Coarse-Grained Sentiment Analysis Berbasis Natural Language Processing – Ulasan Hotel

by Warnia Nengsih

Submission date: 22-Mar-2021 08:31PM (UTC-0700)
Submission ID: 1539977720
File name: paper_jnteti.pdf (1.32M)
Word count: 2594
Character count: 16412

Coarse-Grained Sentiment Analysis Berbasis *Natural Language Processing* – Ulasan Hotel

(*Coarse-Grained Sentiment Analysis Based on Natural Language Processing - Hotel Review*)

Warnia Nengsih[1], M. Mahrus Zein[2], Nazifa Hayati[3]

***Abstract*—Sentiment analysis is a method for obtaining data from various platforms available on the internet. Advances in technology enable the machine to recognize a term that is considered a positive opinion and vice versa. These data and opinions play an important role as product, services, or other topic feedback. Without the need to obtain an opinion directly from the public, the provider has obtained an important evaluation to develop themselves. Hospitality business is a field related to services, providing services to customers. Indicators of business continuity also depend on customer feedback and serve as a reference for strategic policy. Sentiment analysis techniques based on Natural Language Processing are expected to overcome these problems. In this study, the prediction uses a temporary Random Forest (RF) classifier to summarize the quality of the classifier then it can be done using the Receiver Operating Characteristic (ROC) curve. The ROC curve is a good graphic to summarize the quality of the classifier. The higher the curve is above the diagonal line, the better the prediction, with the ROC Curve value of 0.90. The result shows that positive reviews are more than the negative reviews, i.e., 68% and 32%, respectively.**

***Intisari*—*Sentiment analysis* adalah metode untuk memperoleh data dari berbagai *platform* yang tersedia di internet. Kemajuan teknologi memungkinkan mesin untuk mengenali suatu istilah yang dianggap sebagai opini positif maupun sebaliknya. Data-data dan opini tersebut berperan penting sebagai umpan balik produk, layanan, dan topik lainnya. Tanpa perlu memperoleh opini secara langsung dari masyarakat, pihak penyedia telah mendapatkan evaluasi yang penting guna mengembangkan diri. Bisnis perhotelan merupakan bidang yang terkait dengan jasa memberikan layanan pada pelanggan. Indikator keberlangsungan bisnis ini juga bergantung pada umpan balik pelanggannya dan dijadikan sebagai acuan untuk pengambilan kebijakan strategis. Teknik *sentiment analysis* berbasis *Natural Language Processing* dapat mengatasi permasalahan tersebut. Pada makalah ini prediksi dilakukan menggunakan *classifier Random Forest* (*RF*), sementara untuk merangkum kualitas *classifier*, digunakan kurva *Receiver Operating Characteristic* (ROC). Kurva ROC berupa grafik yang baik untuk merangkum kualitas *classifier*. Semakin tinggi kurva berada di atas garis diagonal, semakin baik prediksinya, dengan nilai kurva ROC yang diperoleh sebesar 0,90. Terlihat hasil ulasan terhadap opini pelanggan terhadap jasa dan pelayanan yang diberikan oleh hotel untuk kategori positif lebih banyak daripada kategori negatif. Polaritas dari ulasan diperoleh 68% ulasan pelanggan berada pada area positif dan 32% berada pada area negatif.**

***Kata Kunci*—*Coarse Grained*, *Sentiment Analysis*, NLP, Hotel.**

[1,2,3] *Jurusan Teknik Informatika, Politeknik Caltex Riau, Jl. Umbansari No Rumbai Pekanbaru Riau 28265 (tlp: 0761-53939; fax: 0761-554224; e-mail: [1]warnia@pcr.ac.id, [2]mahrus@pcr.ac.id)*

I. PENDAHULUAN

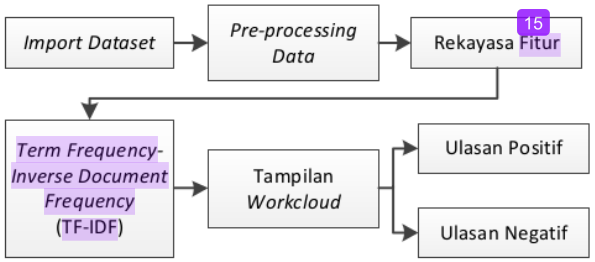
Sentiment analysis merupakan bagian dari *Natural Language Processing* (NLP). Teknik ini sangat baik untuk menentukan hasil ulasan (*review*) dari sebuah opini pengguna (*user*). *Sentiment analysis* adalah salah satu teknik ekstraksi dari sebuah informasi terhadap sebuah isu dan kejadian. Teknik ini digunakan untuk menemukan opini dan paparan terhadap suatu isu atau kejadian dalam bentuk teks. Sentimen merupakan pernyataan subjektif yang mencerminkan persepsi seseorang terhadap suatu peristiwa [1]. Ekstraksi dari opini masyarakat mengenai produk atau layanan dari hampir semua bidang. Secara umum, *sentiment analysis* terbagi menjadi dua kategori besar, yaitu:

- *coarse-grained sentiment analysis*,
- *fined-grained sentiment analysis*.

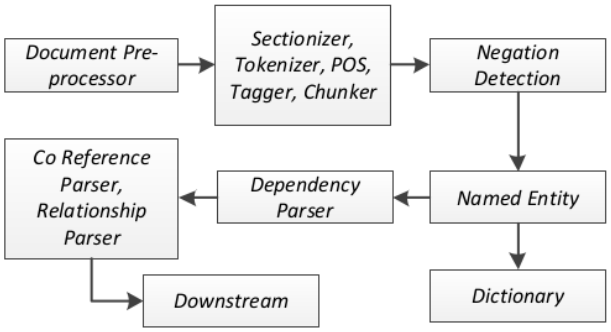
Coarse-grained sentiment analysis melakukan proses analisis pada level dokumen. Pengklasifikasian berorientasi pada sebuah dokumen secara keseluruhan, yaitu positif, netral, dan negatif. Sementara itu, *fined-grained sentiment analysis* melakukan proses analisis sebuah kalimat [2].

Tujuan makalah ini adalah menghasilkan *sentiment analysis* pada ulasan pelanggan hotel sehingga pola yang diperoleh dapat dijadikan sebagai acuan dalam pengambilan kebijakan berikutnya, agar dapat membantu meningkatkan kualitas pelayanan dengan nilai akurasi yang lebih baik. Ulasan produk yang dibuat oleh pengguna *online* dapat berdampak terhadap keputusan membeli pelanggan lain [3]. Ulasan dapat membantu pelanggan dalam membentuk kriteria untuk mengevaluasi produk dan mengurangi biaya kognitif dalam membuat keputusan pembelian [4]. Ulasan produk *online* juga dapat membantu pelanggan untuk: (1) membentuk pemahaman tentang suatu produk; (2) membangun kriteria untuk mengevaluasi produk; (3) membantu membuat keputusan yang tepat; dan (4) mengurangi biaya kognitif dalam membuat keputusan.

Berikut merupakan ulasan penelitian terdahulu terkait dengan makalah ini. Beberapa penelitian menggunakan metode *Naive bayes* dan *rapid miner* serta *data miner* sebagai *tools* pengolahannya. Pengujian validitas data menggunakan 10-*fold cross validation* dengan rata-rata nilai akurasi sebesar 89%. Sumber *dataset* yang digunakan diperoleh dari *Hotels.com*,



Gbr. 1 Alur perancangan *coarse grained sentiment analysis*.



Gbr. 2 Mekanisme *Natural Langunge Processing* (NLP).

booking.com, dan *agoda.com* serta menggunakan Python untuk pengolahannya. *Dataset* yang digunakan bersumber dari *dataset hotel_reviews.csv*, berbasis *coarse-grained sentiment analysis* berbasis NLP. Sementara itu, untuk menghitung kinerja algoritme klasifikasi yang ditampilkan dalam bentuk grafik, digunakan kurva *Receiver Operating Characteristic* (*ROC*) [4]-[6].

Bisnis perhotelan merupakan bidang yang erat kaitannya dengan kepuasan pelanggan. Kepuasan pelanggan menjadi tolok ukur keberhasilan dari rumusan capaian yang diperoleh pada sebuah bisnis. Indikator keberlangsungan bisnis ini juga bergantung pada respons pelanggannya dan dijadikan sebagai acuan untuk pengambilan kebijakan strategis. Permasalahan yang terjadi adalah level manajemen tidak mengetahui dan belum menemukan pola dari respons pelanggan hotel terkait layanan produk dan jasa yang diberikan. Penerapan teknik *sentiment analysis* berbasis NLP diharapkan dapat mengatasi permasalahan yang berhubungan dengan umpan balik pelanggan terhadap layanan produk atau jasa yang diberikan. Aktivitas untuk merangkum kualitas *classifier* dilakukan menggunakan kurva ROC. Kurva ROC berupa grafik yang baik untuk merangkum kualitas *classifier*. Semakin tinggi kurva berada di atas garis diagonal, semakin baik hasil prediksi [7]-[10].

II. METODOLOGI

Dataset yang digunakan adalah *dataset hotel_reviews.csv*. *Dataset* ini merupakan kumpulan ulasan pelanggan dari salah satu situs perjalanan terkemuka. Pada *dataset* ini terdapat 51.574 baris dan 3.840 atribut data [11]. *Dataset* ini berisi komentar pelanggan terhadap layanan yang terdapat pada sebuah hotel. Tentunya komentar dan hasil ulasan memengaruhi keputusan pelanggan untuk melakukan *order* terhadap produk yang ditawarkan, sehingga manajemen hotel harus dapat melakukan ekstraksi dan menggali informasi kecenderungan komentar atau ulasan tersebut berada dalam kelompok atau kategori tetentu. Dengan hasil tersebut, diperlukan sebuah *knowledge* untuk dijadikan sebagai acuan terkait dengan pengambilan keputusan strategis masa depan. Jenis *sentiment analysis* yang digunakan adalah *coarse grained sentiment analysis*, yang merupakan jenis *sentiment analysis* yang dilakukan pada level dokumen, dengan seluruh isi dokumen dianggap sebagai sebuah sentimen positif dan negatif. Isi dokumen pada *dataset* ini diperoleh dari beberapa variabel yang digunakan, di antaranya ulasan deskripsi dari setiap pelanggan, *browser*, serta *device* yang digunakan.

III. PERANCANGAN

Berikut merupakan alur perancangan *coarse grained sentiment analysis*. *Dataset* yang akan diolah bersumber dari *hotel_reviews.csv*, yang selanjutnya dikenai *data pre-processing*, *term frequency-inverse document frequency*, untuk menentukan kategori ulasan positif dan ulasan negatif yang dihasilkan.

Gbr. 1 memberikan gambaran tentang alur perancangan *coarse grained sentiment analysis*. *Dataset* yang digunakan harus melalui tahapan *data pre-processing*. Bagian ini dikerjakan dengan tahapan *text pre-processing*. Proses *pre-processing* ini meliputi: (1) *case folding*, (2) *tokenizing*, (3) *filtering*, dan (4)

Out[1]:

	review	is_bad_review
0	I am so angry that i made this post available...	1
1	No Negative No real complaints the hotel was g...	0
2	Rooms are nice but for elderly a bit difficul...	0
3	My room was dirty and I was afraid to walk ba...	1
4	You When I booked with your company on line y...	0

Gbr. 3 Keluaran pengelompokan kategori.

In[4]:

```
%%time

# return the wordnet object value corresponding to the POS tag
from nltk.corpus import wordnet

def get_wordnet_pos(pos_tag):
    if pos_tag.startswith('J'):
        return wordnet.ADJ
    elif pos_tag.startswith('V'):
        return wordnet.VERB
    elif pos_tag.startswith('N'):
        return wordnet.NOUN
    elif pos_tag.startswith('R'):
        return wordnet.ADV
    else:
        return wordnet.NOUN

import string
from nltk import pos_tag
from nltk.corpus import stopwords
from nltk.tokenize import WhitespaceTokenizer
from nltk.stem import WordNetLemmatizer
```

Gbr. 4 *Data pre-processing*.

stemming. Tahap akhir dari *text pre-processing* adalah *term-weighting*. *Term-weighting* merupakan proses pemberian bobot *term* pada dokumen. Pembobotan ini digunakan untuk klasifikasi dokumen menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

Gbr. 2 menjelaskan tentang mekanisme NLP. Mekanisme ini dimulai dari tahapan *document pre-processor*, *sectionizer*, *tokenizer*, *POS tagger*, *chunker*, sampai pada tahapan *downstream application*.

IV. PEMBAHASAN

Dari deskripsi ulasan yang terdapat pada *dataset* yang digunakan, terlebih dahulu perlu dilakukan pengelompokan kategori. Pandas berguna dalam memasukkan *file* data ke dalam

Python.reviews_df["review"]=reviews_df["Negative_Review"]+reviews_df["Positive_Review"]

yang berfungsi untuk membagi setiap ulasan tekstual, yaitu dibagi menjadi bagian positif dan negatif dan mengelompokkannya.

Gbr. 3 memperlihatkan keluaran yang dihasilkan dalam pengelompokan kategori positif dan negatif. Pada gambar tersebut terlihat hasil pengelompokan setiap ulasan dari komentar yang ada. Ulasan positif diinisialisasi dengan nilai 1 dan ulasan negative diinisialisasi dengan nilai 0.

Proses pembersihan data dilakukan dengan melakukan: *import wordnet, import String, import pos_tag, import stopword, import WhitespaceTokenizer*, dan *import WordNetLemmatizer*. *Pos_tag* adalah simbol yang mewakili

```
def clean_text(text):
    # lower text
    text = text.lower()
    # tokenize text and remove puncutation
    text = [word.strip(string.punctuation) for word in text.split(" ")]
    # remove words that contain numbers
    text = [word for word in text if not any(c.isdigit() for c in word)]
    # remove stop words
    stop = stopwords.words('english')
    text = [x for x in text if x not in stop]
    # remove empty tokens
    text = [t for t in text if len(t) > 0]
    # pos tag text
    pos_tags = pos_tag(text)
    # lemmatize text
    text = [WordNetLemmatizer().lemmatize(t[0], get_wordnet_pos(t[1])) for t in pos_tags]
    # remove words with only one letter
    text = [t for t in text if len(t) > 1]
    # join all
    text = " ".join(text)
    return(text)

# clean text data
reviews_df["review_clean"] = reviews_df["review"].apply(lambda x: clean_text(x))
```

```
CPU times: user 1min 46s, sys: 2.1 s, total: 1min 48s
Wall time: 1min 55s
```

Gbr. 5 Proses pembersihan data tekstual.

```
In[6]:
# add number of characters column
reviews_df["nb_chars"] = reviews_df["review"].apply(lambda x: len(x))

# add number of words column
reviews_df["nb_words"] = reviews_df["review"].apply(lambda x: len(x.split(" ")))
```

Gbr. 6 Penambahan karakter dan teks.

kategori leksikal: NN (*noun*), VB (*verb*), JJ (*adjective*), dan AT (*article*).

Gbr. 4 merupakan proses untuk tahapan *data pre-processing*. Ada beberapa tahapan yang dilakukan, mulai dari tahapan pengelompokan kata sampai ke proses *tokenizer*.

Gbr. 5 menjelaskan tentang pembersihan data tekstual menggunakan fungsi *'clean text'* dan dilakukan beberapa transformasi, seperti *lower text* (membuat huruf menjadi kecil); *tokenize the text* (memisahkan teks menjadi kata-kata); menghapus tanda baca; menghapus kata-kata yang berisi angka; menghapus kata-kata *stop word*, seperti '*the*','*a*', dan '*this*'; menghapus tanda yang kosong; menandai *Part-of-Speech* (POS): menetapkan *tag* untuk setiap kata untuk didefinisikan jika sesuai dengan kata benda, kata kerja; serta *lemmatize* teks (mengubah setiap kata menjadi bentuk dasarnya).

Tahap selanjutnya adalah rekayasa fitur, dimulai dengan menambahkan fitur *sentiment analysis*. Proses ini menggunakan Vader, yang merupakan bagian dari modul NLTK yang dirancang untuk *sentiment analysis*. Vader menggunakan kamus kata untuk menemukan kata-kata yang masuk ke dalam kategori positif atau negatif.

Gbr. 6 menjelaskan penambahan beberapa metrik sederhana untuk setiap teks. Penambahan teks ini dapat berupa penambahan kolom untuk jumlah karakter dalam teks dan penambahan kolom untuk jumlah kata dalam teks.

Pada Gbr. 7 diperlihatkan modul Gensim membuat representasi vektor numerik dari setiap kata dalam korpus dengan menggunakan konteks *Word2Vec*. Selanjutnya, ditambahkan nilai TF-IDF untuk setiap kata dan setiap dokumen, dengan *term frequency* menghitung jumlah

```
In[7]:
# create doc2vec vector columns
from gensim.test.utils import common_texts
from gensim.models.doc2vec import Doc2Vec, TaggedDocument

documents = [TaggedDocument(doc, [i]) for i, doc in enumerate(reviews_df["review_clean"].apply(lambda x: x.split(" ")))]

# train a Doc2Vec model with our text data
model = Doc2Vec(documents, vector_size=5, window=2, min_count=1, workers=4)

# transform each document into a vector data
doc2vec_df = reviews_df["review_clean"].apply(lambda x: model.infer_vector(x.split(" "))).apply(pd.Series)
doc2vec_df.columns = ["doc2vec_vector_" + str(x) for x in doc2vec_df.columns]
reviews_df = pd.concat([reviews_df, doc2vec_df], axis=1)
```

Gbr. 7 Representasi vektor numerik.

```
In[12]:
# wordcloud function

from wordcloud import WordCloud
import matplotlib.pyplot as plt

def show_wordcloud(data, title = None):
    wordcloud = WordCloud(
        background_color = 'white',
        max_words = 200,
        max_font_size = 40,
        scale = 3,
        random_state = 42
    ).generate(str(data))

    fig = plt.figure(1, figsize = (20, 20))
    plt.axis('off')
    if title:
        fig.suptitle(title, fontsize = 20)
        fig.subplots_adjust(top = 2.3)

    plt.imshow(wordcloud)
    plt.show()

# print wordcloud
show_wordcloud(reviews_df["review"])
```

Gbr. 8 Fungsi *wordcloud*.

kemunculan kata klasik dalam teks, sementara *inverse document frequency* menghitung kepentingan relatif dari kata, yang tergantung pada banyak teks kata dapat ditemukan.

Untuk menampilkan jumlah ulasan negatif, menggunakan *reviews_df*, Hasil dari *reviews df* menunjukkan hasil untuk nilai negatif dan nilai positif dengan persentase 95% untuk nilai positif dan 0,43% untuk komentar negatif.

Gbr. 8 menunjukkan tahapan untuk menampilkan *wordcloud*. Tampilan *wordcloud* digunakan untuk melihat sekilas kata-kata yang muncul di ulasan. *Wordcloud* adalah teknik visualisasi data yang digunakan untuk mewakili data teks dengan ukuran setiap kata menunjukkan frekuensi. Pada tampilan ini terdapat kata dengan frekuensi kemunculan yang berulang, di antaranya: "*staff*", "*excellent*", "*nothing*", "*hotel*", "*friendly*", dan "*breakfast*".

Gbr. 9 merupakan hasil dari fungsi *wordcloud* yang sudah dilakukan. Terdapat beberapa kata yang mewakili kemunculan kata. Ulasan positif tertinggi lebih dari lima kata untuk sepuluh data, seperti ditunjukkan pada Gbr. 10. Ada beberapa kata positif yang terdeteksi, seperti: "*perfect*" ,"*clean*", "*lovely staff*", "*great place*","*beautiful*", dan "*good value*".

Gbr. 11 menampilkan ulasan negatif tertinggi lebih dari lima kata untuk sepuluh data. Ada beberapa kata negatif yang terdeteksi, seperti: "*dislikes*", "*noisy*", "*very bad*", dan "*nothing great*".

Gbr. 12 menunjukkan visualisasi yang dihasilkan, dengan warna biru pada grafik menunjukkan ulasan positif dan warna oranye menunjukkan ulasan negatif. Pada grafik tersebut terlihat hasil ulasan terhadap opini pelanggan untuk kategori positif lebih banyak daripada kategori negatif.

Pada *good reviews* terlihat bahwa statistik yang ditampilkan naik turun, sementara *bad reviews* masih berada pada ulasan atau komentar negatif dalam kondisi yang normal.

Selanjutnya adalah memilih fitur yang akan digunakan pada data *training* lalu mengelompokkannya menjadi data *training* dan *testing*. Untuk melakukan prediksi, digunakan *classifier Random Forest* (RF).

Gbr. 9 Hasil *wordcloud function*.

Out[13]:

	review	pos
43101	A perfect location comfortable great value	0.931
211742	Clean comfortable lovely staff	0.907
175551	Friendly welcome Comfortable room	0.905
365085	Good location great value	0.904
109564	Clean friendly and comfortable	0.902
145743	Good value amazing location	0.901
407590	breakfast excellent Clean comfort	0.899
407546	Great place I enjoyed	0.881
218571	Beautiful Quirky Comfortable	0.878
436901	Lovely comfortable rooms	0.877

Gbr. 10 Ulasan positif.

Out[14]:

	review	neg
193086	No dislikes LOCATION	0.831
318516	A disaster Nothing	0.804
29666	A bit noisy No	0.796
426057	Dirty hotel Smells bad	0.762
263187	Very bad service No	0.758
181508	Window blind was broken	0.744
174178	no bad experience location	0.740
291281	nothing great clean comfortable quite hotel	0.733
233344	It was awful No	0.722
384771	Very bad atmosphear noisy weird smells unfrie...	0.713

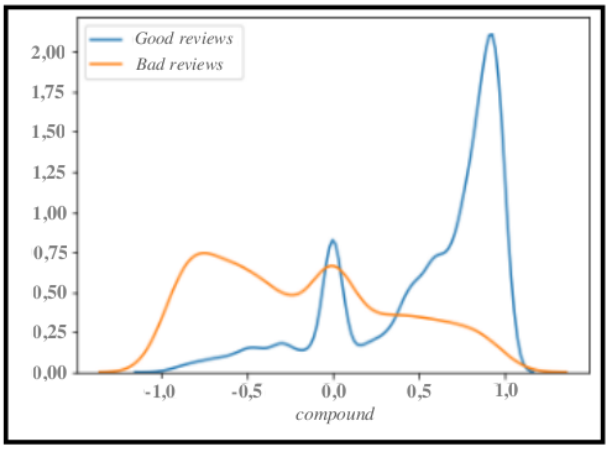
Gbr. 11 Ulasan negatif.

Gbr. 13 menunjukkan pengelompokan pada data *training* secara acak untuk melakukan prediksi dengan nilai estimasi sebesar 100 dan *random states* 42, dengan fitur *yang* ditampilkan hanya dua puluh *record* saja.

Untuk merangkum kualitas *classifier* digunakan kurva ROC. Kurva ROC merupakan represen[14] grafis untuk menggambarkan keakuratan diagnostik. Kurva berada di atas garis diagonal. Semakin tinggi kurva berada di atas garis diagonal, semakin baik nilai prediksi yang dihasilkan. Diperoleh nilai kurva ROC sebesar 0,90. Polaritas dari ulasan yang diperoleh yaitu 68% ulasan pelanggan berada pada area positif dan 32% berada pada area negatif, dengan nilai akurasi prediksi 90%.

V. KESIMPULAN

Sentiment analysis pada ulasan pelanggan hotel menghasilkan klasifikasi ulasan untuk kategori positif dan negatif dari komentar yang diberikan pelanggan. Pola yang diperoleh dapat dijadikan sebagai acuan dalam pengambilan kebijakan bagi level manajemen. Rangkuman kualitas *classifier* pada makalah ini menggunakan kurva ROC dengan nilai sebesar 0,90. Semakin tinggi kurva berada di atas garis diagonal, semakin baik prediksi. Polaritas dari ulasan yang diperoleh adalah 68% ulasan pelanggan berada pada area positif dan 32% berada pada area negatif, dengan nilai akurasi prediksi sebesar 90%.



Gbr. 12 Ulasan warna positif dan negatif.

Out[17]:

	feature	importance
3	compound	0.038355
2	pos	0.024718
0	neg	0.023400
8	doc2vec_vector_2	0.020461
6	doc2vec_vector_0	0.019230
7	doc2vec_vector_1	0.017899
9	doc2vec_vector_3	0.016996
10	doc2vec_vector_4	0.016483
4	nb_chars	0.016391
1	neu	0.014577
5	nb_words	0.013880
2239	word_nothing	0.009913
2853	word_room	0.009530
950	word_dirty	0.009411
285	word_bad	0.008820
3202	word_staff	0.006799
1945	word_location	0.006683
1639	word_hotel	0.006280
3216	word_star	0.006267
2867	word_rude	0.005898

Gbr. 13 Pengelompokan pada *data training* secara acak.

REFERENSI

[1] K. Zvarevashe dan O. O. Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews," *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, 2018, hal. 1–4.

[2] G. Walsh, K.P. Gwinner, dan S.R. Swanson, "What Makes Mavens Tick? Exploring the Motives of Market Mavens' Initiation of Information Diffusion, *Journal of Consumer Marketing*, Vol. 21, No. 2, hal. 109-122, 2014.

[3] Y. Liu, P. Li, dan S. Liu, "Opinion Mining and Sentiment Analysis," *Zhonghua Bing Li Xue Za Zhi*, Vol. 24, No. 2, hal. 72–74. 1995.

[4] T. Ghorpade dan L. Ragha, "Featured Based Sentiment Classification for Hotel Reviews Using NLP and Bayesian Classification," *Proc. Int. Conf. Commun. Inf. Comput. Technol. (ICCICT)*, 2012, hal. 1–5.

[5] V.B. Raut dan D.D. Londhe, "Opinion Mining and Summarization of Hotel Reviews," *Proc. - 2014 6th Int. Conf. Comput. Intell. Commun. Networks, (CICN)*, 2014, hal. 556–559.

[6] P. Prameswari, I. Surjandari, dan E. Laoh, "Opinion Mining from Online Reviews in Bali Tourist Area," *Proc. 3rd Int. Conf. Sci. Inf. Technol. (ICSITech)*, 2017, hal. 226–230.

[7] P. Juneja dan U. Ojha, "Casting Online Votes: To Predict Offline Results Using Sentiment Analysis by Machine Learning Classifiers," *8th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT2017)*, 2017, hal. 1-6.

[8] M. Abbas, K.A. Memon, A.A. Jamali, S. Memon, dan A. Ahmed, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," *Int. Journal of Computer Science and Network Security (IJCSNS)*, Vol. 19, No. 3, 2019, hal. 62–67.

[9] S. George dan S. Joseph, "Text Classification by Augmenting Bag of Words (BOW) Representation with Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature", *IOSR Journal of Computer Engineering*, Vol. 16, No. 1, hal. 34-38, 2014.

[10] E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swam Optimization," *J. Evolusi*, Vol. 4, No. 2, hal. 20-27, 2016.

[11] (2019) "Hotel Reviews", [Online], https://www.kaggle.com/datafiniti/hotelreviews, tanggal akses: 1-Sep-2019.

Coarse-Grained Sentiment Analysis Berbasis Natural Language Processing – Ulasan Hotel

ORIGINALITY REPORT

14%	14%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.dikamp.web.id Internet Source	3%
2	opac.lib.pcr.ac.id Internet Source	2%
3	ammograde97.blogspot.com Internet Source	2%
4	123dok.com Internet Source	1%
5	jurnal.untan.ac.id Internet Source	1%
6	sekolahmalam.wordpress.com Internet Source	1%
7	www.coursehero.com Internet Source	1%
8	yunusmuhammad007.medium.com Internet Source	1%
9	dance.csc.ncsu.edu	

	Internet Source	1%
10	mti.binus.ac.id Internet Source	<1%
11	Muhammad Romy Firdaus, Fikri Muhammad Rizki, Favian Muhammad Gaus, Indra Kusumajati Susanto. "Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru", J-SAKTI (Jurnal Sains Komputer dan Informatika), 2020 Publication	<1%
12	text-id.123dok.com Internet Source	<1%
13	www.neliti.com Internet Source	<1%
14	journal.fkm.ui.ac.id Internet Source	<1%
15	libraryeproceeding.telkomuniversity.ac.id Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		