

**ANALISIS JARINGAN SEMANTIK PADA FORUM
TRIPADVISOR UNTUK PENINGKATAN DAYA TARIK
WISATA
Tugas Akhir**

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat Sarjana
Komputer**



Dibuat Oleh:

RANDYA MAHESWARA

16 07 08841

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2020**

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS JARINGAN SEMANTIK PADA FORUM TRIPADVISOR UNTUK PENINGKATAN DAYA
TARIK WISATA

yang disusun oleh

RANDYA MAHESWARA

160708841

dinyatakan telah memenuhi syarat pada tanggal 24 September 2020

		Keterangan
Dosen Pembimbing 1	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah menyetujui
Dosen Pembimbing 2	: Ir. Sinta Dewi, M.Sc	Telah menyetujui
Tim Penguji		
Penguji 1	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah menyetujui
Penguji 2	: Dr. Pranowo, S.T., M.T.	Telah menyetujui
Penguji 3	: Prof. Ir. Suyoto, MSc., PhD	Telah menyetujui

Yogyakarta, 24 September 2020
Universitas Atma Jaya Yogyakarta
Fakultas Teknologi Industri
Dekan

ttd

Dr. A. Teguh Siswanto, M.Sc

PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Randy Maheswara
NPM : 16 07 0884
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : Analisis Jaringan Semantik Pada Forum Tripadvisor
Untuk Peningkatan Daya Tarik Wisata

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 28 Agustus 2020

Yang menyatakan,

Randy Maheswara

16 07 08841

HALAMAN PERSEMBAHAN



"Live to learn and you will really learn to live"

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir “Analisis Jaringan Semantik Pada Forum Tripadvisor Untuk Peningkatan Daya Tarik Wisata” ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana Teknik Informatika dari Program Studi Teknik Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Allah SWT yang selalu membimbing dan memberikan rahmat-Nya, dan menyertai penulis selalu.
2. Bapak Dr. A. Teguh Siswanto, M.Sc., selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Martinus Maslim, S.T., M.T., selaku Kepala Program Studi Teknik Informatika, Universitas Atma Jaya Yogyakarta
4. Bapak Prof. Ir. Djoko Budiyanto SHR A, M.Eng., Ph.D, selaku dosen pembimbing I sekaligus dosen pembimbing akademik penulis yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
5. Ibu Ir. Sinta Dewi, M.Sc., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
6. Keluarga penulis Ratna Endah Winarti, Arendya Nariswari, Alm. Suryo Nindiatmo yang telah mendoakan dan memberikan dukungan baik secara moral

maupun materi kepada penulis selama berkuliah sehingga penulis boleh melewatinya dengan baik.

7. Para sahabat dan teman yang telah banyak membantu dan memberikan semangat kepada penulis dalam seluruh proses pengerjaan tugas akhir ini dan proses perkuliahan.
8. Semua pihak yang telah membantu penulis selama proses pengerjaan tugas akhir yang tidak dapat penulis sebutkan satu per satu.

Demikian laporan tugas akhir ini dibuat, dan penulis mengucapkan terima kasih kepada semua pihak. Semoga laporan ini dapat bermanfaat bagi pembaca.

Yogyakarta, 28 Agustus 2020

Randya Maheswara

16 07 08841

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH.....	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
INTISARI.....	xi
BAB I. PENDAHULUAN.....	13
1.1. Latar Belakang	13
1.2. Rumusan Masalah.....	16
1.3. Batasan Masalah.....	16
1.4. Tujuan Penelitian	16
1.5. Metode Penelitian.....	16
BAB II. TINJAUAN PUSTAKA.....	4
BAB III. LANDASAN TEORI.....	6
BAB IV. ANALISIS DAN PERANCANGAN EKSPERIMEN	21
4.1. Deskripsi Problem.....	21
4.2. Analisis Kebutuhan Eksperimen.....	22
4.3. Perancangan Eksperimen	23
BAB V. HASIL EKSPERIMEN DAN PEMBAHASAN	34
5.1. Deskripsi Eksperimen	34

5.2. Hasil Eksperimen	34
5.3. Pembahasan Eksperimen.....	38
BAB VI. PENUTUP	45
6.1. Kesimpulan	45
6.2. Saran.....	45
LAMPIRAN.....	46
DAFTAR PUSTAKA.....	59



DAFTAR GAMBAR

Gambar 3.1 <i>Tree of Porphyry</i> , oleh Peter of Spain (1239)	12
Gambar 3.2 <i>Relational Graph</i>	13
Gambar 3.3 <i>Implication Network</i>	14
Gambar 3.4 <i>Executorial Network</i>	15
Gambar 3.5 <i>Neural Network</i>	16
Gambar 4.1 Hasil Pengambilan Data <i>Link Url</i> dan Judul <i>Post</i>	22
Gambar 4.2 Hasil Pengambilan Data Balasan	22
Gambar 4.3 Gambar 4.3 Pelanggaran Aturan Forum	24
Gambar 4.4 Post Dihapus Oleh Pengunggah	24
Gambar 4.5 Data <i>Co-occurrence</i>	29
Gambar 4.6 Visual Awal Data <i>Co-occurrence</i>	30
Gambar 4.7 Filter <i>Nodes</i> Berdasarkan <i>Degree</i>	31
Gambar 4.8 Hasil Filtering Degree Lombok	32
Gambar 4.9 Hasil Filtering Degree Jakarta.....	32
Gambar 4.10 Hasil Filtering Degree Yogyakarta	33
Gambar 5.1 Clustering Jaringan Semantik Lombok.....	34
Gambar 5.2 Jaringan Semantik Jakarta.....	36
Gambar 5.3 Jaringan Semantik Yogyakarta.....	38

DAFTAR TABEL

Tabel 3.1 Matriks <i>Co-occurrence</i> Nilai Default	8
Tabel 3.2 Matriks <i>Co-occurrence</i> Iterasi Pertama	9
Tabel 3.3 Matriks <i>Co-occurrence</i> Iterasi Kedua	9
Tabel 3.4 Matriks <i>Co-occurrence</i> Iterasi Akhir	9
Tabel 5.1 Topik dan Kata Destinasi Lombok	35
Tabel 5.2 Topik dan Kata Destinasi Jakarta	37
Tabel 5.3 Topik dan Kata Destinasi Yogyakarta	39
Tabel 5.4 Komponen Daya Tarik Wisata Lombok	41
Tabel 5.5 Komponen Daya Tarik Wisata Jakarta	42
Tabel 5.6 Komponen Daya Tarik Wisata Yogyakarta	42



INTISARI

ANALISIS JARINGAN SEMANTIK PADA FORUM TRIPADVISOR UNTUK PENINGKATAN DAYA TARIK WISATA

Intisari

Randy Maheswara

16 07 08841

Negara Indonesia merupakan negara yang terkenal akan kekayaan ragam kebudayaan dan tradisi serta berbagai destinasi wisata tersebar di berbagai kota. Jakarta, Lombok, dan Yogyakarta merupakan destinasi wisata terbesar setelah Bali sebagai penyumbang dalam industri pariwisata. Industri pariwisata merupakan industri yang kompleks dimana didalamnya melibatkan berbagai macam pihak mulai dari pengelola tempat wisata, hotel, restoran, penyedia jasa transportasi, hingga pedagang area destinasi wisata. Melalui internet para wisatawan dapat mencari informasi destinasi wisata seperti tempat menginap termurah, transportasi dan rute jalan terpendek untuk menuju lokasi wisata, informasi seputar atraksi yang ada di lokasi wisata, dan berbagai macam informasi ataupun penawaran menarik lainnya yang ada di lokasi wisata tersebut.

Dengan ketersediaan informasi yang melimpah dari berbagai situs penyedia informasi pariwisata, para wisatawan saat ini dapat mencari informasi dari berbagai sumber salah satunya adalah forum diskusi Tripadvisor. Interaksi antar wisatawan juga menumbuhkan rasa puas ketika dapat membagikan pengalamannya dan dengan membantu orang lain. Dengan demikian, kumpulan ulasan dan diskusi tersebut menjadi hal yang penting untuk melihat tren dan juga komponen daya tarik wisata. Penelitian ini ingin mengungkap bagaimana text mining dari sebuah data yang diambil dari forum diskusi mengandung informasi penting terkait pengembangan daya tarik industri pariwisata. Data tersebut diolah kedalam bentuk matriks *co-occurrence* yang kemudian divisualisasikan kedalam bentuk jaringan semantik.

Visualisasi tersebut selanjutnya diidentifikasi dan diklasifikasikan sesuai dengan teori komponen daya tarik wisata serta akan dilakukan komparasi terhadap jaringan semantik untuk daerah Jakarta, Lombok, dan Yogyakarta. Hasil dari komparasi tersebut diharapkan mengandung informasi yang dapat membantu pemerintah terkait peningkatan daya tarik wisata untuk daerah tersebut.

Kata Kunci: *Text Mining*, *Co-occurrence*, Analisis Jaringan Semantik, Daya Tarik Wisata

Dosen Pembimbing I : Prof. Ir. Djoko Budiyanto SHR A, M.Eng., Ph.D
Dosen Pembimbing II : Ir. Sinta Dewi, M.Sc.
Jadwal Sidang Tugas Akhir : xxx



BAB I. PENDAHULUAN

1.1. Latar Belakang

Negara Indonesia merupakan negara yang terkenal akan kekayaan ragam kebudayaan dan tradisi serta berbagai destinasi wisata tersebar di berbagai kota. Selain kota Bali yang menjadi kota unggulan di Indonesia sebagai penyumbang pemasukan terbesar di industri pariwisata terdapat tempat populer lainnya di Indonesia yaitu Jakarta, Lombok, dan Yogyakarta. merupakan destinasi yang populer bagi kalangan wisatawan mancanegara maupun domestik serta termasuk destinasi dengan penyumbang wisatawan terbanyak bagi Indonesia. Hal tersebut tidak terlepas dari kesan dan pengalaman terhadap daya tarik wisata yang didapat oleh para wisatawan setelah berkunjung ke tempat tersebut [1].

Namun, berdasarkan data dari Badan Pusat Statistik Indonesia, jumlah kunjungan turis asing ke Indonesia sepanjang bulan Januari hingga Oktober 2019 sudah menggapai 13,62 juta. Informasi BPS tersebut menampilkan terdapat peningkatan kunjungan wisatawan asing kurang lebih 2,85 persen dibanding periode yang sama tahun 2018. Badan Pusat Statistik mencatat, pada 2018 lalu, dalam kurun setahun ada 15.810.305 kunjungan turis asing ke Indonesia. Berdasarkan informasi tersebut, terlihat jumlah pengunjung itu jauh di dasar sasaran pemerintah, adalah 20 juta kunjungan. Belum tercapainya target kunjungan tersebut menandakan dibutuhkan evaluasi dari pihak yang terlibat dalam industri pariwisata dan dibutuhkan analisa serta *feedback* untuk melihat komponen daya tarik wisata yang perlu dioptimalkan [1].

Industri pariwisata merupakan industri yang kompleks dimana didalamnya melibatkan berbagai macam pihak mulai dari pengelola tempat wisata, hotel, restoran, penyedia jasa transportasi, hingga pedagang area destinasi wisata. Mereka yang terlibat dalam industri tersebut berusaha memasarkan produk masing-masing dengan kemasan dan berbagai penawaran yang menarik melalui berbagai macam media baik cetak maupun elektronik. Informasi seputar destinasi wisata dan berbagai macam penawaran menarik tersebut merupakan hal yang penting bagi para

calon wisatawan untuk mendapatkan pengalaman terbaik dari destinasi wisata yang dikunjungi. Melalui internet para wisatawan dapat mencari informasi destinasi wisata seperti tempat menginap termurah, transportasi dan rute jalan terpendek untuk menuju lokasi wisata, informasi seputar atraksi yang ada di lokasi wisata, dan berbagai macam informasi ataupun penawaran menarik lainnya yang ada di lokasi wisata tersebut [2].

Situs penyedia informasi seputar pariwisata telah banyak tersebar dan selain menyediakan informasi, situs tersebut juga menjadi penghubung bagi para penyedia jasa seperti penginapan dan transportasi serta restoran untuk menawarkan produk mereka kepada wisatawan. Berbagai macam teknik pemasaran tentu dilakukan agar dapat meraup keuntungan lebih bagi situs penyedia informasi dan penyedia jasa. Dengan ketersediaan informasi yang melimpah dari berbagai situs penyedia informasi pariwisata, para wisatawan saat ini dapat membandingkan harga penyedia jasa antar situs yang lain agar dapat mendapatkan harga yang termurah [3]. Selain hal tersebut para wisatawan saat ini mulai menyadari trik pemasaran yang dilakukan di berbagai situs sehingga lebih mempercayai ulasan dari seorang wisatawan yang memiliki pengalaman memakai jasa tersebut daripada apa yang ditawarkan langsung melalui penyedia jasa. Interaksi antar wisatawan juga menumbuhkan rasa puas ketika dapat membagikan pengalamannya dan dengan membantu orang lain. Dengan demikian, kumpulan ulasan dan diskusi tersebut menjadi hal yang penting untuk melihat tren dan juga komponen daya tarik wisata [3].

Tripadvisor merupakan salah satu dari banyak situs yang menyediakan berbagai macam informasi di bidang pariwisata seperti harga hotel, reservasi restoran, obyek wisata yang dapat dikunjungi di suatu daerah, hingga hal-hal yang perlu diketahui sebelum mengunjungi destinasi wisata. Selain informasi tersebut, saat ini Tripadvisor telah menjadi wadah bagi para wisatawan untuk berbagi pengalaman seputar destinasi wisata yang telah mereka kunjungi. Tripadvisor menyediakan fitur yaitu sebuah forum dimana para wisatawan dapat bertanya, berdiskusi, dan berbagi pengalaman mengenai destinasi wisata yang telah mereka kunjungi. Sebanyak 60 persen pengguna Tripadvisor yang berasal dari Amerika mempercayai situs tersebut ketika akan melakukan rencana berlibur dan pada tahun 2014 Tripadvisor sudah

mencapai 100 juta pengguna yang menandakan bahwa aplikasi tersebut sangat berguna bagi industri pariwisata [4].

Dalam berbagai penelitian mengungkapkan bahwa ulasan dan pengalaman pengguna dalam bentuk teks mengandung informasi penting yang dapat dimanfaatkan untuk dapat mengoptimalkan performa industri pariwisata. Pendekatan analisis *big data* terutama *text mining* memiliki kapabilitas untuk mengambil data dalam jumlah yang besar untuk memecahkan berbagai macam masalah di kehidupan nyata. Pada industri pariwisata, *feedback* yang diberikan oleh para wisatawan terbukti membantu mengembangkan industri tersebut. Dengan hadirnya pendekatan analisis *big data* pada industri pariwisata membuka peluang untuk melihat solusi yang efektif dalam skala lebih besar dan memungkinkan untuk dapat melihat permasalahan lain yang mungkin tidak akan ditemui menggunakan pendekatan konvensional [5].

Interaksi antar pengguna dalam bentuk ulasan dan balasan yang terdapat pada situs layanan informasi pariwisata mengandung potensi informasi mengenai hal berkaitan dengan objek yang didiskusikan. Menggunakan pendekatan *big data* dan melalui *text mining*, informasi yang diperoleh dalam jumlah besar membutuhkan perhitungan yang berkaitan dengan jumlah kemunculan kata beserta relasi antar kata didalamnya. Melalui perhitungan *co-occurrence*, potensi informasi tersebut dapat teridentifikasi dan muncul dalam bentuk matriks beserta nilai relasinya. Jumlah data yang besar dan kolom serta baris yang banyak tidak memungkinkan untuk hasil tersebut dapat dibaca secara jelas dalam bentuk matriks. Dibutuhkan proses representasi data dalam bentuk lain untuk dapat melihat informasi mengenai frekuensi kemunculan dan relasi antar kata yang ada didalam matriks tersebut. Melalui representasi data dalam bentuk jaringan semantik, nilai kemunculan dan keterkaitan antar kedua kata dapat terlihat sehingga dapat ditarik informasi melalui proses visualisasi data tersebut.

Secara garis besar penelitian ini ingin mengungkap bagaimana *text mining* dari sebuah data yang diambil dari forum diskusi mengandung informasi penting terkait pengembangan daya tarik industri pariwisata. Data tersebut akan diolah kedalam bentuk matriks *co-occurrence* yang kemudian akan divisualisasikan kedalam

bentuk jaringan semantik. Visualisasi tersebut selanjutnya akan diidentifikasi dan diklasifikasikan sesuai dengan teori komponen daya tarik wisata serta akan dilakukan komparasi terhadap jaringan semantik untuk Jakarta, Lombok, dan Yogyakarta. Hasil dari komparasi tersebut diharapkan mengandung informasi yang dapat membantu pemerintah terkait peningkatan daya tarik wisata untuk daerah tersebut.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka dapat dirumuskan masalah yaitu bagaimana penerapan *text mining* dalam menemukan, menganalisis, dan membandingkan komponen daya tarik wisata kota berdasarkan visualisasi jaringan semantik data forum diskusi situs Tripadvisor untuk daerah Jakarta, Lombok, dan Yogyakarta?

1.3. Batasan Masalah

Berdasarkan rumusan masalah yang telah dijelaskan diatas, maka terdapat beberapa batasan masalah yang diterapkan, yaitu:

1. Analisis dan perbandingan komponen daya tarik wisata yang dilakukan adalah khusus untuk *post* dan balasan yang terdapat pada forum diskusi situs Tripadvisor untuk daerah Jakarta, Lombok, dan Yogyakarta.
2. Data yang diambil merupakan *post* diskusi dan balasan berbahasa Inggris.
3. Untuk setiap *post* diskusi, diambil 10 balasan teratas dan jika balasan tidak mencapai 10 maka diambil balasan yang tersedia.
4. Data yang diambil data *post* 5 tahun terakhir yaitu tahun 2015 hingga 2020.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah untuk menemukan, menganalisis, dan membandingkan komponen daya tarik wisata berdasarkan visualisasi jaringan semantik data forum diskusi situs Tripadvisor untuk daerah Jakarta, Lombok, dan Yogyakarta.

1.5. Metode Penelitian

Penelitian ini menggunakan data yang diambil dari situs Tripadvisor. Data berupa post dan balasan yang terdapat pada forum diskusi situs Tripadvisor khusus daerah Jakarta, Lombok, dan Yogyakarta.

Berikut adalah metode yang digunakan dalam penelitian ini:

1. Pengumpulan data dari situs Tripadvisor

Data yang diambil yaitu data teks dari tiap *post* diskusi di forum diskusi Tripadvisor khusus daerah Jakarta, Lombok, dan Yogyakarta.

2. *Pre-processing* data

Beberapa proses *pre-processing* atau pembersihan data yang diterapkan untuk menyingkirkan data yang tidak digunakan yaitu penghapusan teks *post* diskusi yang melanggar aturan forum tersebut menggunakan *Microsoft Excel* kemudian dilanjutkan dengan tokenisasi (memisahkan teks menjadi kata per kata), *case folding* (mengubah seluruh kata menjadi *lowercase*), kemudian menghapus angka, tanda baca, spasi berlebih, dan menghilangkan *stopwords*.

3. Pembuatan matriks *co-occurrence*

Data selanjutnya akan dikonversi menjadi sebuah matriks *co-occurrence*. Matriks tersebut berisi kata yang telah di *pre-process* dan angka yang mewakili frekuensi kemunculan antar kedua kata. Matriks ini akan digunakan untuk pembuatan jaringan semantik di proses berikutnya.

4. Pembuatan jaringan semantik

Data akan diproses menggunakan perangkat lunak *gephi* untuk membantu visualisasi data *co-occurrence* yang telah dibuat di proses sebelumnya. Matriks akan dimasukkan kedalam perangkat lunak *gephi* dan diterapkan parameter *degree* untuk menghilangkan *nodes* yang memiliki *degree* rendah.

5. Identifikasi dan klasifikasi *nodes* komponen daya tarik wisata

Data visualisasi berupa *nodes* berlabel akan diidentifikasi dan akan dikelompokkan berdasarkan keempat komponen daya tarik wisata yaitu *attraction*, *accessibility*, *amenities*, dan *anciliary*

6. Analisis dan komparasi jaringan semantik

Nodes visualisasi ketiga jaringan semantik (Jakarta, Lombok, dan Yogyakarta) yang telah dikasifikasikan berdasarkan komponen daya tarik wisata tersebut akan dikomparasi dan dianalisis dengan tujuan mendapatkan informasi komponen yang berpengaruh dari ketiga destinasi wisata tersebut

7. Penarikan kesimpulan berdasarkan hasil eksperimen

Membuat kesimpulan berdasarkan komparasi ketiga jaringan semantik tersebut. Lalu membuat analisis bagaimana pemerintah dapat mengoptimalkan kinerja dari komponen daya tarik wisata tersebut untuk meningkatkan pelayanan dari industri pariwisata di ketiga daerah tersebut.

1.6. Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi konsep dasar dalam pembuatan dan penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini berisi beberapa penelitian terdahulu beserta tabel pembandingnya yang menyangkut penelitian yang dilakukan penulis.

BAB III LANDASAN TEORI

Bab ini berisi penjelasan tentang teori-teori yang digunakan penulis pada penelitian yang dilakukan.

BAB IV ANALISIS DAN PERANCANGAN EKSPERIMEN

Bab ini berisi deskripsi problem, analisis kebutuhan, dan perancangan eksperimen yang dilakukan.

BAB V HASIL EKSPERIMEN DAN PEMBAHASAN

Bab ini berisi deskripsi eksperimen, hasil eksperimen, dan pembahasan eksperimen yang dilakukan.

BAB VI PENUTUP

Bab ini berisi kesimpulan dan saran yang didapatkan dari penelitian yang dilakukan.

BAB II. TINJAUAN PUSTAKA

Layanan pada *hospitality industry* membutuhkan *feedback* dari para pengguna jasa layanannya. Hal ini dibutuhkan agar pihak yang terlibat di industri tersebut dapat mengevaluasi serta meningkatkan jasa pelayanan menjadi lebih nyaman bagi konsumen. Penerapan analisis *big data* dan *text mining* yang dilakukan oleh Zheng dkk (2014) pada *hospitality industry* khususnya pada hotel membuktikan bahwa terbuktinya korelasi yang kuat antara pengalaman pengguna jasa dengan kepuasan yang dituliskan dalam bentuk ulasan dapat membantu meningkatkan kualitas pelayanan bagi pihak hotel. Data tersebut diambil dari 60.648 ulasan dari 10.537 hotel yang terletak di 100 kota besar Amerika Serikat. Data diambil melalui situs *www.expedia.com* dengan periode 18-29 Desember 2007. Hadirnya penerapan analisis *big data* dan *text mining* ini juga ternyata mampu memberikan pandangan baru atau prediksi terhadap permasalahan yang ada di masa mendatang [1].

Ulasan dari pengguna ternyata menyimpan informasi yang dalam mengenai produk atau jasa yang digunakan. Berdasarkan penelitian Andrea Fronzetti dkk (2019) sebuah penelitian dilakukan terhadap 2,6 juta *post* dari forum Tripadvisor. Data tersebut diolah untuk memprediksi permintaan para wisatawan di sektor pariwisata. Dalam pengolahan data tersebut, analisis semantik dilibatkan untuk mengetahui informasi mengenai permintaan di sektor pariwisata. Prediksi dari hasil penelitian tersebut berhasil memprediksi 75% dari total kasus yang diberikan berdasarkan *Google Trends* sebagai pembanding hasil prediksi [2].

Penelitian lainnya juga menggunakan opini yang terdapat pada ulasan situs *online travel* yang ada di China. Sebanyak 165.429 ulasan dikumpulkan untuk melihat topik yang dibicarakan pada ulasan tersebut. Dengan menggunakan visualisasi jaringan semantik, topik dengan frekuensi kemunculan terbesar diidentifikasi. Dari ulasan ketiga situs yaitu Tongcheng, Tuniu, dan Ctrip tersebut ditemukan topik seputar *tour guide*, hotel, pelayanan, pemandangan dan pengalaman [3].

Sebuah sistem pakar berbasis semantik juga diciptakan dengan

menggunakan data dasar pengalaman dari para pengguna hotel. *Sem-Fit* merupakan sistem pakar dengan kombinasi analisis semantik yang bertujuan untuk dapat merekomendasikan hotel sesuai dengan karakteristik yang diinginkan oleh pengguna. Rekomendasi tersebut disesuaikan berdasarkan data ulasan pengalaman pengguna hotel sebelumnya. Data sample yang dilakukan yaitu dengan melakukan simulasi 50 murid yang ada di tahun terakhir mengakses sistem tersebut untuk mendapatkan rekomendasi hotel untuk liburan mereka. Data dilatih menggunakan sepuluh data hotel yang ada di Mallorca, Spanyol. Dengan mengkombinasikan karakteristik hotel, pengalaman pengguna, dan pendapat para ahli mengenai hotel tersebut maka hasil dari rekomendasi yang dihasilkan memiliki tingkat kepuasan lebih baik dibandingkan rekomendasi yang hanya menggunakan pendapat pakar [4].

Prediksi juga telah dilakukan oleh peneliti Andrea de la Hoz-Correa dkk (2017) yang menggunakan text mining dan juga visualisasi dari data tersebut menggunakan analisis co-word. Penelitian tersebut bertujuan untuk mengumpulkan tren tema medical tourism di masa lalu. Setelah dilakukan pemrosesan terhadap data tersebut, visualisasi co-word didapatkan dan menghasilkan enam cluster dalam tren tema medical tourism. Hasilnya dalam keenam cluster tersebut mengindikasikan bahwa praktik mengenai studi “sensitif” menjadi calon tren yang kuat di masa mendatang [5].

Penelitian terakhir dilakukan oleh penulis yaitu untuk mengidentifikasi, menganalisis, dan membandingkan komponen daya tarik wisata yang terdapat pada jaringan semantik. Jaringan semantik tersebut dibentuk berdasarkan data diskusi forum Tripadvisor untuk daerah Lombok, Jakarta, dan Yogyakarta. Penelitian ini menggunakan metode *text mining* dan perhitungan *co-occurrence* dimana data yang dihasilkan berupa matriks yang akan divisualisasikan menggunakan jaringan semantik. Dengan hasil visualisasi tersebut, komponen daya tarik wisata yang teridentifikasi dapat digunakan sebagai bahan pertimbangan keputusan dalam memperbaiki atau meningkatkan daya tarik wisata di daerah tersebut.

BAB VI. PENUTUP

6.1 Kesimpulan

Melalui penelitian mengenai analisis jaringan semantik untuk peningkatan daya tarik wisata ini, dapat disimpulkan bahwa pengunjung yang menggunakan forum diskusi ini cenderung untuk mendiskusikan komponen *amenity* atau fasilitas mengenai transportasi umum seputar harga dan juga alternatifnya. Selain hal tersebut, pengguna forum juga mendiskusikan hal mengenai obyek wisata spesifik yang juga berkaitan dengan harga. Berdasarkan hasil juga disimpulkan bahwa komponen daya tarik wisata *attraction* dan *amenity* menjadi komponen yang paling sering dibahas dan memiliki frekuensi kemunculan yang tinggi dibanding komponen yang lain. Dapat disimpulkan juga bahwa dua komponen lainnya cenderung jarang dibahas dan namun berperan penting sebagai solusi untuk menjawab diskusi wisatawan yang terfokus pada komponen *attraction* dan *amenity*.

Wisatawan memiliki ekspektasi untuk menikmati kenyamanan dan keamanan ketika berlibur sehingga komponen *accessibility* dan *anciliary* menjadi hal yang penting bagi para pengelola destinasi wisata untuk memberikan informasi dan rekomendasi yang tepat untuk wisatawan. Pihak pengelola destinasi wisata dapat memanfaatkan hasil analisis ini untuk dapat memperbaiki atau meningkatkan komponen daya tarik wisata mereka secara tepat sesuai dengan kebutuhan dan ekspektasi wisatawan.

6.2 Saran

Setelah penelitian dilakukan, terdapat beberapa saran yang dapat dilakukan untuk penelitian selanjutnya yang juga meneliti *text mining* khususnya analisis jaringan semantik berdasarkan data *co-occurrence* yaitu sebagai berikut:

1. Analisis jaringan semantik dapat dilakukan pada data forum diskusi atau ulasan obyek atau *platform* lain termasuk media sosial.
2. Pada bagian *pre-processing* dapat dilakukan lebih banyak normalisasi kata untuk berbagai kata yang merupakan singkatan, kesalahan pengetikan dan sebagainya karena dapat lebih membantu ketika proses pengolahan data.

LAMPIRAN

1. Source Code Data Scraping (Pengumpulan Data)

```
#Kota Yogyakarta
from bs4 import BeautifulSoup
import requests
import pandas

url      =      'https://www.tripadvisor.com/ShowForum-g294230-i9252-
                Yogyakarta_Region_Java.html'
front = url[:52]
back = url[51:]
req = requests.get(url)
soup = BeautifulSoup(req.text, 'html.parser')

# Temp Var Tripadvisor
post_link_list = []
post_title_list = []
post_replies_list = []

# Pagination Page
currentPage = 0
# Loop
while currentPage in range(0, 600):
    res = soup.find_all('table', {'class': 'topics'})
    for x in res:
        x_res = x.find_all('b')

        for y in x_res:
            y_res = y.find_all('a')

            for z in y_res:
                z_res = z.get('href')

                link_post = 'https://www.tripadvisor.com'+z_res
                post_link_list.append(link_post)
                print(link_post)
```

```

currentPage += 20

urlMove = front+'o'+str(currentPage)+back
# Request to next page
req = requests.get(urlMove)
soup = BeautifulSoup(req.text, 'html.parser')

jml_link = len(post_link_list)
print("\n" + str(jml_link) + " tripadvisor link crawled")

for each in post_link_list:
    url_post = each
    new_req = requests.get(url_post)
    new_soup = BeautifulSoup(new_req.text, 'html.parser')

    # Get Title
    getTitle = new_soup.find('span', {'class': 'topTitleText'})
    post_title_list.append(getTitle.text)
    print("\n\nTitle : " + getTitle.text)

    # Get 10 Latest Replies
    getReplies = new_soup.find_all('div', {'class': 'postBody'})

    for reply in getReplies:
        rep = reply.find_all('p')

        for reply_add in rep:
            repp = reply_add.text

            if (repp != ""):
                post_replies_list.append(repp)
                print(repp)

createTablePost = {
    'Post Link': post_link_list,
    'Post Title': post_title_list
}

```

```

}

createTableReplies = {
    'Post Replies': post_replies_list
}

dataFramePost = pandas.DataFrame(createTablePost, columns=['Post
    Link', 'Post Title'])
dataFrameReplies = pandas.DataFrame(createTableReplies,
    columns=['Post Replies'])

dataFramePost.to_csv('tripadvisor_yogyakarta_post.csv')
dataFrameReplies.to_csv('tripadvisor_yogyakarta_replies.csv')

print("\nScraping Done! " + str(len(post_link_list)) + " link
    crawled :)")

#Lombok
from bs4 import BeautifulSoup
import requests
import pandas

url = 'https://www.tripadvisor.com/ShowForum-g297733-i9237-
    Lombok_West_Nusa_Tenggara.html'
front = url[:52]
back = url[51:]
req = requests.get(url)
soup = BeautifulSoup(req.text, 'html.parser')

# Temp Var Tripadvisor
post_link_list = []
post_title_list = []
post_replies_list = []

# Pagination Page
currentPage = 0
# Loop
while currentPage in range(0, 3880):
    res = soup.find_all('table', {'class': 'topics'})

```



```

for x in res:
    x_res = x.find_all('b')

    for y in x_res:
        y_res = y.find_all('a')

        for z in y_res:
            z_res = z.get('href')

            link_post = 'https://www.tripadvisor.com'+z_res
            post_link_list.append(link_post)
            print(link_post)

currentPage += 20

urlMove = front+'o'+str(currentPage)+back
# Request to next page
req = requests.get(urlMove)
soup = BeautifulSoup(req.text, 'html.parser')

jml_link = len(post_link_list)
print("\n" + str(jml_link) + " tripadvisor link crawled")

for each in post_link_list:
    url_post = each
    new_req = requests.get(url_post)
    new_soup = BeautifulSoup(new_req.text, 'html.parser')

    # Get Title
    getTitle = new_soup.find('span', {'class': 'topTitleText'})
    post_title_list.append(getTitle.text)
    print("\n\nTitle : " + getTitle.text)

    # Get 10 Latest Replies
    getReplies = new_soup.find_all('div', {'class': 'postBody'})

    for reply in getReplies:
        rep = reply.find_all('p')

```

```

    for reply_add in rep:
        repp = reply_add.text

        if (repp != ""):
            post_replies_list.append(repp)
            print(repp)

createTablePost = {
    'Post Link': post_link_list,
    'Post Title': post_title_list
}

createTableReplies = {
    'Post Replies': post_replies_list
}

dataFramePost = pandas.DataFrame(createTablePost, columns=['Post
Link', 'Post Title'])
dataFrameReplies = pandas.DataFrame(createTableReplies,
columns=['Post Replies'])

dataFramePost.to_csv('tripadvisor_lombok_post.csv')
dataFrameReplies.to_csv('tripadvisor_lombok_replies.csv')

print("\nScraping Done! " + str(len(post_link_list)) + " link
crawled :)")

#Kota Jakarta
from bs4 import BeautifulSoup
import requests
import pandas

url = 'https://www.tripadvisor.com/ShowForum-g294229-i9444-
Jakarta_Java.html'
front = url[:52]

```

```

back = url[51:]
req = requests.get(url)
soup = BeautifulSoup(req.text, 'html.parser')

# Temp Var Tripadvisor
post_link_list = []
post_title_list = []
post_replies_list = []

# Pagination Page
currentPage = 0
# Loop
while currentPage in range(0, 940):
    res = soup.find_all('table', {'class': 'topics'})
    for x in res:
        x_res = x.find_all('b')

        for y in x_res:
            y_res = y.find_all('a')

            for z in y_res:
                z_res = z.get('href')

                link_post = 'https://www.tripadvisor.com'+z_res
                post_link_list.append(link_post)
                print(link_post)

    currentPage += 20

    urlMove = front+'o'+str(currentPage)+back
    # Request to next page
    req = requests.get(urlMove)
    soup = BeautifulSoup(req.text, 'html.parser')

jml_link = len(post_link_list)
print("\n" + str(jml_link) + " tripadvisor link crawled")

for each in post_link_list:

```

```

url_post = each
new_req = requests.get(url_post)
new_soup = BeautifulSoup(new_req.text, 'html.parser')

# Get Title
getTitle = new_soup.find('span', {'class': 'topTitleText'})
post_title_list.append(getTitle.text)
print("\n\nTitle : " + getTitle.text)

# Get 10 Latest Replies
getReplies = new_soup.find_all('div', {'class': 'postBody'})

for reply in getReplies:
    rep = reply.find_all('p')

    for reply_add in rep:
        repp = reply_add.text

        if (repp != ""):
            post_replies_list.append(repp)
            print(repp)

createTablePost = {
    'Post Link': post_link_list,
    'Post Title': post_title_list
}

createTableReplies = {
    'Post Replies': post_replies_list
}

dataFramePost = pandas.DataFrame(createTablePost, columns=['Post
    Link', 'Post Title'])
dataFrameReplies = pandas.DataFrame(createTableReplies,
    columns=['Post Replies'])

dataFramePost.to_csv('tripadvisor_jakarta_post.csv')

```

```
dataFrameReplies.to_csv('tripadvisor_jakarta_replies.csv')

print("\nScraping Done! " + str(len(post_link_list)) + " link
      crawled :)")
```

2. Source Code Data Pre-Processing dan Co-occurrence

```
import csv
import re
import string
import nltk
# from nltk.stem import PorterStemmer

corpus = []
with open('tripadvisor_lombok_replies.csv', 'r', errors='ignore')
as csv_file:
    csv_reader = csv.reader(csv_file)
    next(csv_reader)
    for line in csv_reader:
        corpus.append(line[1])

# initialize
clean_text = []
# ps = PorterStemmer()

for row in corpus:
    # tokenize
    tokens = nltk.tokenize.word_tokenize(row)

    # lowercase
    tokens = [token.lower() for token in tokens]

    # hapus angka
    tokens = [re.sub(r"\d+", "", token) for token in tokens]

    # hapus tanda baca
```

```

tokens      =      [token.translate(str.maketrans("", "",
string.punctuation)) for token in tokens]

# hapus whitespace
tokens = [token.strip() for token in tokens]

# # stemming
# tokens = [ps.stem(token) for token in tokens]

# isword
tokens = [token for token in tokens if token.isalpha()]
clean_sentence = ''
clean_sentence = ' '.join(token for token in tokens)
clean_text.append(clean_sentence)

all_text = clean_text

my_stopwords=frozenset(["mon","articl","0o", "0s", "3a", "3b",
"3d", "6b", "6o", "a", "a1", "a2", "a3", "a4", "ab", "able",
"about", "above", "abt", "ac", "accordance", "according",
"accordingly", "across", "act", "actually", "ad", "added", "adj",
"ae", "af", "affected", "affecting", "affects", "after",
"afterwards", "ag", "again", "against", "ah", "ain", "ain't", "aj",
"al", "all", "allow", "allows", "almost", "alone", "along",
"already", "also", "although", "always", "am", "among", "amongst",
"amoungst", "amount", "an", "and", "announce", "another", "any",
"anybody", "anyhow", "anymore", "anyone", "anything", "anyway",
"anyways", "anywhere", "ao", "ap", "apart", "apparently", "appear",
"appreciate", "appropriate", "approximately", "ar", "are", "aren",
"arent", "aren't", "arise", "around", "as", "a's", "aside", "ask",
"asking", "associated", "at", "au", "auth", "av", "available",
"aw", "away", "awfully", "ax", "ay", "az", "b", "b1", "b2", "b3",
"ba", "back", "bc", "bd", "be", "became", "because", "become",
"becomes", "becoming", "been", "before", "beforehand", "begin",
"beginning", "beginnings", "begins", "behind", "being", "believe",
"below", "beside", "besides", "best", "better", "between",
"beyond", "bi", "bill", "biol", "bj", "bk", "bl", "bn", "both",
"bottom", "bp", "br", "brief", "briefly", "bs", "bt", "bu", "but",

```

"bx", "by", "c", "c1", "c2", "c3", "ca", "call", "came", "can",
"cannot", "cant", "can't", "cause", "causes", "cc", "cd", "ce",
"certain", "certainly", "cf", "cg", "ch", "changes", "ci", "cit",
"cj", "cl", "clearly", "cm", "c'mon", "cn", "co", "com", "come",
"comes", "con", "concerning", "consequently", "consider",
"considering", "contain", "containing", "contains",
"corresponding", "could", "couldn", "couldnt", "couldn't",
"course", "cp", "cq", "cr", "cry", "cs", "c's", "ct", "cu",
"currently", "cv", "cx", "cy", "cz", "d", "d2", "da", "date", "dc",
"dd", "de", "definitely", "describe", "described", "despite",
"detail", "df", "di", "did", "didn", "didn't", "different", "dj",
"dk", "dl", "do", "does", "doesn", "doesn't", "doing", "don",
"done", "don't", "down", "downwards", "dp", "dr", "ds", "dt", "du",
"due", "during", "dx", "dy", "e", "e2", "e3", "ea", "each", "ec",
"ed", "edu", "ee", "ef", "effect", "eg", "ei", "eight", "eighty",
"either", "ej", "el", "eleven", "else", "elsewhere", "em", "empty",
"en", "end", "ending", "enough", "entirely", "eo", "ep", "eq", "er",
"es", "especially", "est", "et", "et-al", "etc", "eu", "ev", "even",
"ever", "every", "everybody", "everyone", "everything",
"everywhere", "ex", "exactly", "example", "except", "ey", "f",
"f2", "fa", "far", "fc", "few", "ff", "fi", "fifteen", "fifth",
"fifty", "fill", "find", "fire", "first", "five", "fix", "fj", "fl",
"fn", "fo", "followed", "following", "follows", "for", "former",
"formerly", "forth", "forty", "found", "four", "fr", "from",
"front", "fs", "ft", "fu", "full", "further", "furthermore", "fy",
"g", "ga", "gave", "ge", "get", "gets", "getting", "gi", "give",
"given", "gives", "giving", "gj", "gl", "go", "goes", "going",
"gone", "got", "gotten", "gr", "greetings", "gs", "gy", "h", "h2",
"h3", "had", "hadn", "hadn't", "happens", "hardly", "has", "hasn",
"hasnt", "hasn't", "have", "haven", "haven't", "having", "he",
"hed", "he'd", "he'll", "hello", "help", "hence", "her", "here",
"hereafter", "hereby", "herein", "heres", "here's", "hereupon",
"hers", "herself", "hes", "he's", "hh", "hi", "hid", "him",
"himself", "his", "hither", "hj", "ho", "home", "hopefully", "how",
"howbeit", "however", "how's", "hr", "hs", "http", "hu", "hundred",
"hy", "i", "i2", "i3", "i4", "i6", "i7", "i8", "ia", "ib", "ibid",
"ic", "id", "i'd", "ie", "if", "ig", "ignored", "ih", "ii", "ij",
"il", "i'll", "im", "i'm", "immediate", "immediately",

"importance", "important", "in", "inasmuch", "inc", "indeed",
"index", "indicate", "indicated", "indicates", "information",
"inner", "insofar", "instead", "interest", "into", "invention",
"inward", "io", "ip", "iq", "ir", "is", "isn", "isn't", "it", "itd",
"it'd", "it'll", "its", "it's", "itself", "iv", "i've", "ix", "iy",
"iz", "j", "jj", "jr", "js", "jt", "ju", "just", "k", "ke", "keep",
"keeps", "kept", "kg", "kj", "km", "know", "known", "knows", "ko",
"l", "l2", "la", "largely", "last", "lately", "later", "latter",
"latterly", "lb", "lc", "le", "least", "les", "less", "lest", "let",
"lets", "let's", "lf", "like", "liked", "likely", "line", "little",
"lj", "ll", "ll", "ln", "lo", "look", "looking", "looks", "los",
"lr", "ls", "lt", "ltd", "m", "m2", "ma", "made", "mainly", "make",
"makes", "many", "may", "maybe", "me", "mean", "means", "meantime",
"meanwhile", "merely", "mg", "might", "mightn", "mightn't", "mill",
"million", "mine", "miss", "ml", "mn", "mo", "more", "moreover",
"most", "mostly", "move", "mr", "mrs", "ms", "mt", "mu", "much",
"mug", "must", "mustn", "mustn't", "my", "myself", "n", "n2", "na",
"name", "namely", "nay", "nc", "nd", "ne", "near", "nearly",
"necessarily", "necessary", "need", "needn", "needn't", "needs",
"neither", "never", "nevertheless", "new", "next", "ng", "ni",
"nine", "ninety", "nj", "nl", "nn", "no", "nobody", "non", "none",
"nonetheless", "noone", "nor", "normally", "nos", "not", "noted",
"nothing", "novel", "now", "nowhere", "nr", "ns", "nt", "ny", "o",
"oa", "ob", "obtain", "obtained", "obviously", "oc", "od", "of",
"off", "often", "og", "oh", "oi", "oj", "ok", "okay", "ol", "old",
"om", "omitted", "on", "once", "one", "ones", "only", "onto", "oo",
"op", "oq", "or", "ord", "os", "ot", "other", "others", "otherwise",
"ou", "ought", "our", "ours", "ourselves", "out", "outside",
"over", "overall", "ow", "owing", "own", "ox", "oz", "p", "p1",
"p2", "p3", "page", "pagecount", "pages", "par", "part",
"particular", "particularly", "pas", "past", "pc", "pd", "pe",
"per", "perhaps", "pf", "ph", "pi", "pj", "pk", "pl", "placed",
"please", "plus", "pm", "pn", "po", "poorly", "possible",
"possibly", "potentially", "pp", "pq", "pr", "predominantly",
"present", "presumably", "previously", "primarily", "probably",
"promptly", "proud", "provides", "ps", "pt", "pu", "put", "py",
"q", "qj", "qu", "que", "quickly", "quite", "qv", "r", "r2", "ra",
"ran", "rather", "rc", "rd", "re", "readily", "really",

"reasonably", "recent", "recently", "ref", "refs", "regarding",
"regardless", "regards", "related", "relatively", "research",
"research-articl", "respectively", "resulted", "resulting",
"results", "rf", "rh", "ri", "right", "rj", "rl", "rm", "rn", "ro",
"rq", "rr", "rs", "rt", "ru", "run", "rv", "ry", "s", "s2", "sa",
"said", "same", "saw", "say", "saying", "says", "sc", "sd", "se",
"sec", "second", "secondly", "section", "see", "seeing", "seem",
"seemed", "seeming", "seems", "seen", "self", "selves", "sensible",
"sent", "serious", "seriously", "seven", "several", "sf", "shall",
"shan", "shan't", "she", "shed", "she'd", "she'll", "shes",
"she's", "should", "shouldn", "shouldn't", "should've", "show",
"showed", "shown", "shows", "shows", "si", "side", "significant",
"significantly", "similar", "similarly", "since", "sincere", "six",
"sixty", "sj", "sl", "slightly", "sm", "sn", "so", "some",
"somebody", "somehow", "someone", "somethan", "something",
"sometime", "sometimes", "somewhat", "somewhere", "soon", "sorry",
"sp", "specifically", "specified", "specify", "specifying", "sq",
"sr", "ss", "st", "still", "stop", "strongly", "sub",
"substantially", "successfully", "such", "sufficiently", "suggest",
"sup", "sure", "sy", "system", "sz", "t", "t1", "t2", "t3", "take",
"taken", "taking", "tb", "tc", "td", "te", "tell", "ten", "tends",
"tf", "th", "than", "thank", "thanks", "thanx", "that", "that'll",
"thats", "that's", "that've", "the", "their", "theirs", "them",
"themselves", "then", "thence", "there", "thereafter", "thereby",
"thered", "therefore", "therein", "there'll", "thereof", "therere",
"theres", "there's", "thereto", "thereupon", "there've", "these",
"they", "theyd", "they'd", "they'll", "theyre", "they're",
"they've", "thickv", "thin", "think", "third", "this", "thorough",
"thoroughly", "those", "thou", "though", "thoughh", "thousand",
"three", "throug", "through", "throughout", "thru", "thus", "ti",
"til", "tip", "tj", "tl", "tm", "tn", "to", "together", "too",
"took", "top", "toward", "towards", "tp", "tq", "tr", "tried",
"tries", "truly", "try", "trying", "ts", "t's", "tt", "tv",
"twelve", "twenty", "twice", "two", "tx", "u", "u201d", "ue", "ui",
"uj", "uk", "um", "un", "under", "unfortunately", "unless",
"unlike", "unlikely", "until", "unto", "uo", "up", "upon", "ups",
"ur", "us", "use", "used", "useful", "usefully", "usefulness",
"uses", "using", "usually", "ut", "v", "va", "value", "various",

```

"vd", "ve", "ve", "very", "via", "viz", "vj", "vo", "vol", "vols",
"volumtype", "vq", "vs", "vt", "vu", "w", "wa", "want", "wants",
"was", "wasn", "wasnt", "wasn't", "way", "we", "wed", "we'd",
"welcome", "well", "we'll", "well-b", "went", "were", "we're",
"weren", "werent", "weren't", "we've", "what", "whatever",
"what'll", "whats", "what's", "when", "whence", "whenever",
"when's", "where", "whereafter", "whereas", "whereby", "wherein",
"wheres", "where's", "whereupon", "wherever", "whether", "which",
"while", "whim", "whither", "who", "whod", "whoever", "whole",
"who'll", "whom", "whomever", "whos", "who's", "whose", "why",
"why's", "wi", "widely", "will", "willing", "wish", "with",
"within", "without", "wo", "won", "wonder", "wont", "won't",
"words", "world", "would", "wouldn", "wouldnt", "wouldn't", "www",
"x", "x1", "x2", "x3", "xf", "xi", "xj", "xk", "xl", "xn", "xo",
"xs", "xt", "xv", "xx", "y", "y2", "yes", "yet", "yj", "yl", "you",
"you'd", "you'll", "your", "youre", "you're", "yours",
"yourself", "yourselves", "you've", "yr", "ys", "yt", "z", "zero",
"zi", "zz"]
# sklearn countvectorizer
from sklearn.feature_extraction.text import CountVectorizer
# Convert a collection of text documents to a matrix of token counts
cv = CountVectorizer(ngram_range=(1,1), stop_words = my_stopwords)
# matrix of token counts
X = cv.fit_transform(all_text)
Xc = (X.T * X) # matrix manipulation
Xc.setdiag(0) # set the diagonals to be zeroes as it's pointless to
be 1

import pandas as pd
names = cv.get_feature_names() # This are the entity names (i.e.
keywords)
df = pd.DataFrame(data = Xc.toarray(), columns = names, index =
names)
df.to_csv('word_coocurrence_matrix_test_lombok.csv', sep = ',')

```

DAFTAR PUSTAKA

- [1] Z. Xiang, Z. Schwartz, J. H. Gerdes, and M. Uysal, “What can big data and text analytics tell us about hotel guest experience and satisfaction?,” *Int. J. Hosp. Manag.*, vol. 44, pp. 120–130, 2015.
- [2] A. Fronzetti Colladon, B. Guardabascio, and R. Innarella, “Using social network and semantic analysis to analyze online travel forums and forecast tourism demand,” *Decis. Support Syst.*, vol. 123, no. January, p. 113075, 2019.
- [3] Z. Hou, F. Cui, Y. Meng, T. Lian, and C. Yu, “Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis,” *Tour. Manag.*, vol. 74, no. January, pp. 276–289, 2019.
- [4] Á. García-Crespo, J. L. López-Cuadrado, R. Colomo-Palacios, I. González-Carrasco, and B. Ruiz-Mezcua, “Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain,” *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13310–13319, 2011.
- [5] A. de la Hoz-Correa, F. Muñoz-Leiva, and M. Bakucz, “Past themes and future trends in medical tourism research: A co-word analysis,” *Tour. Manag.*, vol. 65, pp. 200–211, 2018.
- [6] A.-H. Tan, “Text Mining: The state of the art and the challenges,” *Proc. PAKDD 1999 Work. Knowl. Discovery from Adv. Databases*, vol. 8, pp. 65–70, 1999.
- [7] M. J. Kim, K. Ohk, and C. S. Moon, “Trend analysis by using text mining of journal articles regarding consumer policy,” *New Phys. Sae Mulli*, vol. 67, no. 5, pp. 555–561, 2017.
- [8] C. Malegori, L. Franzetti, R. Guidetti, E. Casiraghi, and R. Rossi, “GLCM, an image analysis technique for early detection of biofilm,” *J. Food Eng.*, vol. 185, pp. 48–55, 2016.
- [9] Priyanka and D. Kumar, “Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA,” *Procedia Comput. Sci.*, vol.

- 167, no. 2019, pp. 1722–1731, 2020.
- [10] L. Budiman, “Analisis Bibliometrika Berdasarkan Pendekatan Co-Words: Pemetaan Laporan Hasil Penelitian Pusat Penelitian Biologi-LIPI,” 2012.
- [11] M. H. Davarpour, M. K. Sohrabi, and M. Naderi, “Toward a semantic-based location tagging news feed system: Constructing a conceptual hierarchy on geographical hashtags,” *Comput. Electr. Eng.*, vol. 78, pp. 204–217, 2019.
- [12] J. F. Sowa, “Semantic Networks,” in *Encyclopedia of Artificial Intelligence*, no. 1972, 2018, pp. 1–24.
- [13] UUDRI, “Undang-Undang Dasar Republik Indonesia Nomor 10 Tahun 2009,” *Undang. Dasar Republik Indones.*, 2009.
- [14] Setiawan, “Identifikasi Potensi Wisata Beserta 4a (Attraction, Amenity, Accessibility, Ancilliary) Di Dusun Sumber Wangi, Desa Pemuteran, Kecamatan Gerokgak, Kabupaten Buleleng, Bali Nama,” pp. 1–21, 2015.
- [15] G. Prayag, S. Hosany, and K. Odeh, “The role of tourists’ emotional experiences and satisfaction in understanding behavioral intentions,” *J. Destin. Mark. Manag.*, vol. 2, no. 2, pp. 118–127, 2013.
- [16] M. Nilashi, O. Ibrahim, E. Yadegaridehkordi, S. Samad, E. Akbari, and A. Alizadeh, “Travelers decision making using online review in social network sites: A case on TripAdvisor,” *J. Comput. Sci.*, vol. 28, pp. 168–179, 2018.