehingga penggunaanya dapat menambahkan Komentar langsung pada tempat yang dikunjunginya.

2. METODOLOGI

Proses pembuatan prototipe menggunakan model pendekatan kepada pengembang software yang sistematis dan sekuensial yang meliputi aktivitas sehari – hari seperti analisis kebutuhan, desain, pengkodean, pengujian, dan evaluasi, serta pemeliharaan. Gambar 1. mengilustrasikan proses pembuatan prototipe :



Gambar 1. System Development Live Cycle

3. HASIL DAN PEMBAHASAN

Hasil dan Pembahasan meliputi cara dan langkah – langkah penulis dalam menyelesaikan penelitiannya, adapun langkah – langkahnya sebagai berikut :

3.1. TAHAP PREPROCESSING

Tahap *Preprocessing* diperlukan untuk membersihkan data dari hal yang tidak diperlukan, dengan tujuan pada tahap masuk ke dalam metode *naïve bayes classifier* lebih optimal dalam perhitungannya pada tahap ini melibatkan rekognisi dari isi dan stuktur teksnya. Adapun tahapan – tahapan dari *preprocessing*.

1. *Convert Emoticon*

Emoticon merupakan salah satu cara pengungkapan ekspresi perasaan secara tekstual. Hal tersebut akan membantu dalam menentukan sentimen dalam suatu kalimat atau komentar. Jenis *emoticon* yang akan diproses adalah *emoticon western style*, dikarenakan jenis tersebut banyak digunakan atau menjadi standar di semua platform seperti web ataupun mobile. Pada tabel 1 dijelaskan hasil pengklasifikasian arti dari *emoticon western style* secara umum.

Tabel 1 Emoticon yang akan Dikonversi

Emoticon	Deskripsi
:-) :) :o) :] :3 :o) :> =] 8)	Senang
=) :) ^) :>)	
:D :D 8-D 8D x-D xD X-D XD =- D =D =-3 =3 B^D	Ketawa
: :@ >:(Benci
\$. >[:-((:o :o :< :>C :< : [:[:{ ;(:-(:(D<D: D8 D; D= DX v.v D-'	Kecewa
<3 ;-) ;) *-) *) ;-] ;] ;D ;^' ;,	Suka

2. *Cleansing*

Cleansing merupakan proses membersihkan kata – kata yang tidak diperlukan untuk mengurangi noise. Kata yang dihilangkan adalah URL, *Hastag*(#), *Username* (@username), dan email. Selain itu juga tanda baca seperti titik(.), Koma(,) dan tanda baca yang lainnya akan dihilangkan Pada tabel 2.

Tabel 2 Penerapan Cleansing

Data Latih	
INPUT	OUTPUT
Ulasan Taman Hutan Raya IR. H. Djianda Https://goo.gl/maps/T4rA9NiiT5z Taman hutan raya yang masih terjaga, suasana sejuk dan bebas polusi, parkir sangat luas, full	Taman hutan raya yang masih terjaga, suasana sejuk dan bebas polusi, parkir sangat luas, full.

3. *Case Folding*

Case Folding adalah tahapan untuk merubah bentuk kata – kata menjadi sama bentuknya, Baik semuanya menjadi *lower case* atau *upper case*,

4. *Convert Negation*

Dalam bahasa terdapat kata yang dapat membalikan arti dari kata tersebut atau bersifat negasi. Kata – kata yang bersifat negasi adalah “Kurang”, “Tidak”, “enggak”, “ga”, “nggak”, “tak”, dan “gak”.

5. *Tokenizing*

Tokenizing bekerja untuk mengidentifikasi kata – kata dalam teks menjadi beberapa urutan yang terpotong oleh spasi atau karakter spesial. Berikut contoh penerapan dari *tokenizing*.

Tabel 3 Penerapan Tokenizing

Data Latih	
INPUT	OUTPUT
Taman hutan raya djuanda taman hutan raya djuanda masih terjaga bersih dan sejuk	[taman, hutan, raya, djuanda, taman, hutan, raya, djuanda, masih, terjaga, bersih, dan sejuk]

6. *Stopping*

Stopping berguna untuk membuang kata – kata yang sering muncul dan bersifat umum, Kurang menunjukan relevansinya dengan teks. Kata – kata yang akan dibuang tersebut Didefinisikan dalam *stopword list*. Contoh beberapa kata yang sering masuk ke dalam *Stopword list* adalah “sebuah”, “yang”, “itu”. Berikut adalah beberapa daftar *Stoplist* Yang disimpan dalam *database*.

Tabel 4 Stoplist yang digunakan

masih	Dong	Ke	Ada	Yoi
malam	Ya	Loe	Pada	Yang
Ini	Dan	Juga	Kita	Saya
untuk	Dari	Bagi	Iya	di mana
kapan	Bisa	Mana	Itu	Sih
sudah	Bikin	dengan	Anda	Begitu

7. *Stemming*

Stemming adalah tahapan untuk membuat kata yang berimbuhan kembali ke bentuk asalnya

Contoh kata “memberikan” setelah melewati tahap ini maka akan menjadi “beri” [6].

Tabel 5 Penerapan Stemming

Data Latih	
INPUT	OUTPUT
menginspirasi	inspirasi

3.2. Algoritma Naïve Bayes

Naïve Bayes Classifier adalah penggolong menggunakan statistik sederhana berdasarkan teorema bayes yang mengasumsikan bahwa keberadaan atau ketiadaan dari suatu fitur tertentu dari suatu kelas tidak berhubungan dengan keberadaan atau ketiadaan fitur lainnya. Sifatnya yang sebagai model probabilitas, *naïve bayes classifier* bisa dilatih dengan efisien sebagai *supervised learning*. Pada pengaplikasiannya, parameter estimasi untuk model naïve bayes menggunakan metode kemungkinan maksimum. Dengan kata lain masih bisa bekerja dengan model naïve bayes tanpa harus memperdulikan bayesian probabilitas atau metode bayesian lainnya. Berikut adalah model matematis untuk *naïve bayes classifier*

$$p(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}$$

3.2.1. Sumber Data

Data yang digunakan berbentuk komentar yang diambil langsung memanfaatkan API *Google Maps*. Berikut contoh data yang digunakan

Tabel 6 Contoh Data yang Digunakan

Komentar
@yemimairawanti murah,parkir luas, udara sejuk.
@keizaahmed dekat kota, parkir murah.
@dedisopandi parkirnya luas, mudah.
@vincentius parkirnya mahal, gk sesuai harga.
@akhmadhidayat buang sampah sembarangan, jalan rusak, dan berlubang.
@risstomo beberapa tempat kotor dan banyak coretan di gua.

3.2.2 Implementasi Preprocessing

Sebelum masuk proses utama yaitu pengklasifikasian komentar menggunakan *naïve bayes*, data akan diolah terlebih dahulu pada tahap *preprocessing* agar pada tahap klasifikasi hasilnya bisa jauh lebih optimal pada tabel 7 menunjukan hasil *preprocessing* dari data pada tabel 6

Tabel 7 Data Hasil Preprocessing

Komentar
[murah, parkir, luas, udara, sejuk]
[dekat, kota, jalan, murah]

[parkir_luas]
[parkir_mahal]
[buang_sampah, jalan_rusak, berlubang]
[kotor, banyak_coretan]

3.3. Implementasi Naïve Bayes Classifier

Tahap ini merupakan tahap yang paling penting, pada tahap ini proses pengklasifikasian berdasarkan sentimen yang ada di dalam dokumen dimulai. Tahap ini mempunyai dua proses, berikut prosesnya

1. Proses *learning naïve bayes classiifier*

Naïve bayes classifier sebagai algoritma supervised learning harus diberi pengetahuan awal terlebih dahulu sebagai acuan untuk dapat mengklasifikasikan suatu dokumen berdasarkan sentimennya.

Ada tiga langkah dalam, proses learning. Berikut ketiga tahapan pada proses *learning*.

a. Membentuk Fitur

Dalam penelitian ini yang dimaksud dengan fitur adalah kata kunci yang akan menjadi parameter satuan data latih, yaitu dokumen (komentar) untuk diklasifikasikan ke dalam kelas yang telah ditentukan (positif atau kelas negatif). Dalam kata lain fitur adalah kata yang memiliki nilai sentimen. Tabel 8 berisikan pembentukan fitur dari data latih yang sudah dilakukan proses *preprocessing* terlebih dahulu (Tabel 7).

Tabel 8 Pembentukan Fitur Data Latih

Data	Fitur (kemunculan)	Kelas sentimen
D1	murah(1), luas(1)	Positif
D2	murah(1)	positif
D3	luas(1)	Positif
D4	mahal(1)	negatif
D5	sampah(1), jalan_rusak(2)	negatif
D6	kotor(1)	negatif

b. Menghitung probabilitas $p(c_i)$

Setelah membentuk fitur dengan kemunculannya dari data latih. Selanjutnya menghitung probabilitas dari setiap kelas dengan cara sebagai berikut :

$$p(c_i) = \frac{fd(c_i)}{|D|}$$

Keterangan :

$fd(C_i)$ = Jumlah dokumen yang termasuk c_i

$|D|$ = Jumlah data latih / jumlah komentar

Tabel 9 Probabilitas Kelas *Training Naïve Bayes*

Kelas Sentimen (c)	Data (j)						fd(cj)	p(cj)
	D1	D2	D3	D4	D5	D6		
Positif	1	1	1	0	0	0	3	$\frac{3}{6}$
Negatif	0	0	0	1	1	1	3	$\frac{3}{6}$

- c. Menentukan Probabilitas $p(w_k|c_i)$
 Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas dengan cara sebagai berikut :

$$p(w_k|c_i) = \frac{f(w_{ki}, c_i) + 1}{f(c_i) + |W|}$$

Keterangan :

$f(w_{ki}, c_i)$ = Nilai kemunculan kata w_{ki} pada kelas c_i

$f(c_i)$ = Jumlah keseluruhan kemunculan kata pada kelas c_i

$|W|$ = Jumlah keseluruhan n dari W_k dari rumus 3, maka diperoleh model Probabilistik untuk setiap fitur pada 10.

Tabel 10 Model Probabilitas dari Data Latih

Data $f(W_{ki}, C_i)$	Kelas Sentimen	
	Positif	Negatif
Murah	$\frac{0+1}{5+7} = \frac{1}{12}$	$\frac{2+1}{4+7} = \frac{3}{11}$
Luas	$\frac{0+1}{5+7} = \frac{1}{12}$	$\frac{2+1}{4+7} = \frac{3}{11}$
mahal	$\frac{2+1}{5+7} = \frac{3}{12}$	$\frac{0+1}{4+7} = \frac{1}{11}$
sampah	$\frac{1+1}{5+7} = \frac{2}{12}$	$\frac{0+1}{4+7} = \frac{1}{11}$
rusak	$\frac{1+1}{5+7} = \frac{2}{12}$	$\frac{0+1}{4+7} = \frac{1}{11}$
kotor	$\frac{1+1}{5+7} = \frac{2}{12}$	$\frac{0+1}{4+7} = \frac{1}{11}$

2. Proses klasifikasi *naïve bayes classifier*
 Berikut alur proses dari klasifikasi menggunakan *naïve bayes classifier*.



Gambar 2 Flowchart Klasifikasi Naïve Bayes

Berikut adalah contoh satu komentar yang akan dijadikan data uji menggunakan model probabilitas.

Tabel 11 Data Uji Klasifikasi

Komentar	
Sebelum <i>preprocessing</i>	Sesudah Preprocessing
tahura tempat yg asik buat bersantai udaranya sejuk parkirnya luas syg tiket parkirnya mahal dan bnyk coretan di gua.	tahura udara_sejuk parkir_mahal coretan gua

Dari alur proses klasifikasi ada dua tahapan penting. Berikut tahapannya :

- a. Menghitung Vmap
 Vmap adalah perhitungan yang digunakan *naïve bayes classifier* untuk menentukan probabilitas data uji dari masing – masing kelas berdasarkan dari proses *learning*. Nilai probabilitas yang terbesar akan dipilih. Berikut perhitungannya :

$$V_{map} = \operatorname{argmax}_c p(c) \prod p(w_k | c) \times p(c)$$

Berdasarkan dari hasil *training*, berikut hasil perhitungannya :

$$V_{map} = \operatorname{argmax}_{\{positif, negatif\}} p(w_k | c) \times P(c)$$

$$V_{map} = \operatorname{argmax}_{\{positif, negatif\}} p(w_k | c) \times P(c)$$

$$P(\text{"tahura"} | c) p(\text{"udara"} | c) p(\text{"sejuk"} | c) p(\text{"parkir"} | c) p(\text{"mahal"} | c) p(\text{"coretan"} | c) \\ P(\text{"gua"} | c).$$

1. V_{map} untuk sentimen positif

$$V_{map}(\text{"positif"}) =$$

$$P(\text{"tahura"} | \text{positif}) p(\text{"udara"} | \text{positif}) p(\text{"sejuk"} | \text{positif}) p(\text{"parkir"} | \text{positif}) \\ p(\text{"mahal"} | \text{positif}) p(\text{"coretan"} | \text{positif}) P(\text{"gua"} | \text{positif}).$$

$$= \frac{1}{2} \times \frac{1}{12} \times \frac{1}{12} \times \frac{3}{12} \times \frac{1}{12} \times \frac{2}{12} \times \frac{1}{12} \times \frac{1}{12} = 0,0084$$

2. V_{map} untuk Sentimen negatif

$$V_{map}(\text{"negatif"}) =$$

$$P(\text{"tahura"} | \text{negatif}) p(\text{"udara"} | \text{negatif}) p(\text{"sejuk"} | \text{negatif}) p(\text{"parkir"} | \text{negatif}) \\ p(\text{"mahal"} | \text{negatif}) p(\text{"coretan"} | \text{negatif}) P(\text{"gua"} | \text{negatif}).$$

$$= \frac{1}{2} \times \frac{1}{11} \times \frac{1}{11} \times \frac{1}{11} \times \frac{3}{11} \times \frac{1}{11} \times \frac{1}{11} \times \frac{1}{11} = 0,0064$$

b. Menentukan V_{map} maksimum

Dari hasil perhitungan V_{map} diatas didapatkan bahwa nilai V_{map} Positif lebih besar dibanding dengan V_{map} negatif. Bisa disimpulkan bahwa komentar tersebut di klasifikasikan ke dalam sentimen positif.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan terlihat bahwa algoritma *naïve bayes classifier* dapat mengklasifikasikan suatu opini berupa komentar ke dalam dua kelas yaitu positif dan negatif dengan akurat. Tingkat keakurasian dari pengklasifikasian tersebut sangat dipengaruhi oleh proses *training*. Sehingga dapat disimpulkan dari pengklasifikasian yang dihasilkan dapat terlihat dengan jelas informasi sentimen public terhadap tempat wisata

UCAPAN TERIMA KASIH

Selama proses penyusunan penelitian ini, penulis terkadang menghadapi berbagai permasalahan dan hambatan. Selain kekuatan dan kemudahan yang diberikan oleh Allah Subhanahuwataa'ala, terdapat beberapa dorongan, masukan dan bantuan baik secara moril maupun materil yang diberikan oleh dari berbagai pihak sebagai faktor penting demi terselesaikannya penelitian ini. Untuk itu, dengan segala ketulusan dan kerendahan hati, penulis menghaturkan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. Orang Tua saya Bapak yang sangat saya sayangi Kusnandar dan Ibu Saya Nanik Pujiastuti, dan Adik beserta keluarga besar yang telah memberikan dukungan motivasi serta do'a yang tulus sehingga penulis dapat menyelesaikan penelitian ini.
2. Yth. Bapak Tacbir Hendro Pudjiantoro, S.Si., M.T., selaku Dosen Pembimbing 1 yang telah memberikan ide solusi, waktu, dan sumbangan pikiran yang sangat berharga dalam membimbing dan mengarahkan penulis selama penyusunan penelitian ini.
3. Yth Bapak Fajri Rakhmat Umbara S.T., M.T., selaku Dosen Pembimbing 2 yang telah memberikan bimbingan, motivasi dan pentunjuk yang sangat berguna kepada penulis selama penyusunan penelitian ini.

4. Yth. Bapak Hernandi Sujono, S.Si., M.Si., selaku Dekan Fakultas MIPA Universitas Jenderal Achmad Yani.
5. Yth. Bapak Gunawan Abdillah, S.Si., M.Cs., selaku Ketua Jurusan Informatika Universitas Jenderal Achmad Yani.
6. Semua dosen beserta staf karyawan di Jurusan Informatika Universitas Jenderal Achmad Yani yang telah mendidik dan membekali penulis dengan ilmu pengetahuan yang sangat berharga selama masa perkuliahan.
Di lain pihak penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada:
 1. Sahabat-sahabat seperjuangan Informatika 2011 Universitas Jenderal Achmad Yani khususnya saudara Pujo Sulardi, Madzid Hidayat, Ira Anggraeni, Desta Gumilar, masih banyak lagi yang lainnya yang telah berbagi suka dukanya selama menempuh perkuliahan.

Akhir kata semoga amal baik Bapak/Ibu/Kerabat/Sahabat sekalian mendapat balasan dan karunia dari Allah Subhanahuwataa'la.

DAFTAR PUSTAKA

- D. I Gede Surya Rahayuda, "Identifikasi Jenis Obat Berdasarkan Gambar Logo Pada Kemasan Menggunakan Metode Naive Bayes," *Jurnal Sisfo*, vol. 06 no.01, pp. 17-32, Agustus 2016.
- F. Alfa Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, vol. 2 no.3, pp. 2354-5771, Mei-Juli 2015.
- B Denny Nathaniel Chandra, Gede Indrawan, and I Nyoman Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes Dengan Fitur N-Gram," *Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA)*, vol. 10 no.1, pp. 0852-730X, Februari 2016.
- N. Amir Hamzah, "Klasifikasi Teks Dengan Naive Bayes Classifier(NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, pp. 1979-911X, Nopember 2012.
- I. B. M. Yogie Adnyana dan R. Efendi, "Rancang Bangun Sistem Informasi Geografis Persebaran Lokasi Objek Parawisata Berbasis Web dan Mobile Android (Studi Kasus di Dinas Parawisata Kabupaten Gianyar)," 2012.