

ANALISIS KONTEN PELANGGAN AIRBNB PADA NETWORK SOSIAL MEDIA TWITTER

CONTENT ANALYSIS OF AIRBNB CUSTOMER BASED ON TWITTER SOCIAL MEDIA

Ditya Dwi Adhi Nugroho, Andry Alamsyah²

1 Prodi S1 Manajemen Bisnis Telekomunikasi Informatika, Fakultas Ekonomi dan Bisnis, Universitas Telkom

2 Prodi S1 Manajemen Bisnis Telekomunikasi Informatika, Fakultas Ekonomi dan Bisnis, Universitas Telkom

Email: ¹dityaadhi@student.telkomuniversity.ac.id, ²andrya@telkomuniversity.ac.id

Abstrak

Arus digital sekarang telah memberikan dampak yang cukup besar. Saat ini Internet telah menjadi salah satu hal yang penting, sosial media sudah menjadi salah satu alat komunikasi yang umum digunakan untuk orang-orang saling berkomunikasi, bertukar ide, bahkan membagikan informasi. Hal tersebut didukung oleh data yang mengatakan bahwa pertumbuhan pengguna sosial media sudah mencapai 2.7 miliar pengguna aktif sosial media. Banyaknya pengguna sosial media artinya arus informasi sangat banyak yang merupakan sebuah tambang informasi bagi perusahaan. Banyaknya informasi ini harus dibarengi dengan pengolahan data yang cepat agar dapat di manfaatkan oleh perusahaan. Tujuan penelitian ini adalah untuk mengali informasi yang ada pada media sosial. Untuk mendukung pengolahan informasi dengan cepat penelitian ini menggunakan *naïve bayes* sebagai *classifier* sentiment dan menggunakan *Latent Dirichlet Allocation (LDA)* untuk memodelkan topik. Penelitian ini diharapkan dapat membantu perusahaan dalam mengolah informasi secara *near real-time*. Peneliti menggunakan *Airbnb* sebagai studi kasus, sebab dipandang sebagai salah satu perusahaan penyewaan tempat yang banyak digunakan oleh pelancong. Periode pengambilan data dimulai pada 7 Januari 2018 hingga 28 Maret 2018 menggunakan kata kunci "*Airbnb*"

Kata kunci : *bigdata, naïve bayes, latent dirichlet allocation, social media*

Abstract

Digital currents have now made a considerable impact. Currently the Internet has become one of the important things, social media has become one of the common communication tools used for people communicate with each other, exchange ideas, and even share information. This is supported by data, saying that the growth of social media users has reached 2.7 billion active users of social media. The abundance of social media users means an enormous flow of information that is an information mine for the company. The amount of this information should be coupled with fast data processing in order to be utilized by the company. The purpose of this study is to multiply existing information on social media. To support rapid information processing this research uses *naïve bayes* as a sentiment classifier and uses *Latent Dirichlet Allocation (LDA)* to model topics. This research is expected to help the company in processing information in near real-time. Researchers use *Airbnb* as a case study, because it is seen as one of the rental companies that are widely used by travelers. The data retrieval period starts on January 7, 2018 to March 28, 2018 using the keyword "*Airbnb*".

Keywords: *bigdata, naïve bayes, latent dirichlet allocation, social media*

1. Pendahuluan

Industri pariwisata merupakan salah satu sektor industri yang terus menunjukkan trend positif setiap tahunnya. Hal tersebut didukung oleh laporan tahun dari *United Nation World Tourism Organization (UNWTO)* pada akhir tahun 2016, bahwa pertumbuhan industri pariwisata terus meningkat 3.9% setiap tahunnya serta memberikan proporsi sebesar 10% pada GDP dunia. Pendapatan dari industri ini pada tahun 2016 mencapai 1220 miliar USD [1]. Salah satu pendapatan terbesar dari industri ini adalah sektor perhotelan atau penyewaan tempat.

Salah satu *startup* penyewaan tempat yang mulai banyak digunakan oleh kalangan pelancong adalah *Airbnb*. Aplikasi berbasis *online* *Airbnb* sudah digunakan di lebih dari 191 negara dan 65.000 kota di seluruh dunia [2]. *Airbnb* adalah sebuah aplikasi *online* yang dapat memesan sebuah apartemen, rumah, kamar, atau villa. Aplikasi ini telah di unduh 10 juta kali di *platform* android dengan 180.000 pengguna aktif aplikasi *Airbnb*. Menurut penulis fenomena ini cukup menarik karena dari tahun berdirinya *Airbnb* pada tahun 2008 hingga saat ini perusahaan tersebut sudah dapat memasuki pasar global di berbagai negara bahkan terdapat 3.000.000 properti diseluruh dunia, dan telah digunakan oleh 200 juta orang dalam kurun waktu 9 tahun [2]. Serta memiliki

valuasi nilai perusahaan sebesar 31 miliar USD [3] dan menjadikan *Airbnb* perusahaan *startup* kedua terbesar setelah *uber* yang paling pesat pertumbuhannya [4].

Di lain sisi *Airbnb* memiliki pesaing seperti *homeaway*, *hoteltonight*, *holidayinn*, dan lain-lain. Pangsa pasar perusahaan-perusahaan tersebut sama, yaitu para pelancong atau *traveler* baik dalam penyewaan kamar dalam jangka waktu pendek maupun panjang. Semakin meningkatnya jumlah pelancong yang berpergian maka kebutuhan kamar akan meningkat. *Tophotelprojects* yang merupakan sebuah perusahaan yang menyediakan data perkembangan hotel memperkirakan akan adanya ledakan pembangunan hotel yang akan terus terjadi di hampir setiap tahunnya di seluruh dunia [5]. Menurut data dari *Tophotelprojects* di Amerika tengah dan selatan, telah dibangun sejumlah 1695 hotel dengan kamar sebanyak 352190 unit, di Asia Pasifik telah dibangun 1536 hotel yang mencakup 409120 kamar, di Eropa dibangun 1520 hotel dengan 253453 kamar, Timur Tengah dengan 525 proyek hotel yang meliputi 149915 kamar, serta Afrika dengan 323 hotel dan 77784 kamar [5]. Persaingan perhotelan semakin ketat dengan semakin banyaknya hotel-hotel baru yang bermunculan. Salah satu kebutuhan perusahaan untuk dapat menyaingi pesaing adalah dengan mengetahui apa yang menjadi kebutuhan konsumen.

Saat ini para pelancong atau *traveller* dan lainnya dapat memperoleh informasi tambahan dari mana saja seperti brosur, *word of mouth*, dan internet termasuk dari sosial media. Menurut data dari *wearesocial.com* pada bulan April 2017 jumlah pengguna internet sudah masuk diangka 3.8 miliar orang, sudah melwati setengah populasi dunia. Sehingga tidak heran jika sekarang banyak orang mencari informasi melalui internet atau sosial media. Hal ini penting karena *potential buyer* atau orang-orang memiliki sejumlah informasi penting di sosial media atau internet yang berguna untuk perusahaan agar dapat mengetahui apa yang menjadi keinginan konsumen atau mengetahui apa yang menjadi topik pembicaraan terhadap perusahaan di sosial media. Dianalogikan, jika di sosial media banyak orang memperbincangkan sebuah pengalaman mereka saat menginap di sebuah tempat, hal tersebut adalah tambang informasi bagi perusahaan, sebuah data yang dapat dimanfaatkan oleh perusahaan.

Di era yang serba cepat dan dimana pergeseran gaya hidup orang-orang pada zaman sekarang yang mulai segalanya berbentuk digital atau menggunakan internet. Serta cara berkomunikasi orang-orang sudah menjadi digital menggunakan beberapa platform sosial media. Maka dibutuhkan metode yang tepat untuk dapat mengekstraksi data secara cepat dan mendekati tepat. Hal tersebut berguna bagi perusahaan dalam mengetahui kebutuhan maupun keluhan yang pengguna gunakan agar dapat dimanfaatkan sebagai keuntungan perusahaan.

2. Tinjauan Literatur

2.1 Text Classification

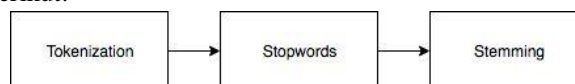
Text classification yang juga merupakan metode untuk mengelaskan suatu kalimat atau kumpulan teks [6]. Pengklasifikasian teks adalah untuk menentukan satu atau lebih label-label kelas yang ditentukan sebelumnya pada sebuah dokumen tekstual [7].

2.1.1 Training Data

Mempersiapkan data latih dan data uji berasal dari corpus yang berisi kumpulan opini publik *twitter* dengan konten sentimen adalah suatu keharusan dalam penelitian mengenai *sentiment analysis* [8]. Data latih berfungsi untuk membuat *machine learning* dapat bekerja dengan baik, karena komputer akan belajar jenis kata tersebut masuk kedalam sentimen pada data uji yang sudah ada di data latih. Pembagian data latih dan data uji yang optimal adalah 30% data latih dan 70% data uji [9].

2.1.2 Data Preprocessing

Preprocessing merupakan salah satu tahapan penting dalam penelitian ini karena pada tahap ini, data akan dipersiapkan sebelum di analisis. Tugas *preprocessing* adalah untuk membersihkan, dan menstrukturisasi data agar dapat dianalisis dalam proses selanjutnya dalam *text mining* [10]. Tahapan *preprocessing* sebagai berikut:



Gambar 1 Tahapan *Preprocessing*

Tokenization adalah membagi suatu teks menjadi elemen-elemen kecil dan bermakna seperti kalimat dan kata-kata [10]. *Stopwords* adalah kata-kata yang sering muncul dalam dokumen teks dan memiliki nilai yang rendah-tidak berguna [10]. *Stemming* adalah proses normalisasi teks menjadi bahasa baku, kata-kata didalam dokumen teks akan dimodifikasi sesuai dengan konteks tata bahasa [10].

2.1.3 Pembobotan TF-IDF

Pada *Naive bayes Classifier* diberikan pembobotan dengan teknik *Term frequency-inverse document frequency* atau disingkat TF-IDF, yang merupakan pengukuran gabungan: (TF); memberikan bobot nilai pada kata yang sering muncul dalam sebuah dokumen. dan (IDF); memberikan bobot nilai pada kata yang jarang muncul pada dalam dokumen tetapi dianggap penting [11]. Menurut Hofmann [10], rumus pembobotan *tf-idf* untuk bobot kata *t* pada dokumen *d*, yaitu:

$$W_{t,d} = t_{f_{t,d}} \times i_{d,f_t} \quad (1)$$

Dimana, (d) adalah dokumen d, (t) adalah kata t dari kata kunci, (W) adalah bobot dokumen d terhadap kata t, (tf) adalah banyaknya kata yang dicari pada sebuah dokumen, (idf) adalah *inverse document frequency*. Setelah bobot (W) dari dokumen telah diketahui, maka dilakukan proses pengurutan dimana jika semakin besar nilai W, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian juga sebaliknya.

2.1.4 Klasifikasi dengan *naïve bayes*

Menurut Chapman, NBC mempergunakan pelatihan data untuk mempelajari probabilitas dari keanggotaan kelasnya sebagai sebuah fungsi dari tiap variable prediktor dipertimbangkan secara independen "*naïve*". Ketika diaplikasikan pada data baru, keanggotaan kelas dilekatkan pada kategori yang dianggap paling memungkinkan berdasarkan probabilitas gabungan yang ditentukan oleh kombinasi prediktor - predictor [12]. Teorema dasar *naïve bayes* yang dipakai dalam penggunaan *machine learning*, dengan menggunakan aturan Bayes:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (2)$$

Peluang kejadian y terhadap x ditentukan dari peluang x saat y, peluang y, dan peluang x. Contoh pengaplikasiannya rumus ini dalam studi kasus pengompokan spam, menjadi :

$$p(\text{spam}|\text{word}) = \frac{p(\text{word}|\text{spam})p(\text{spam})}{p(\text{word})} \quad (3)$$

Bagian sisi kanan dari persamaan ini dapat dikomputerisasi dengan menggunakan data yang diberi label terlebih dahulu. Jika kita mengacu pada *nonspam* sebagai "ham" maka kita hanya perlu untuk mengkomputerisasi $p(\text{word}|\text{spam})$, $p(\text{word}|\text{ham})$, $p(\text{spam})$, dan $p(\text{ham}) = 1 - p(\text{spam})$, karena kita dapat menyusun penyebutnya dengan menggunakan formula yang kita gunakan sebelumnya dalam contoh tes berikut:

$$p(\text{word}) = p(\text{word}|\text{spam})p(\text{spam}) + p(\text{word}|\text{ham})p(\text{ham}) \quad (4)$$

Dengan kata lain, kita telah menghasilkan sebuah latihan berhitung: p spam akan menghitung email - email spam terhadap seluruh email, p word spam akan menghitung kemunculan email - email spam yang mengandung "word," dan p word ham menghitung kemunculan email - email ham yang mengandung "word".

2.2 Validation and Text Classification Performance Evaluation

2.2.1 Cross Validation

Cross validation merupakan sebuah metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi dua bagian, yang pertama adalah data latih yang digunakan sebagai data latih bagi mesin dan yang kedua digunakan untuk memvalidasi metode klasifikasi. *Cross Validation K-fold* digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi [13]. *10 fold cross validation* adalah salah satu *k fold* yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang tidak bias [13]. Dalam metode tersebut set data dibagi menjadi subset k dan di jalankan dalam perulangan k. Salah satu dari subset data k digunakan sebagai data pengujian sedangkan yang lainnya digunakan sebagai data pembelajaran dalam tiap perulangan. Ukuran performa dikalkulasi kepada seluruh perulangan k. Teknik ini dapat digunakan untuk menganalisa keakuratan klasifikasi terhadap data yang akan digunakan [14]

2.2.2 Evaluasi Performa Metode Klasifikasi Teks

Pengukuran performa klasifikasi teks ditujukan untuk menguji hasil klasifikasi dengan mengukur nilai performa yang telah di klasifikasi oleh mesin. Pengukuran tersebut dapat dilakukan dengan metode *Confusion Matrix* [14].

		Predicted Label	
		Positive	Negative
Actual Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Gambar 2 Performa Matriks

TP (*true positive*), klasifikasi kelas yang diprediksi positif dan diprediksi oleh mesin klasifikasi kelas positif. TN (*true negative*), klasifikasi kelas yang diprediksi negatif dan diprediksi oleh mesin klasifikasi kelas negative. FP (*false positive*), klasifikasi kelas negatif namun diprediksi oleh mesin klasifikasi kelas positif. FN (*false negative*), klasifikasi kelas positif namun diprediksi oleh mesin klasifikasi kelas negative.

Pengukuran akurasi menurut Syarif, untuk menguji tingkat akurasi dari metode klasifikasi teks yang telah dibuat dan dapat dihitung menggunakan rumus akurasi sebagai berikut:

$$Akurasi = \frac{TP + FN}{TP + FN + FP + FN} \quad (3.5)$$

Recall, menurut G. Williams pengukuran ini untuk mengidentifikasi seberapa besar metode dapat mengingat dengan pengelompokan teks yang tepat, rumus *recall* sebagai berikut [15]:

$$Recall = \frac{TP}{TP + FN} \quad (3.6)$$

Precision menurut G.Williams merupakan rasio pengukuran seberapa tepat metode klasifikasi teks dalam memprediksi teks. Rumus *precision* sebagai berikut [15]:

$$precision = \frac{TP}{TP + FP} \quad (3.7)$$

F-Measure menurut Targo, L merupakan pengabungan pengukuran *recall* dan *precision*, merupakan nilai tengah dari *recall* dan *precision*. Rumus *F-Measure* sebagai berikut [16]:

$$F - Measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (3.8)$$

Kappa menurut Carenini, G., (2011:21) digunakan untuk mengukur konsistensi antara tiap setiap pasang *annotator* dimana *annotator* tersebut digunakan untuk menilai pembuatan metode klasifikasi teks. Rumus *kappa* sebagai berikut:

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (3.9)$$

Dimana P(A) merupakan merupakan *annotator* benar (TP/(TP+TN)), dan nilai P(E) merupakan proporsi yang diperkirakan berdasarkan kesepakatan dari hasil penelitian terdahulu. Nilai Interpretasi kappa [18].

Tabel 1 Interpretasi Nilai Kappa

Nilai Kappa	Keterangan
<0	Kurang tepat
0.01 - 0.20	Sedikit tepat
0.21 - 0.40	Cukup tepat
0.41 - 0.60	Layak
0.61 - 0.80	Tepat
0.81 - 0.99	Hampir Sempurna

2.3 Topic modeling

Topic modeling, dokumen teks terdiri dari kata-kata, topik yang dituliskan dalam banyak dokumen dapat dinyatakan dengan kombinasi kata-kata yang saling terkait. *Topic modeling* merupakan teknik yang digunakan untuk menyimpulkan topik yang tersembunyi dalam dokumen teks. Karena *topic modeling* mewakili setiap dokumen sebagai kombinasi kompleks dari beberapa topik dan setiap topik sebagai kombinasi kompleks dari beberapa kata, itu juga digunakan sebagai alat *text mining* untuk mengklasifikasikan dokumen berdasarkan hasil lkesimpulan topik[18].

3 Pembahasan

3.1 Sentimen Analisis

3.1.1 Klasifikasi Data

Pada tahap ini data akan dibagi menjadi , yaitu data latih dan data uji. Pembagian data uji sebesar 70% dan data latih sebanyak 30% dari jumlah total data yang ada [19]. Selanjutnya peneliti mengkategorikan data latih menjadi 3 label yaitu *sentiment positive, negative, dan neutral*. Pada hasil dari pengkategorian tersebut label *neutral* berisi hal-hal seperti iklan, *update location*, atau memuat tautan tertentu. Contoh teks yang dikategorikan oleh peneliti sebagai berikut,

Tabel 2 Contoh Klasifikasi teks berdasarkan sentiment

No.	Text	Sentiment
1	SEC Chief @MichaelPiowowar says everyone should be able to invest in startups like Uber and Airbnb	Positive
2	Who's looking for a place to stay in Truckee North Lake Tahoe!	Neutral
3	@FinallyEven @Airbnb I hope more people find out what a scam Airbnb is. Destroying property seems to be a common thread w many people.	Negative

3.1.2 Performa Klasifikasi *Naïve bayes*

Dengan menggunakan metode klasifikasi sentiment menggunakan *Naïve bayes*, diperoleh nilai performa dari metode tersebut, seperti dibawah ini:

Tabel 3 Hasil Performa Klasifikasi *Naïve bayes*

No	Accuracy	Kappa	Precision	Recall	F-Measure
1	77.70%	66.10%	77.74%	76.80%	77.26%

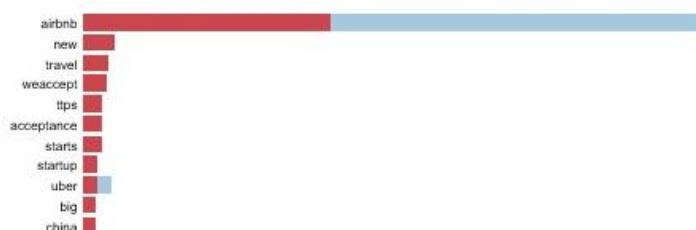
Dari hasil pada table 3 bahwa performa klasifikasi menggunakan metode *naïve bayes* menghasilkan nilai *accuracy* sebesar 77.70%, nilai *kappa* sebesar 66.10%, nilai *precision* sebesar 77.74%, dan nilai *recall* sebesar 76.80%. Nilai *kappa* lebih 60% sudah memasuki kategori *substantial agreemen* [20]. *Precision* adalah seberapa tepat sebuah model untuk dapat memprediksi hasil. Sedangkan *recall* adalah mengukur nilai sebuah label yang teridentifikasi oleh model [21].

3.1.3 Hasil Klasifikasi Sentimen

Untuk dapat melihat kelebihan maupun kekurangan dari *Airbnb*, peneliti memperdalam analisis dengan menggunakan analisis sentiment. Peneliti membagi data menjadi 3 kategori sentiment *positive, negative, dan neutral*. Dari total data sebanyak 374241 *tweets* terdapat 117453 *tweets* bersentimen positive, 48787 *tweets* bersentimen *negative*, 208001 *tweets* bersentimen *neutral*. Selanjutnya klasifikasi ini akan dimasukan dimodelkan kedalam *topic modeling* guna memunculkan topik-topik yang ada pada jaringan komunikasi *twitter*.

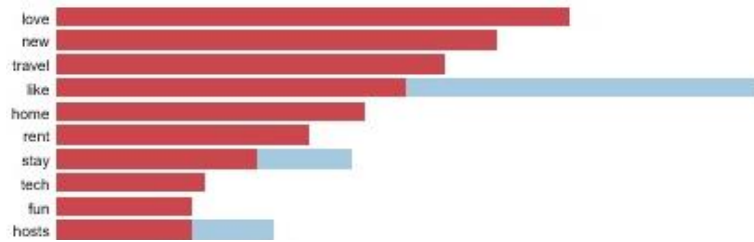
3.2 Topic Modeling

Untuk dapat mempertajam analisis data percakapan mengenai *Airbnb*, peneliti ingin memunculkan topik-topik yang muncul pada data percakapan tersebut. Untuk dapat memetakan topik yang dibahas, peneliti hanya menggunakan dua data sentimen yaitu sentiment positif dan negatif. Sentimen netral tidak digunakan karena setelah peneliti membaca text pada sentiment netral, kebanyakan dari data tersebut adalah *data noise* seperti iklan, *update location*, hal-hal yang tidak memiliki kecenderungan sentiment. Dari hasil klasifikasi data yang akan digunakan dalam topik modeling, untuk mengetahui topik yang ada pada data tersebut, peneliti menggunakan metode penghitungan *saliency (term's overall frequency)*. Metode penghitungan tersebut menghitung probabilitas seringnya sebuah topik muncul dengan menggunakan komputasi *Kullback-Liebler divergence* [22].



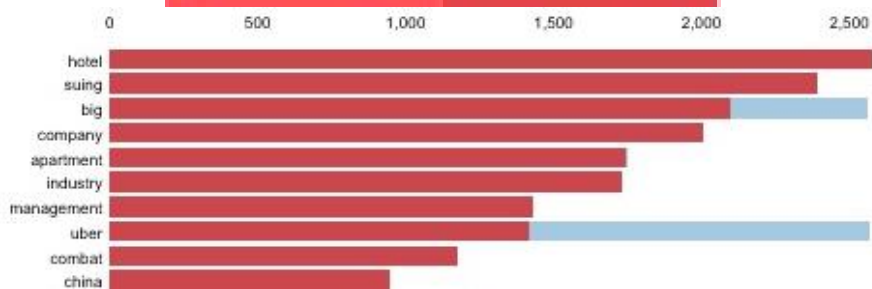
Gambar 3 Topik kelesuruhan *airbnb*

Pada gambar 3 terlihat kata-kata yang sering muncul seperti *airbnb*, *new*, *travel*, *weaccept*, *tips*, *startup*, dll. Pada topik kedua ini kebanyakan orang membahas mengenai kontroversi penutupan *airbnb* yang di nilai tidak tepat. Karena kebanyakan orang merasa bahwa *airbnb* adalah salah satu alternatif yang mempermudah para pelancong untuk mencari tempat tinggal dan kebanyakan orang berpendapat bahwa mereka setuju dengan adanya *airbnb*.



Gambar 4 Topik positif *airbnb*

Pada gambar 4 adalah topik kedua pada sentiment positif, banyak orang menyukai *airbnb* karena banyak pelancong merasa tertolong dengan adanya hal baru seperti *airbnb* yang menyediakan layanan sewa tempat dengan pengalaman tidak biasa seperti di hotel melainkan seperti di rumah atau apartemen sendiri sehingga nuansa lokal tempat berlibur lebih terasa, hal tersebut tercermin dari kata-kata *love*, *new*, *travel*, *like*, *home*, *rent*, *stay*, *fun*, *hosts*.



Gambar 5 Topik negatif *airbnb*

Gambar 5 adalah topik negative ketiga. Pada topik ini kebanyakan membahas mengenai banyaknya dukungan orang-orang kepada perusahaan-perusahaan perhotelan yang menuntut perusahaan *airbnb*. Hal ini tergambar dari kata-kata *hotel*, *suing*, *big*, *company*, *apartment*, *industry*.

4. Kesimpulan

Peneliti telah berhasil menerapkan analisis konten dengan jumlah data yang besar. Dalam penelitian ini mengenai *airbnb* telah berhasil mengungkap topik-topik yang ada pada sosial media *twitter*. Dan pada penelitian ini mengenai, *naïve bayes* untuk *sentiment analysis*, dan *Latent Dirichlet Allocation* untuk *topic modeling*. Berdasarkan pengamatan peneliti bahwa metode-metode diatas cocok untuk menangani analisis konten secara cepat dan tepat. Berdasarkan analisis dan pengujian yang dilakukan pada bab sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Media sosial yang digunakan saat ini cukup banyak seperti *twitter*, *facebook*, *Instagram*, dll. Pada penelitian ini hanya mengambil data *twitter* dan *facebook*. *Instagram* tidak digunakan karena inti dari sosial media *intagram* itu sendiri adalah bertukar gambar (foto). Pada saat pengambilan data sosial media *twitter* digunakan lebih banyak oleh orang untuk berinteraksi mengenai *airbnb*. Pada pola jaringan sosial media *twitter* orang-orang kebanyakan membicarakan mengenai masalah penutupan *airbnb* yang dinilai tidak tepat bagi sebagian orang.
2. Untuk memperdalam analisis diperlukan sentimen analisis agar dapat mengetahui apa kelebihan dan kekurangan mengenai produk atau jasa layanan. Untuk melakukan analisis sentimen dibutuhkan metode klasifikasi yang tepat untuk dapat memperoleh hasil yang baik. Peneliti menggunakan metode *naïve bayes* yang memberikan nilai akurasi klasifikasi yang cukup baik dan cepat.
3. Dalam kasus penelitian mengenai *airbnb* mampu mengetahui apa yang menjadi kekurangan atau inti complain dari konsumen secara tepat. Dari hasil penelitian sebanyak 117452 *tweets* bersentimen positif dan 48786 *tweets* bersentimen negatif. Pada data sentimen positif tersebut dikarenakan saat periode pengambilan data banyak orang yang memberikan dukungan maupun pujian kepada *airbnb* yang sedang dituntut oleh perusahaan-perusahaan perhotelan. Sedangkan sentimen negatif kebanyakan menganggap bahwa *airbnb* hanya menjadi masalah keamanan di daerah mereka dan kenyamanan mereka yang terganggu oleh adanya *airbnb*.

DaftarPustaka

- [1] United Nation World Tourism Organization, 2017, "UNWTO tourism highlights 2017 edition"
- [2] Airbnb, 2017. *Overview*. [Online] Available: <https://www.airbnb.com/trust>
- [3] Statistics portal, 2017, diakses 5 Maret 2018, "company value and equity funding of Airbnb from 2014-2017". [Online] Available: <https://www.statista.com/statistics/339845/company-value-and-equity-funding-of-airbnb/>
- [4] Businessinsider, 2017, diakses 5 Maret 2018, "the \$10 billion club: meet the most valuable startups in the US" , [Online] Available: <http://www.businessinsider.com/most-valuable-us-startups-10-billion-decacorns-2017-12/?IR=T/#7-dropbox-1>
- [5] Alexander, Hilda. 2016. Diakses 11 Maret 2018. "Kunjungan wisata meningkat, pembangunan hotel melesat". [Online] Available: <https://properti.kompas.com/read/2016/03/07/060000121/Kunjungan.Wisata.Meningkat.Pembangunan.Hotel.Melesat>
- [6] Feldman, R., & Sanger. J. 2007. *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- [7] Zhao, Yanchang *et al.* 2014. *Data Mining Applications with R*. Amsterdam: Elsevier.
- [8] Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. San Rafael, USA: Morgan&Claypool.
- [9] Ting, S., W.H. Ip., Tsang, Albert., 2015., "Is Naive Bayes a Good Classifier for Document Classification?"., *International Journal of Software Engineering and its Applications* Vol.5, No.3. pp 37-46
- [10] Hofmann, M., Chisholm, A., Areal, N., Buckley, P., Carter, B., Colton, D., . . . Silipo, R. (2016). *Text Mining and Visualization: Case Studies Using Open-Source Tools*. CRC Press.
- [11] Ingersoll, G.S., Morton, T.S., Farris, A.L., 2013. *Taming Text How to Find, Organize and Manipulate It*. Shelter Island, NY : Manning.
- [12] Chapman, C., & Feit, E. Mc. 2015. *R for Marketing Research and Analytics*. Switzerland: Springer.
- [13] Wibowo, Antoni. 2017. 10 Folf-Cross Validation. Master of Information Technology, Binus University. Didapat dari: <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>
- [14] Syarif, Iwan. 2014. comprehensive review of classification alhorithms for high dimensional datasets. University of southhampton, Electronics and Computer Science, PdD thesis, 19-25, 39-40.
- [15] G. Williams. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. LLC: Springer Science Business Media
- [16] Targo, Luis. 2011. *Dara Mining with R with Case Studies*. New York: Chapman&Hall
- [17] Carenini, G., Murray, G., & Ng, R. 2011. *Methods for mining and summarizing text conversations*. Morgan&Claypool : Colombia.
- [18] Jeong, Byeongki., Yoon, Janghyeok., Lee, Jae-min. 2017. "Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis". *International Journal of Information Management*, <http://dx.doi.org/10.1016/j.ijinfomgt.2017.09.009>
- [19] Ting, S., W.H. Ip., Tsang, Albert., 2015., "Is Naive Bayes a Good Classifier for Document Classification?"., *International Journal of Software Engineering and its Applications* Vol.5, No.3. pp 37-46
- [20] Landis, Richard., Koch, Gary., 1977., "An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers", *International Biometric Society* Vol.33, No.2. pp 363-374
- [21] Williams, G. (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. Springer Science Business Media.
- [22] Sievert, Carson., Shirley, Kenneth., 2014., "LDAvis : A method for visualizing and interpreting topics", *proceedings of the workshop on interaactive Language Learning, Visualization, and Interfaces*, pp 63-70, June 2014.