ANALISIS SENTIMEN REVIEW HOTEL MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*

Suryadi[1], Abdurrahman Ridho[2], Murhaban[3]
Program Studi Teknologi Informasi Fakultas Teknik
Universitas Teuku Umar
suryadi@utu.ac.id[1]
abdurrahman.ridho@utu.ac.id[2]
murhaban@utu.ac.id[3]

Abstrak

Abstrak- Wisatawan saat melakukan liburan ke suatu kota pastinya memesan sebuah hotel. Wisatawan dalam memesan hotel seringkali mengalami kesulitan dalam menentukan hotel mana yang akan dipilih. Traveloka merupakan salah satu jenis situs pemesanan hotel yang menawarkan berbagai fitur untuk memudahkan pengunjung dalam menentukan hotel yang akan dipilih. Salah satu fitur yang ditawarkan adalah adanya ulasan yang menampilkan berbagai komentar pengunjung tentang suatu hotel. Namun, semakin banyak komentar atau ulasan tentang suatu hotel maka pengunjung memerlukan waktu yang lama untuk membaca dan memilih hotel yang diinginkan. Berdasarkan permasalahan tersebut, maka dibutuhkan analisis sentimen yang dapat mengekstraksi sejumlah komentar untuk memperoleh informasi yang bermanfaat bagi pengunjung. Sistem analisis sentimen yang dibangun memiliki tujuan membuat sebuah model sentimen untuk menentukan ulasan komentar suatu hotel. Proses analisis sentimen dilakukan dengan menggunakan algoritma *Naive Bayes Classifier*. Hasil pengujian memperlihatkan bahwa klasifikasi sentimen menggunakan *Naïve Bayes Classifier* memperoleh hasil akurasi sebesar 90,61%, presisi sebesar 93,03%, *recall* sebesar 89,52% dan *f-measure* sebesar 90,99%.

Key Words : *Text Mining*, Analisis Sentimen, *Naïve Bayes Classifier*

1. Pendahuluan

Saat ini banyak perusahaan yang memberikan layanan untuk memudahkan siapa saja yang hendak mencari dan menentukan

akomodasi di berbagai tempat di seluruh dunia dengan mempergunakan layanan tersebut. Layanan yang dimaksud adalah berupa situs atau website yang menyediakan informasi mengenai akomodasi dan pemesanan hotel yang terdapat di berbagai kota di seluruh dunia, sebagai contoh adalah http://www.traveloka.com. Melalui situs http://www.traveloka.com, pengunjung dapat memperoleh berbagai informasi terkait akomodasi hotel untuk beberapa kota yang ada di Indonesia maupun luar negeri dengan sangat mudah dan cepat. Situs http://www.traveloka.com memiliki beberapa kelebihan dan dilengkapi beberapa fitur yang dapat membantu dan memudahkan pengunjung dalam menentukan hotel mana yang akan dipilih.

Fitur yang dimaksud adalah terdapat ulasan yang memuat berbagai ulasan-ulasan dan komentar dari setiap pengunjung yang menginap di hotel-hotel tertentu yang terdapat dalam situs http://www.traveloka.com. Dengan adanya ulasan-ulasan yang ada di situs http://www.traveloka.com, pengunjung dapat memperoleh gambaran secara lebih objektif sehingga dapat memudahkan pengunjung mengambil keputusan dalam memilih hotel yang akan dijadikan tujuan menginap. Keberadaan ulasan-ulasan yang dituliskan oleh para pengunjung di situs http://www.traveloka.com kemudian akan dianalisa dan diolah sehingga bisa menghasilkan sebuah keluaran yang bisa bermanfaat untuk pengunjung baru. Salah satu model analisa yang dilakukan untuk mengolah ulasan-ulasan komentar dari pengunjung adalah analisis sentimen.

Analisis sentimen merupakan bagian dari salah satu *opinion mining* [1], adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi[2]. Dilakukan untuk melihat pendapat terhadap sebuah masalah, atau dapat juga digunakan untuk identifikasi kecenderungan hal di pasar[3]. Analisis sentimen dalam penelitian ini adalah proses klasifikasi komentar ke dalam dua kelas, yaitu kelas sentimen positif dan kelas negatif.

Beberapa penelitian sebelumnya yang terkait dengan klasifikasi sentimen review. Indrayuni dan Wahyudi [4], pada penelitian ini menganalisa *review* hotel yang terdapat pada situs www.tripadvisor.com dan www.virtualtourist.com dengan menggunakan algoritma Naive Bayes. Penelitian tentang klasifikasi sentimen terhadap *review* film juga telah dilakukan oleh Dhande dan

Patnaik [5] dengan menggunakan algoritma *Naive Bayes, Neural Network*, dan *Naïve Bayes Neural Classifier*. Analisis sentimen tentang review pelanggan pada situs penjualan online menggunakan algoritma naïve bayes classifier, Purwanto dan Santoso[6].

Pengklasifikasian menggunakan algoritma *Naïve Bayes* sangat sederhana dan efisien [7]. Selain itu, pengklasifikasian menggunakan Naïve Bayes adalah teknik *machine learning* yang sangat populer yang digunakan untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain [8]. Namun, Algoritma *Naïve Bayes* juga terdapat beberapa kekurangan yaitu sangat sensitive dalam pemilihan fitur [7]. Terlalu banyak jumlah fitur yang diproses, tidak hanya meningkatkan waktu penghitungan tapi juga dapat menurunkan akurasi klasifikasi [9].

Berdasarkan berbagai hal yang sudah dikemukakan dalam penjelasan di atas, maka dalam penelitian ini akan menerapkan algoritma *Naïve Bayes* untuk sentimen analisis ulasan komentar hotel dengan menggunakan review pada situs pemesanan hotel online http://www.traveloka.com.

2. Metode Penelitian

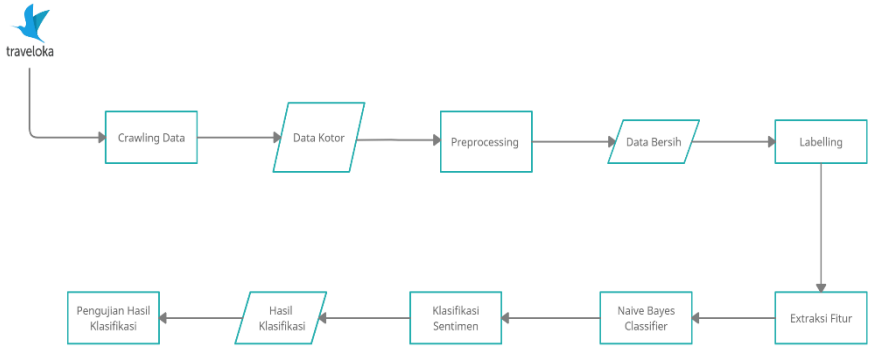
Secara umum metode penelitian pada penelitian ini terdiri dari beberapa tahap diantaranya adalah crawling data, *preprocessing*, *labelling*, ekstraksi fitur, pembuatan model algoritma naïve bayes classifier, klasifikasi sentimen, dan pengujian hasil klasifikasi.

Tahap pertama dari alur metode penelitian yaitu pengumpulan data dengan melakukan pendekatan *crawling* data dari server Traveloka memanfaatkan API *Traveloka*, Selanjutnya data hasil *crawling* data tersebut akan dilakukan proses *preprocessing* agar data siap digunakan untuk proses *labelling*. Tahap selanjutnya sebelum proses pelabelan *(labelling)* yaitu *preprocessing,* tahap *proprecessing* ini dilakukan pembersihan data ulasan komentar yang terdiri dari *case folding,* penghapusan simbol-simbol, tokenisasi, konversi *slangword,* penghapusan *stopword,* konversi kata negasi dan pengecekan *multiword*. Kemudian data komentar hasil *preprocessing* disimpan dalam database yang sudah menjadi komentar data bersih.

Pada proses *labelling* ini dilakukan sebagai tahapan sebelum proses training data yaitu melabelkan ulasan komentar secara manual untuk memilih komentar positif dan komentar negative. Kemudian data

ulasan komentar bersih yang sudah diberi label secara manual tersebut akan dijadikan sebagai data *training* untuk dilakukan perhitungan dengan metode *naïve bayes classifier* sehingga menghasilkan sebuah model klasifikasi sentimen. Langkah selanjutnya proses ekstraksi fitur dengan menggunakan *term frequency*, selanjutnya tahap yang dikerjakan pembuatan model klasifikasi sentimen dengan menggunakan algoritma naïve bayes classifier disertai dengan proses klasifikasi sentimen data testing menggunakan naïve bayes classifier.

Tahapan terakhir dari alur metode penilitian adalah pengujian hasil klasifikasi yang meliputi pengujian akurasi, presisi, *recall*, dan *f-measure*. Langkah-langkah proses alur metode penelitian seperti terlihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

3. Hasil dan Pembahasan

A. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data-data yang dibutuhkan dalam penelitian. Data-data yang dimaksud adalah kumpulan dari ulasan -ulasan atau komentar terkait tentang *review* hotel di www.traveloka.com. Pengumpulan data pada penelitian ini dilakukan dengan cara *crawling* pada situs www.traveloka.com. Proses *crawling* dilakukan dengan memanfaatkan *library* Python BeautifulSoup (https://pypi.python.org/pypi/beautifulsoup4) dan Traveloka API.

B. Preprocessing

Preprocessing sangat menentukan dalam proses penentuan klasifikasi sentimen dengan algoritma *naïve bayes classifier*. *Preprocessing*

juga digunakan dalam kedua proses utama, baik proses pelatihan (training) maupun proses penentuan klasifikasi sentimen. Tahap *preprocessing* terdiri dari beberapa proses yang akan dibahas satu per satu secara detail, antara lain:

1. *Case Folding*
   Case folding berfungsi untuk tahapan merubah semua karakter huruf yang ada di dalam komentar menjadi karakter huruf kecil semua.
2. *Penghapusan Simbol-Simbol*
   Penghapusan simbol-simbol berfungsi untuk menghapus karakter khusus dalam yang ada dalam komentar seperti tanda baca (seperti: koma (,), titik(.), tanda tanya (?), tanda seru (!) dan sebagainya), angka numerik (0 - 9), dan karakter lainnya (seperti: $, %, *, dan sebagainya).
3. *Tokenisasi*
   Tokenisasi berfungsi sebagai tahapan memecah komentar dari sekumpulan kalimat menjadi satuan kata. Proses tokenisasi dilakukan dengan memperhatikan setiap spasi yang ada pada setiap komentar maka berdasarkan spasi tersebut kata-kata dapat dipecah.
4. *Konversi Slangword*
   Konversi *slangword* merupakan tahapan mengubah terhadap kata tidak baku ke kata baku. Tahap ini dilakukan dengan menggunakan bantuan kamus *slangword* dan padanannya dalam kata-kata baku. Tahapan ini akan memeriksa kata yang terdapat dalam kamus *slangword* atau tidak. Jika kata tidak baku terdapat dalam kamus *slangword* maka kata tidak baku akan dirubah ke kata baku yang terdapat didalam kamus *slangword.*
5. Penghapusan *Stopword*
   Tahap ini berfungsi untuk menghapus dan menghilangkan kata-kata yang dianggap tidak penting dalam proses klasifikasi sentimen ulasan atau komentar, seperti kata: yang, maka, tetapi, atau, karena, ke, di, namun, dengan, dan sebagainya.
6. Konversi Kata Negasi
   Tahapan ini bertujuan untuk menghilangkan kata negasi dan mengkonversi atau merubah kata setelah kata negasi menjadi antonim yang terdapat dalam komentar. Proses konversi negasi dilakukan dengan mendeteksi kemunculan kata negasi "tidak" dan

"kurang" yang tedapat dalam sebuah komentar, selanjutnya kata setelah negasi dicek keberadaan kedalam kamus lawan kata atau antonim, jika kata tersebut terdapat dalam kamus antonim maka akan dirubah berdasarkan pasangan kata yang ada didalam daftar kata lawan kata.

7. Konversi *Multiword*
   Konversi m*ultiword* berfungsi menggabungkan dua kata yang mempunyai makna tunggal.Proses konversi m*ultiword*dilakukan dengan memeriksa apakah dua kata yang mempunyai makna tunggal terdapat dalam kamus data m*ultiword*atau tidak. Jika kata-kata yang bermakna tunggal ada dalam daftar m*ultiword*maka akan dirubah berdasarkan daftar m*ultiword*yang ada.

C. Labelling

Proses labelisasi data dilakukan sebelum ekstarksi fitur, proses labelisasi data yaitu seluruh data ulasan atau komentar dilabeli secara manual sesuai berdasarkan jenis klasifikasinya (sentimennya). Sebagai contoh ulasan "*hotel bersih*" maka akan dilabeli dengan "+" karena memiliki nilai sentimen positif yang berisi tentang kondisi hotel yang bersih, dan ulasan "*wastafel macet"* maka akan dilabeli dengan "-" karena memiliki nilai sentimen negatif yang menggambarkan bahwa wastafel sedang rusak atau macet.

D. Ekstraksi Fitur

Proses training data diawali dengan proses ekstraksi fitur. Ekstraksi kata bertujuan untuk mengambil kata-kata penting atau kata kunci untuk menentukan sentimen. Kata-kata kunci yang terdapat dalam komentar hotel akan digunakan untuk membuat model klasifikasi dari data *training*. Masukan dari proses ini adalah berupa data *training* yang sudah dilakukan preprocessing dan diberi label secara manual. Pelatihan dilakukan dengan metode *Naïve Bayes Classifier*. Model klasifikasi ini selanjutnya digunakan untuk mengidentifikasi klasifikasi sentimen ulasan komentar pada data komentar yang baru (data *testing*).

E. Perancangan Training

Perancangan proses training dilakukan dengan menggunakan metode *naïve bayes*.

a. *Perancangan Naïve Bayes Classifier*

Perancangan *naïve bayes classifier* dibagi menjadi 2 tahapan yaitu tahap pembelajaran dan tahap klasifikasi. Proses tahapan perhitungan *naïve bayes classifier* pada tahap pembelajaran adalah sebagai berikut :

a. Menghitung jumlah seluruh komentar (data) dalam data *training.*
b. Menghitung jumlah komentar pada masing-masing kelas.
c. Menghitung nilai *prior probability* kemunculan komentar pada masing-masing kelas.

$$P(C_j) = \frac{Nc_j}{M} \qquad (1)$$

d. Melakukan operasi logaritma pada nilai *prior probability* pada masing-masing kelas.
e. Menghitung jumlah kemunculan setiap kata (fitur)dalam masing-masing kelas tertentu dan jumlah total kemunculan kata untuk tiap kelasnya.
f. Mencari jumlah fitur unik *B* yang terdapat pada *vocabulary.*
g. Menentukan nilai probabilitas kondisional tiap fitur menggunakan Persamaan (2) untuk tiap kelas *P*(*fitur*|*kelas*).

$$P(x_i|c_j) = \frac{Nx_i.c_j+1}{\sum_{X\in V} Nx_i.c_j+1} = \frac{Nx_i.c_j+1}{(\sum_{X\in V} Nx_i.c_j)+B} \qquad (2)$$

h. menghitung probabilitas kondisional dengan menggunakan operasi logaritma terhadap setiap kata (fitur) untuk masing-masing kelas.

Proses tahapan perhitungan *naïve bayes classifier* pada tahap klasifikasi adalah sebagai berikut :

a. Menentukan nilai log probabilitas kondisional setiap kata yang terdapat pada komentar (data *testing*) dengan mengambil nilai log probabilitas kondisional setiap kata untuk masing-masing kelas yang terdapat pada model klasifikasi.
b. Mengambil nilai *prior probability* masing-masing kelas pada model klasifikasi.
c. Penentuan kelas sentimen berdasarkan persamaan (3) dengan mencari nilai maksimum dari nilai probabilitas akhir tiap kelas *P*(*kelas*|*komentar*).

$$C_{map} = argmax_{c_j \in C} \log\left(P(c_j)\Pi_{i=1}^{m} P(x_i|c_j)\right)$$

$$= argmax_{c_j \in C} P(c_j) + \sum_{i=1}^{m} P(x_i|c_j) \quad (3)$$

F. Klasifikasi Sentimen

Tahapan keenam dari proses sentimen analisis komentar *review* hotel adalah klasifikasi sentimen. Pada tahapan ini proses klasifikasi terhadap ulasan komentar dilakukan menggunakan model klasifikasi *naïve bayes classifier.* Proses klasifikasi ulasan komentar diklasifikasikan kedalam dua kelas yaitu ulasan komentar positif dan ulasan komentar negatif.

Penentuan klasifikasi atau sentimen ulasan komentar untuk data (komentar) baru menggunakan *naïve bayes classifier* dapat ditentukan sebagai berikut:

1. Melakukan *preprocessing* terhadap komentar kemudian mengekstrak kata-kata unik (kunci) yang ada dalam komentar (data *testing*).
2. Menentukan nilai log probabilitas kondisional setiap kata yang terdapat pada komentar (data *testing*) dengan mengambil nilai log probabilitas kondisional setiap kata untuk masing-masing kelas yang terdapat pada model klasifikasi.
3. Mengambil nilai *prior probability* masing-masing kelas c pada model klasifikasi.
4. Penentuan kelas sentimen dengan mencari nilai maksimum dari nilai probabilitas akhir tiap kelas *P*(*kelas*|*komentar*).

G. Pengujian Hasil Klasifikasi

Bagian ini membahas tentang pengujian dari hasil klasifikasi ulasan komentar yang telah dibangun. Pengujian klasifikasi ulasan komentar dilakukan dengan mengukur akurasi, presisi, *recall*, dan *f-measure* dari hasil perhitungan *naïve bayes classifier.*

Adapun total data komentar yang digunakan untuk pengujian klasifikasi (*testing)* adalah 225 komentar ulasan. Total data komentar untuk proses pengujian tersebut dibagi ke dalam empat hotel yang telah diinputkan sebelumnya oleh *user*. Rincian data komentar untuk proses pengujian klasifikasi ulasan komentar dan pengujian ulasan komentar diperlihatkan pada Tabel 1.

Tabel 1. Rincian Data Testing Komentar

Nama Hotel	Jumlah Komentar	Komentar	
		Positif	Negatif
Hotel Bugis Asri	84	56	27
Blue Safir	63	29	34
Grand Surya	37	13	24
Violet Malioboro	41	24	17
Total	225	122	102

Sebelum dilakukan pengujian terhadap klasifikasi ulasan komentar, seluruh data komentar ulasan dilabeli secara manual sesuai berdasarkan jenis klasifikasinya (sentimennya). Sebagai contoh komentar "*hotel bersih*" maka akan dilabeli dengan "+" karena memiliki nilai sentimen positif yang berisi tentang kondisi hotel yang bersih.

Pengujian klasifikasi ulasan komentar untuk mengukur akurasi, presisi, *recall*, dan *f-measure* diperoleh dengan membandingkan tiap komentar yang telah dilabeli secara manual dengan hasil perhitungan *naïve bayes classifier* yang dilakukan oleh sistem. Jumlah komentar yang sesuai antara hasil perhitungan *naïve bayes classifier* oleh sistem dengan pelabelan secara manual, akan mempengaruhi nilai akurasi, presisi, *recall* dan *f-measure* yang diperoleh. Semakin besar jumlah komentar yang sesuai, maka semakin tinggi pula nilai akurasi, presisi, *recall dan f-measure* yang didapatkan. Rekapitulasi hasil pengujian klasifikasi perhitungan klasifikasi ulasan komentar dengan menggunakan *naïve bayes classifier* untuk tiap hotel diperlihatkan pada Tabel 2.

Tabel 2. Hasil Pengujian Klasifikasi Komentar Menggunakan NBC

Nama Hotel	Akurasi	Presisi	Recall	F-Measure
Hotel Bugis Asri	89,46%	97,53%	89,10%	93,12%
Blue Safir	93,72%	99,43%	93,51%	96,38%
Grand Surya	91,80%	93,85%	91,04%	92,42%
Violet Malioboro	94,90%	98,51%	94,29%	96,35%
Total	92,47%	97,33%	91,98%	94,56%

Berdasarkan Tabel 3 di atas dapat disimpulkan bahwa hasil akurasi perhitungan klasifikasi ulasan komentar secara keseluruhan dengan menggunakan *naïve bayes classifier* yaitu sebesar 92,47 %, presisi 97,33 %,

recall 91,98 %, sedangkan hasil perhitungan untuk *f-measure* yaitu 94,56 %.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut:

1. Hasil pengujian pada system analisis sentimen yang telah dibangun memperoleh bahwa hasil pengujian menggunakan algoritma *naïve bayes classifier* memberikan hasil pengujian klasifikasi dengan akurasi yaitu sebesar 92,47%, presisi 97,33%, *recall*91,98%, dan *f-measure* yaitu 94,56%.
2. Algoritma *naïve bayes classifier* menghasilkan akurasi performansi yang lebih baik dalam penerapan proses klasifikasi.

Daftar Pustaka

Liu. B, "*Sentiment Analysis and Subjectivity, in Handbook of Natural Language Processing*," 2010.

Pang. B, dan Lee. L, "*Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval*," vol. Volume 2, no. Issue 1-2, pp. 1-135, 2008.

Pang. B, Lee. L, dan Vaithyanathan. S, "*Thumbs up? Sentiment Classification using Machine Learning, in Proceedings of the ACL-02 conference on Empirical methods in natural language processing*," vol.Volume 10, pp. 79–86,Morristown, NJ, USA, 2002.

Elly. I, dan Mochamad. W, "Penerapan Character N-Gram untuk Sentiment Analysis Review Hotel Menggunakan Algoritma Naive Bayes," Konferensi Nasional Ilmu dan Teknologi (KNIT), 8 Agustus 2015, Bekasi, 2015.

Lina. L. D, dan Girish. K. P, "Analyzing Sentiment of Movie Review Data using Naive *Bayes Neural Classifier*," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol (3) Issue 4*. ISSN 2278-6856, 2014.

Devi. D. P, dan Joan. S, "*Multinomial Naïve Bayes Classifier untuk Menentukan Review Positif atau Negatif pelanggan Website Penjualan*," Seminar Nasional "Inovasi dalam Desain dan Teknologi" 2015.

Jingnian. C, Houkuan. H, Shengfeng. T, dan Youli. Q, "*Feature selection for text classification with Naïve Bayes. Expert Systems with Applications*," *36*(3), 5432–5435, 2009.

Qaing. Y, Ziqiong. Z, dan Rob. Law, "*Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches*," *Expert Systems With Applications*, *36*(3), 6527–6535, 2009.

Alper. K. U, dan Serkan. G, "A novel probabilistic feature selection method for text classification." *Knowledge-Based Systems36*, 226–235, 2012.

Edmond. K, dan Edi. W, "Penambangan Opini Pada Situs Review Film Berbahasa Indonesia," *Tesis*, Program Magister Ilmu Komputer Universitas Gadjah Mada, Yogyakarta, 2012.