

**PERANCANGAN METODE *TEXT MINING* UNTUK
MENEMUKAN *TRAIT* TERKAIT *BRAND PERSONALITY*
HOTEL BERDASARKAN *ONLINE REVIEW* KONSUMEN**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat guna mencapai gelar
Sarjana dalam bidang ilmu Teknik Industri

Disusun oleh :

Nama : Ilmor Sonata Mahardika

NPM : 6131801145



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG
2023**

**FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG**



Nama : Ilmor Sonata Mahardika
NPM : 6131801145
Jurusan : Teknik Industri
Judul Skripsi : PERANCANGAN METODE *TEXT MINING* UNTUK
MENEMUKAN *TRAIT* TERKAIT *BRAND PERSONALITY*
HOTEL BERDASARKAN *ONLINE REVIEW* KONSUMEN

TANDA PERSETUJUAN SKRIPSI

Bandung, Januari 2023

Ketua Program Studi Sarjana Teknik Industri

(Dr. Ceicalia Tesavrita)

Pembimbing

(Dedy Suryadi, S.T., M.S., Ph.D.)



Program Studi Sarjana Teknik Industri
Jurusan Teknik Industri
Fakultas Teknologi Industri
Universitas Katolik Parahyangan

Pernyataan Tidak Mencontek atau Melakukan Tindakan Plagiat

Saya, yang bertanda tangan di bawah ini,

Nama : Ilmor Sonata Mahardika

NPM : 6131801145

dengan ini menyatakan bahwa skripsi dengan judul :

**“PERANCANGAN METODE *TEXT MINING* UNTUK MENEMUKAN *TRAIT*
TERKAIT *BRAND PERSONALITY* HOTEL BERDASARKAN *ONLINE REVIEW*
KONSUMEN”**

adalah hasil pekerjaan saya dan seluruh ide, pendapat atau materi dari sumber lain telah dikutip dengan cara penulisan referensi yang sesuai.

Pernyataan ini saya buat dengan sebenar-benarnya dan jika pernyataan ini tidak sesuai dengan kenyataan, maka saya bersedia menanggung sanksi yang akan dikenakan kepada saya.

Bandung,

Ilmor Sonata Mahardika
6131801145

ABSTRAK

Era digital saat ini telah mendorong industri pariwisata untuk mengubah strategi organisasi dan *brand personality*. Tripadvisor muncul di web 2.0 memudahkan para wisatawan untuk mendapatkan informasi tentang penyedia pariwisata. Pertumbuhan industri pariwisata dan perhotelan menjadi sangat cepat di dunia dan menunjukkan alasan yang signifikan untuk mengevaluasi *sustainability* sebagai tema penting. Selain itu, *branding* menjadi tema yang penting karena merupakan salah satu aset yang paling berharga yang dimiliki oleh perusahaan. Konsumen saat ini juga semakin terbiasa menggunakan teknologi dalam proses pengambilan keputusan mereka. Oleh karena itu, bisnis perhotelan perlu mengetahui *insight* dari *review* pengguna, sehingga hotel dapat melakukan evaluasi terhadap *brand personality* yang mereka miliki. Penelitian ini memberikan usulan perbaikan metode *text mining* dari penelitian sebelumnya karya (Neto, Silva, Ferreira, & Araújo, 2020). Tujuannya agar *insight* yang diperoleh hotel menjadi lebih akurat.

Penelitian ini dilakukan dengan mencari kata pada tiap *trait* dan dimensi *brand personality* hotel berdasarkan Aaker. Metode yang digunakan adalah *word embedding* untuk mendapatkan kata yang memiliki kemiripan dengan *trait* berdasarkan konteks. Informasi dari kata pada tiap *trait* ini dilanjutkan dengan penelitian menggunakan metode *dependency parsing*. *Dependency parsing* ini menganalisa struktur kalimat berdasarkan konteks. *Insight* dari *dependency parsing* ini berguna untuk hotel dan dapat dikembangkan dengan metode *sentiment analysis* untuk mendapatkan *insight* yang lebih mendalam.

Dari hasil pengolahan suatu hotel, ditemukan 1968 *review* hotel dengan 5 *review* yang kurang tepat. Lalu, dengan metode yang sama, dilakukan pengolahan pada hotel lain dan menghasilkan 1335 *review* dengan 3 *review* yang kurang tepat. Selisih perbedaan sebelum dan sesudah penggunaan metode *dependency parsing* pada dimensi *sincerity* sebesar 4317, dimensi *excitement* sebesar 3792, dimensi *competence* sebesar 2566, dimensi *sophistication* sebesar 1017, dimensi *ruggedness* sebesar 1046, dan dimensi *sustainability* sebesar 5443.

Kata kunci : *Text Mining, Brand Personality, Hotel, Dependency Parsing, NLP, Python.*

ABSTRACT

The digital age has pushed the tourism industry to transform their organization strategies and brand personalities. Tripadvisor emerged on web 2.0, making it easier for travelers to obtain information about tourism providers. The growth of the tourism and hospitality industry has become very fast worldwide and shows significant reasons for assessing sustainability as an important theme. In addition, branding has become an important theme because it is one of the most valuable assets that companies possess. Consumers today are also increasingly using technology in their decision-making processes. Therefore, hotel businesses need to understand the insights from user reviews so that hotels can evaluate their own brand personalities. This study proposes an improvement of the text mining method from the previous study by (Neto, Silva, Ferreira, & Araújo, 2020), in order to make the insights obtained by hotels more accurate.

This study was conducted by searching for words in each trait and dimension of hotel brand personality based on Aaker. The method used is word embedding to obtain words that are similar to traits based on context. Information from the word on each trait is followed by research using the dependency parsing method. This dependency parsing analyzes sentence structure based on context. Insights from this dependency parsing are useful for hotels and can be developed with a sentiment analysis method to obtain deeper insights.

From the processing of one hotel, 1968 reviews of the hotel were found with 5 reviews that were not appropriate. Then, using the same method, processing was carried out on another hotel and resulted in 1335 reviews with 3 reviews that were not appropriate. The difference before and after using dependency parsing method on the dimension of sincerity is 4317, the dimension of excitement is 3792, the dimension of competence is 2566, the dimension of sophistication is 1017, the dimension of ruggedness is 1046, and the dimension of sustainability is 5443.

Keyword : *Text Mining, Brand Personality, Hotel, Dependency Parsing, NLP, Python.*

KATA PENGANTAR

Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa, Pada kesempatan ini, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan dalam penyelesaian skripsi ini. Penulis menyadari bahwa tidak mungkin penyelesaian skripsi ini terlaksana dengan baik tanpa bantuan dari berbagai pihak. Oleh karena itu, penulis berterimakasih sebesar besarnya kepada :

1. Bapak Dedy Suryadi, S.T., M.S., Ph.D. selaku dosen pembimbing skripsi yang telah memberikan bimbingan selama proses penyusunan skripsi.
2. Bapak Hanky Fransiscus, S.T., M.T., dan Bapak Romy Loice, S.T., M.T. selaku dosen penguji yang telah memberikan masukan penelitian agar lebih baik.
3. Ibu Dr. Ceicalia Tesavrita, S.T., M.T. selaku koordinator skripsi.
4. Keluarga yang telah memberikan dukungan dalam proses penelitian.
5. Teman teman teknik industri 2018 kelas C.
6. Leonard, Allif K, Yohannes Mario, Daris Azhar yang telah memberikan inspirasi bagi peneliti.
7. youtuber machine learning atau NLP, dan peneliti NLP yang telah membantu dalam proses pemahaman *coding* dan pemahaman materi penelitian.
8. Pihak lainnya yang telah membantu dalam proses penelitian.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca demi kesempurnaan skripsi ini.

Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca dan memberikan sumbangsih terhadap pengembangan ilmu pengetahuan di bidang *text mining*.

Bandung, 4 Januari 2023

Penulis

DAFTAR ISI

ABSTRAK.....	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN.....	I-1
I.1 Latar Belakang Masalah	I-1
I.2 Identifikasi dan Rumusan Masalah	I-2
I.3 Pembatasan Masalah dan Asumsi.....	I-7
I.4 Tujuan Penelitian.....	I-8
I.5 Manfaat Penelitian.....	I-8
I.6 Metodologi Penelitian	I-8
I.7 Sistematika Penulisan	I-12
BAB II TINJAUAN PUSTAKA.....	II-1
II.1 <i>Dependency Parsing & Dependency Tree</i>	II-1
II.2 <i>Brand Personality</i>	II-2
II.3 <i>Word Embedding</i>	II-4
II.4 <i>Word2vec</i>	II-7
BAB III PENGUMPULAN DAN PENGOLAHAN DATA.....	III-1
III.1 Data Penelitian	III-1
III.2 Pencarian <i>Trait</i>	III-1
III.3 <i>Dependency Parsing & Dependency Tree</i>	III-12
III.4 Performansi Pada RIU Hotel.....	III-21
III.5 Usulan Perbaikan Pada Hotel Lotte New York Palace	III-23
BAB IV ANALISIS.....	IV-1
IV.1 Analisis Pengumpulan Data.....	IV-1
IV.2 Analisis Pencarian <i>Trait</i> (<i>Word2vec</i>).....	IV-2
IV.3 Analisis Metode <i>Dependency Parsing</i>	IV-3

IV.4	Analisis Pengukuran Performansi	IV-4
IV.5	Analisis Usulan Perbaikan Brand Personality Hotel	IV-6
BAB V KESIMPULAN DAN SARAN.....		V-1
V.1	Kesimpulan.....	V-1
V.2	Saran.....	V-2
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR TABEL

Tabel III.1 Sebelum dan Sesudah Penggunaan Dependency Parsing	III-20
Tabel IV.1 Contoh Kalimat yang Tidak Memiliki Konteks Hotel	IV-4

DAFTAR GAMBAR

Gambar I.1 Gambar Hasil <i>Text Mining</i>	I-4
Gambar I.2 Radar <i>Chart</i>	I-5
Gambar I.3 Diagram Alir Metodologi Penelitian	I-9
Gambar II.1 Dependency Parsing & Dependency Tree.....	II-1
Gambar II.2 <i>Grammar Dependency Parsing & Dependency Tree</i>	II-2
Gambar II.3 Deskripsi Dimensi <i>Brand Personality</i>	II-3
Gambar II.4 <i>Cosine Similarity</i>	II-5
Gambar II.5 <i>Word's Move Distance</i>	II-6
Gambar II.6 <i>Euclidean Distance</i>	II-6
Gambar II.7 Arsitektur CBOW	II-8
Gambar II.8 Arsitektur <i>Skip Gram</i>	II-9
Gambar III.1 <i>Flow</i> Pencarian <i>Trait</i>	III-2
Gambar III.2 Tahapan Persiapan dan <i>Preprocessing Data</i>	III-4
Gambar III.3 Tahapan Persiapan dan <i>Preprocessing Data</i>	III-4
Gambar III.4 <i>Training Data Word2vec</i>	III-5
Gambar III.5 <i>Flow Most Similar Words Hotel</i>	III-6
Gambar III.6 <i>Coding For Most Similar Words Hotel</i>	III-7
Gambar III.7 Contoh <i>Coding</i> Menghitung Frekuensi Kata	III-8
Gambar III.8 Jumlah Kata <i>Trait Honest</i>	III-9
Gambar III.9 Frekuensi Kata pada Tiap Dimensi.....	III-10
Gambar III.10 <i>Interval Plot</i> Faktor Unsur Dinamis terhadap Skor <i>Recall</i>	III-11
Gambar III.11 <i>Radar Chart</i> Pada Tiap Dimensi.....	III-11
Gambar III.12 Contoh Hubungan <i>Parent Child Dependency Parsing</i>	III-12
Gambar III.13 <i>Libraries Dependency Parsing</i>	III-13
Gambar III.14 Pencarian Kata Hasil <i>Dependency Parsing</i>	III-14
Gambar III.15 Kalimat Hasil <i>Dependency Parsing</i>	III-16
Gambar III.16 Kata Hasil <i>Dependency Parsing Sibling</i>	III-17
Gambar III.17 Kalimat Hasil <i>Dependency Parsing Sibling</i>	III-17
Gambar III.18 Frekuensi Hasil <i>Dependency Parsing Parent Child</i>	III-19

Gambar III.19 Kata Sinonim Hasil <i>Trait Real</i>	III-20
Gambar III.20 Contoh <i>Code Dependency Parsing RIU</i>	III-20
Gambar III.21 Contoh Hasil <i>Dependency Parsing RIU</i>	III-21
Gambar III.22 <i>Radar Chart</i> Hasil <i>Dependency Parsing RIU</i>	III-22
Gambar III.23 <i>Libraries Sentiment Analysis</i>	III-23
Gambar III.24 Contoh <i>Sentiment Analysis</i> Dimensi <i>Sincerity</i>	III-26
Gambar III.25 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Sincerity</i>	III-27
Gambar III.26 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Excitement</i>	III-28
Gambar III.27 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Competence</i>	III-29
Gambar III.28 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Sophistication</i>	III-30
Gambar III.29 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Ruggedness</i>	III-31
Gambar III.30 Frekuensi <i>Sentiment Analysis</i> Dimensi <i>Sustainability</i>	III-32

DAFTAR LAMPIRAN

Lampiran A Pencarian <i>Trait</i> (<i>Word Embedding</i>)	A-1
Lampiran B <i>Dependency Parsing</i> Hotel Lotte New York Palace.....	B-1
Lampiran C <i>Dependency Parsing</i> Hotel RIU Plaza.....	C-1
Lampiran D <i>Sentiment Analysis</i> Hotel Lotte New York Palace.....	D-1

BAB I

PENDAHULUAN

Pada bab ini akan dijelaskan mengenai pendahuluan skripsi. Bab ini akan menjelaskan mengenai latar belakang, identifikasi dan rumusan masalah, pembatasan dan asumsi penelitian, tujuan dan manfaat penelitian, serta metodologi penelitian. Berikut penjelasannya.

I.1 Latar Belakang Masalah

Brand personality (BP) merupakan hal yang penting agar konsumen membeli produk yang ditawarkan. *Brand personality* berfungsi untuk diferensiasi dengan produk pesaing. *Brand personality* yang baik dapat memiliki ciri khas tersendiri bagi perusahaan. Selain itu, dapat memberikan pengalaman dan memori terbaik bagi pelanggan terhadap produk yang ditawarkan. Sehingga, dapat mempertahankan nilai perusahaan dalam waktu yang lama dan dapat memiliki nilai lebih di mata konsumen.

Menurut Tarver (2021), *Brand personality* adalah sebuah *framework* yang berguna untuk perusahaan dalam meningkatkan *value* produk, jasa, misi kepada masyarakat. Masyarakat akan menggunakan produk atau jasa ketika mereka merasa sesuai dengan karakteristik produk atau jasa yang ditawarkan. *Brand personality* sangat penting di dunia digital seperti sekarang. Dengan adanya teknologi otomatisasi mempermudah masyarakat dalam belanja *online*. Selain itu, perusahaan dapat memprediksi preferensi masyarakat ketika mereka belanja *online*.

Pentingnya *brand personality* ini berlaku pada *brand* hotel. Saat ini hotel membutuhkan data *feedback* atau *review* dari konsumen. Hal ini dikarenakan melalui *review* dari konsumen, hotel dapat menilai apakah hotel tersebut telah memenuhi kepuasan konsumen atau belum. Selain itu, dengan menilai dari *review*, hotel dapat mengevaluasi terhadap *brand* yang dimiliki. Salah satu cara untuk melihat dan menilai *review* hotel adalah melalui *website* Tripadvisor. Tripadvisor adalah sebuah *website* dan aplikasi *travel online* dengan menyediakan pemesanan penginapan, hotel, restoran, dan *review* pengalaman pengguna.

Menurut Moscado (2015), terdapat hubungan kompleks antara *tourism* dan *sustainability*. Dan kunci untuk menyelesaikan ini adalah dengan adanya rasa tanggung jawab antar *stakeholder*. Faktanya, terdapat 3 permasalahan yang belum bisa dicapai oleh sektor *tourism* saat ini. Permasalahan tersebut adalah permasalahan ekonomi, sosial, dan lingkungan. Menurut Gimenez, Sierra, & Rodon (2012), perusahaan atau hotel tidak hanya harus berfokus pada sosial dan lingkungan saja, tetapi harus berfokus pada keuangan yang positif selama berproses, karena ketika hotel mampu mengintegrasikan ketiga permasalahan tersebut, maka hotel mampu mencapai *sustainability*. Saat ini, perilaku masyarakat telah berubah, sama seperti perilaku konsumen yang berubah karena kebebasan berpendapat di Internet (Kotler, 2013). Menurut Amatulli (2017), konsumen saat ini peduli dengan isu sosial dan lingkungan. Sehingga waktu yang tepat bagi sektor *tourism* termasuk hotel untuk memperhatikan *sustainability* dalam menjalankan operasi dan bisnis hotel.

Penelitian terhadap *brand personality* hotel ini dilakukan mengingat semakin banyaknya pesaing. Hotel harus dapat beradaptasi dengan cepat untuk melakukan perubahan yang lebih baik. Untuk membuat *brand personality* hotel yang baik, hotel memerlukan data hasil *review* pengguna dengan cara metode *text mining*. Metode ini berguna untuk menilai *review* pengguna karena kebutuhan akan informasi tentang produk pariwisata saat ini lebih tinggi, dan proses pengambilan keputusan untuk memilih hotel terutama didasarkan pada pernyataan yang dipublikasikan oleh pengguna sebelumnya di *platform* Tripadvisor. Oleh karena itu, bisnis perhotelan perlu mengetahui *insight* dari *review* pengguna, sehingga hotel dapat melakukan evaluasi terhadap *brand personality* yang mereka miliki. Menambahkan *brand personality* hotel dapat membuat konsumen merekomendasikan hotel tersebut, menyebarkan *Word of Mouth* (WOM) positif, dan mendorong konsumen lain untuk membelinya.

I.2 Identifikasi dan Rumusan Masalah

Penelitian *text mining* ini berasal dari penelitian sebelumnya yaitu "*Discovering The Sustainable Hotel Brand Personality on Tripadvisor*" (Neto, Silva, Ferreira, & Araújo, 2020). Pada penelitian ini terdapat 3 hotel yang diteliti berdasarkan lokasi dan destinasi turis yaitu *London*, *New York*, dan *Sydney*, namun sayangnya penelitian tidak menyebutkan secara spesifik mengenai nama

hotel dan lokasi hotel yang diteliti. Pada *website* Tripadvisor terdapat fitur pencarian yang dapat berguna untuk memilih lokasi hotel yang dituju. Selain itu, terdapat fitur *sort by* yang berfungsi untuk menyesuaikan berdasarkan harga, jarak, *rating*, dan gabungan semuanya (*best value*). Setelah memilih berdasarkan pilihan, maka *website* akan menampilkan daftar hotel - hotel yang tersedia. Lalu ketika memilih salah satu hotel, maka akan terdapat *review* dari pengguna. *Review* ini sangat berguna untuk informasi bagi perusahaan hotel dan calon konsumen terkait kualitas hotel tersebut. Berdasarkan penelitian dari (Neto, Silva, Ferreira, & Araújo, 2020), terdapat 6 atribut yang dimodifikasi dari penelitian Aaker (1997) dengan menambahkan satu atribut terkait *brand personality* hotel yaitu *sustainability*. Atribut *sustainability* ditambahkan karena seperti yang dikatakan oleh Amatulli (2017), perilaku masyarakat saat ini sudah mulai berubah, saat ini masyarakat mulai *concern* atau *aware* terhadap permasalahan sosial dan lingkungan, dengan demikian, atribut – atribut yang diteliti pada penelitian ini terdiri dari *sincerity*, *excitement*, *sophistication*, *competence*, *sustainability*, dan *ruggedness*.

Penelitian ini berharap agar hotel menjadi lebih peduli terhadap *sustainability* pada *brand personality* hotel. Berdasarkan penelitian sebelumnya, didapatkan jumlah *trait* dari penelitian Aaker (1997) dan Opoku (2006), dan dilakukan penyesuaian sehingga mendapatkan total 213 *trait* untuk dimensi *sustainability*. Hasil dari 213 *trait* ini tidak dijelaskan secara spesifik terkait proses dan metodologi nya. Sehingga *trait trait* pada tiap dimensi tidak diketahui proses mendapatkannya. Contohnya pada dimensi *ruggedness* terdapat *trait challenge*. Peneliti tidak tahu bagaimana metode memperoleh *trait* tersebut, karena tidak adanya transparansi dalam proses pencarian *trait*.

Pada penelitian ini dilakukan *text mining*. Dari hasil *text mining* ini, diperoleh *insight* yang dapat berguna untuk hotel dalam mengevaluasi *review* dari konsumen. *Text mining* pada penelitian sebelumnya dilakukan dengan menghitung frekuensi terhadap *keyword* yaitu *trait* dari tiap dimensi tanpa memperhatikan konteks kalimat, lalu hasil frekuensi tiap *trait* ini direkap dan dihitung jumlahnya. Berikut ini merupakan gambar 1 hasil *text mining* yang dilakukan terhadap *brand personality* adalah sebagai berikut.

Dimension	Trait	Freq.	(%)
Sincerity	Friendly	3352	21.43
	Helpful	3025	19.34
	Standard	1023	6.54
	Pleasant	767	4.90
	Open	572	3.66
	Decent	476	3.04
	Kind	438	2.80
	Polite	417	2.67
	Warm	414	2.65
	Welcoming	413	2.64
Excitement	Modern	1165	19.74
	Early	1039	17.60
	New	955	16.18
	Fresh	449	7.61
	Awesome	417	7.06
	Recent	196	3.32
	Young	196	3.32
	Cool	195	3.30
	Crazy	112	1.90
	Moving	102	1.73
Sophistication	Expensive	941	19.79
	Pretty	707	14.87
	Beautiful	680	14.30
	Fabulous	350	7.36
	Lady	294	6.18
	Spectacular	245	5.15
	Luxurious	142	2.99
	Top notch	138	2.90
	Stunning	131	2.75
	First class	101	2.12
Competence	Able	819	21.27
	Safe	420	10.91
	Outstanding	346	8.99
	Experienced	253	6.57
	In front	170	4.42
	Superior	164	4.26
	Complete	153	3.97
	Secure	105	2.73
	Trade	100	2.60
	Knowledgeable	94	2.44
Sustainability	Lovely	1060	28.25
	Care	487	12.98
	Fairly	236	6.29
	Giving	177	4.72
	Fair	146	3.89
	Hospitality	145	3.86
	Aware	127	3.38
	Share	106	2.83
	Woke	82	2.19
	Peaceful	80	2.13

(continued)

Dimension	Trait	Freq.	(%)
Ruggedness	Outside	766	32.74
	Hard	561	23.97
	Difficult	224	9.57
	Uncomfortable	202	8.63
	Unpleasant	59	2.52
	Tough	39	1.67
	Rough	31	1.32
	Extreme	30	1.28
	Struggle	28	1.20
Challenge	27	1.15	

Gambar I.1 Hasil *Text Mining*

Sumber : Neto, A. P., Silva, E. A., Ferreira, L. V., & Araújo, J. F. (2020).
Discovering The Sustainable Hotel Brand Personality on TripAdvisor.

Berdasarkan Gambar I.1, untuk dimensi *sincerity*, *trait friendly* merupakan *trait* dengan memiliki frekuensi terbanyak. Sedangkan *trait welcoming* merupakan *trait* paling sedikit. Lalu, pada dimensi *excitement*, *trait modern* merupakan *trait* dengan frekuensi terbanyak. Lalu untuk *trait* paling sedikit pada dimensi *excitement* adalah *moving*. Pada dimensi *sophistication*, *trait expensive* merupakan *trait* dengan frekuensi terbanyak. Sedangkan *trait* dengan frekuensi paling sedikit adalah *first class*. Selanjutnya, pada dimensi *competence*, *trait*

dengan frekuensi terbanyak yaitu *able*, sedangkan frekuensi paling sedikit adalah *knowledgeable*. Lalu pada dimensi *sustainability*, *trait lovely* merupakan *trait* dengan frekuensi terbanyak. Sedangkan *trait* paling sedikit adalah *peaceful*. Terakhir, pada dimensi *ruggedness*, *trait* paling banyak adalah *outside*, sedangkan paling sedikit adalah *trait challenge*. Selain itu, terdapat *radar chart* yang bertujuan untuk memudahkan pembaca dalam melihat hasil data. *Chart* ini disajikan dalam bentuk visualisasi data. Berikut ini merupakan *radar chart*.



Gambar 1.2. *Radar Chart*

Sumber : Neto, A. P., Silva, E. A.,
Ferreira, L. V., & Araújo, J. F. (2020).

*Discovering The Sustainable Hotel Brand Personality
on TripAdvisor.*

Berdasarkan *chart* diatas, dimensi *sustainability* tidak terlalu signifikan dalam frekuensi *reviewer*. Hal ini dikarenakan data yang diambil dari data dari tahun lama hingga 2019. Sehingga ada kemungkinan masyarakat belum terlalu memperhatikan isu *sustainability* pada waktu itu. Namun disisi lain, ada kemungkinan metode perhitungan frekuensi dari penelitian sebelumnya kurang tepat. Oleh karena itu, masih ada kemungkinan masyarakat akan semakin *aware* dengan *sustainability* saat ini.

Hasil *text mining* di atas, masih memiliki kekurangan, karena proses yang dilakukan pada saat *text mining* hanya menghitung frekuensi saja tanpa memperhatikan konteks kalimatnya. Contohnya, kata "*cool*" bisa saja tidak memiliki konteks dengan hotel, namun kata tersebut terhitung sebagai frekuensi.

Hal ini dapat menyebabkan kekeliruan dalam mendapatkan *insight* dari hasil frekuensi.

Selain itu, jika diperhatikan pada *trait* tiap dimensi terdapat beberapa *trait* yang mungkin kurang sesuai untuk dijadikan representasi pada dimensi yang ada. Contohnya adalah pada dimensi *sophistication* terdapat *lady*, padahal sepertinya tidak terdapat hubungan antara dimensi *sophistication* dari sebuah hotel dengan kata *lady*. Sehingga hasil rekap frekuensi yang ada belum tentu akurat, karena frekuensi yang diperoleh belum tentu akurat, maka *radar chart* pun bisa jadi kurang akurat. Alhasil, kesimpulan terkait *sustainability* terhadap *brand personality* pun bisa jadi kurang akurat.

Pada penelitian sebelumnya menggunakan metode perhitungan yang sederhana, sehingga ada peluang untuk memperbaiki dengan metode yang baru. Akan tetapi perbaikan dari metode yang baru ini pun tidak akan sempurna karena interpretasi yang dibuat berdasarkan metode *text mining*. Metode yang akan digunakan dalam penelitian ini adalah metode pencarian *trait* menggunakan *word embedding* dan *word2vec*. Metode tersebut dapat memperoleh kemiripan kata atau sinonim kata dari setiap *trait*, tujuannya agar mendapatkan informasi sebanyak mungkin dari setiap *trait*. Lalu, dilakukan analisa menggunakan *dependency parsing* dan *dependency tree*. Tujuannya adalah untuk menganalisa kalimat berdasarkan konteks dengan cara memperhatikan stuktur kalimatnya. Setelah itu, dilakukan *sentiment analysis* untuk memberikan usulan perbaikan berdasarkan *review* konsumen.

Penelitian ini lebih baik dari penelitian sebelumnya karena terdapat beberapa perbaikan. Pertama, penelitian ini transparan. Metode dan proses pencarian *trait*, *dependency parsing*, *sentiment analysis* hingga usulan perbaikan dijelaskan secara lengkap dan jelas, sehingga peneliti lain dapat memahami penelitian ini dengan baik. Kedua, proses dalam mengekstrak informasi berdasarkan frekuensi dilakukan berdasarkan konteks menggunakan *dependency parsing & dependency tree*. Sehingga *insight* yang diperoleh tidak keliru. Ketiga, proses pencarian *trait* menggunakan metode *word embedding* (*word2vec*). Metode ini bertujuan untuk mengubah teks menjadi bentuk *vector* agar dapat diproses dan mendapatkan *most similar words* untuk mendapatkan sinonim kata dari setiap *trait* yang dibantu dengan *library wordnet* sehingga mendapatkan informasi sebanyak mungkin dan akurat. Terakhir, pada penelitian ini terdapat usulan perbaikan

kualitas hotel berdasarkan metode *sentiment analysis*. Dengan menggunakan metode ini, peneliti dapat menganalisis *review* dari konsumen terhadap kualitas hotel berdasarkan hasil sentimen pada setiap *trait* dan dimensi.

Fokus pada penelitian ini adalah hotel Lotte New York Palace, adapun hotel RIU Plaza digunakan dalam penelitian ini sebagai pembuktian bahwa metode yang dilakukan dalam penelitian ini dapat direplikasi dan diimplementasikan pada hotel yang berbeda dengan tujuan untuk menjawab penelitian sebelumnya yang tidak menyebutkan lokasi penelitian hotel secara spesifik. Berdasarkan identifikasi tersebut, maka dapat dirumuskan masalah sebagai berikut:

1. Bagaimana metode untuk menemukan *trait* berdasarkan metode *word embedding* dan *word2vec* dari data *review* yang dimiliki?
2. Bagaimana metode untuk menentukan kepentingan dimensi *sustainability* terhadap dimensi – dimensi lain yang sudah ada pada Aaker (1997) ?
3. Bagaimana performansi yang diperoleh dari metode *text mining* yang diusulkan?
4. Bagaimana usulan perbaikan terhadap dimensi *brand personality* hotel yang dijadikan studi kasus yaitu hotel Lotte New York Palace berdasarkan *sentiment analysis*?

I.3 Pembatasan Masalah dan Asumsi Penelitian

Setelah mengidentifikasi masalah, maka langkah selanjutnya adalah dibuat batasan masalah. Batasan masalah bertujuan untuk membatasi ruang lingkup penelitian sehingga penelitian lebih terfokus pada permasalahan yang akan diselesaikan. Pembatasan masalah pada penelitian ini adalah sebagai berikut.

1. Penelitian dilakukan menggunakan data dari website Tripadvisor pada hotel Lotte New York Palace dari tahun 2013 sampai 2022 dan Hotel RIU Plaza dari tahun 2016 hingga 2022.
2. Data ini diperoleh dari review konsumen pada website Tripadvisor terhadap hotel Lotte New York Palace dan Hotel RIU Plaza.

Selain Batasan masalah, dalam penelitian ini pun terdapat asumsi. Asumsi yang digunakan dalam penelitian ini adalah data diambil satu kali dari *review* hotel di Tripadvisor bukan dari hasil pengulangan.

I.4 Tujuan Penelitian

Tujuan penelitian merupakan sesuatu yang ingin dicapai berdasarkan identifikasi masalah dan rumusan masalah penelitian. Terdapat tiga poin tujuan dalam penelitian ini. Berikut ini merupakan tujuan penelitian ini.

1. Merancang metode untuk menemukan *trait* berdasarkan metode *text mining* dan menggunakan *review* yang dimiliki.
2. Merancang metode untuk menentukan kepentingan dimensi *sustainability* terhadap dimensi – dimensi lain yang sudah ada pada Aaker.
3. Menghitung performansi yang diperoleh dari metode *text mining* saat ini.
4. Memperoleh usulan perbaikan pada *brand personality* hotel Lotte New York Palace berdasarkan hasil dari *sentiment analysis*.

I.5 Manfaat Penelitian

Penelitian ini dapat berdampak positif bagi pihak yang terkait. Hal ini dikarenakan penelitian ini dilakukan guna untuk memperbaiki dari penelitian sebelumnya. Berikut ini merupakan manfaat penelitian yang dilakukan.

1. Dapat mengetahui metode yang dapat diimplementasikan pada bisnis hotel dengan menggunakan metode *text mining*.
2. Mengetahui kelebihan dari metode *text mining* penelitian saat ini yang digunakan dalam mencari *insight*.
3. Memberikan usulan metode terbaik dalam melakukan *text mining* terhadap *brand personality* hotel.

I.6 Metodologi Penelitian

Pada subbab ini akan dijelaskan mengenai metodologi penelitian. Metodologi penelitian ini akan dijadikan pedoman untuk tata cara pelaksanaan penelitian. Berikut ini tahapan – tahapan yang dapat dilihat beserta gambar diagram alir metodologi penelitian dibawah ini.

1. Identifikasi dan Rumusan Masalah

Identifikasi permasalahan merupakan proses awal untuk mengumpulkan informasi secara umum. Identifikasi masalah dilakukan melalui studi literatur atau penelitian sebelumnya. Setelah identifikasi masalah terkumpul, maka langkah selanjutnya adalah menentukan rumusan masalah. Rumusan masalah ini adalah

permasalahan penelitian yang menjadi fokus untuk diteliti dan diselesaikan. Rumusan masalah dalam penelitian ini mencakup mengenai metode untuk menentukan trait berdasarkan text mining dari review yang dimiliki, metode untuk menentukan level dimensi sustainability terhadap dimensi lain, dan perbandingan antara metode penelitian saat ini dengan penelitian sebelumnya.

2. Pembatasan Masalah dan Asumsi

Pembatasan masalah dilakukan agar ruang lingkup penelitian tidak terlalu luas dan melebar. Hal ini dilakukan agar dapat fokus untuk menyelesaikan permasalahan secara mendalam. Sedangkan asumsi bertujuan untuk mengurangi kompleksitas dengan mengurangi faktor – faktor tertentu.

3. Penentuan Tujuan Penelitian

Tujuan penelitian merupakan hal yang ingin dicapai dalam penelitian. Tujuan penelitian berhubungan dengan identifikasi masalah dan rumusan masalah penelitian. Pada penelitian ini terdapat tiga tujuan penelitian yang ingin dicapai yaitu merancang metode untuk menemukan trait berdasarkan proses *text mining*, merancang metode untuk menentukan kepentingan tiap dimensi pada *brand personality hotel*, dan menghitung performansi dengan menggunakan metode *text mining*.

4. Tinjauan Pustaka

Tinjauan pustaka merupakan tahap mengumpulkan dan mencari informasi yang berhubungan dengan penelitian sebagai landasan teori. Tinjauan pustaka dapat ditemukan pada buku, jurnal, ataupun refrensi dari internet lainnya. Pada bagian tinjauan pustaka ini, terdapat teori yang menjadi landasan dalam penelitian ini seperti *word embedding*, *word2vec* serta teori *dependency parsing & dependency tree*.

5. Pengumpulan Data

Pengumpulan data pada penelitian ini merupakan data dari website Tripadvisor. Data ini diperoleh dari review pengguna dan diambil dari tahun 2013 hingga tahun 2022 pada hotel Lotte New York Palace. Selain itu, pengumpulan data dilakukan untuk hotel RIU Plaza dari tahun 2016 hingga tahun 2022. Pengumpulan data dilakukan satu kali pengambilan. Dari hasil pengumpulan data ini akan diolah untuk proses selanjutnya dalam penelitian.

6. Pencarian Trait

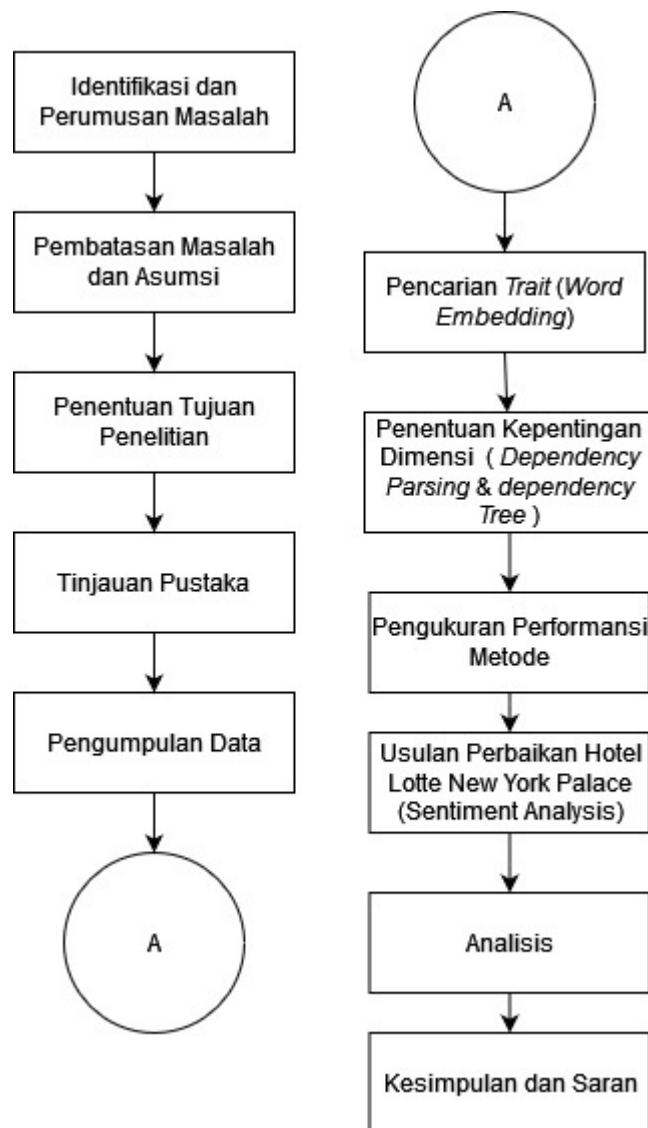
Pencarian *trait* dilakukan untuk menentukan *trait* yang sesuai dengan dimensi yang ada dari penelitian sebelumnya. Pencarian *trait* dilakukan dengan metode *word embedding*. Hal ini dilakukan untuk mendapatkan kata yang sesuai dengan memprediksi kata-kata disekitarnya atau dengan mendapatkan *output* kata-kata di sekitarnya berdasarkan *input* atau target. Pencarian *trait* ini memerlukan teknik *Natural Language Processing* (NLP) seperti *word2vec* dan *wordnet*.

7. Penentuan Kepentingan Dimensi *Sustainability*

Salah satu tujuan penelitian ini adalah meneliti level kepentingan dari dimensi *sustainability* terhadap dimensi – dimensi yang lain berdasarkan metode *text mining*. Hal ini dikarenakan Hotel perlu mengetahui seberapa penting dimensi *sustainability* terhadap *brand personality* hotel. Untuk menentukan level dari dimensi *sustainability* digunakan metode *dependency parsing* dan *dependency tree*.

8. Pengukuran dan Performansi Metode

Pengukuran performansi dilakukan menggunakan metode *dependency parsing* dan *dependency tree*. Pengukuran dilakukan dengan mengambil sampel data penelitian untuk memeriksa performansi metode *dependency parsing* dan *dependency tree*. Selain itu, akan dibandingkan metode *dependency parsing* dan *dependency tree* pada hotel Lotte New York Palace dengan hotel RIU Plaza.



Gambar I.3. Diagram Alir Metodologi Penelitian

9. Usulan Perbaikan Hotel Lotte New York Palace

Pada penelitian ini terdapat usulan perbaikan pada hotel Lotte New York Palace. Usulan perbaikan ini dilakukan setelah proses *dependency parsing*. Usulan perbaikan *brand personality* hotel diperoleh berdasarkan *review* konsumen menggunakan metode *sentiment analysis*. Metode ini dapat mengetahui respon konsumen terhadap *brand personality* hotel Lotte New York Palace.

10. Analisis

Pada bagian ini akan dijelaskan analisis dari proses pengumpulan data, lalu dilanjutkan dengan analisis pencarian trait menggunakan *word embedding*,

setelah itu adalah analisis metode *dependency parsing*, lalu analisis pengukuran performansi dari kedua metode *dependency parsing* yaitu *dependency parsing parent child* dan *sibling* pada hotel Lotte New York Palace dan hotel RIU Plaza. Setelah itu, dilakukan analisis terkait usulan perbaikan *brand personality* hotel Lotte New York Palace berdasarkan *sentiment analysis*.

11. Kesimpulan dan Saran

Terakhir adalah kesimpulan. Kesimpulan ini akan menjawab dari rumusan masalah penelitian. Sehingga tujuan penelitian dapat tercapai. Kesimpulan merupakan jawaban dari penelitian berdasarkan proses dan metodologi yang sistematis. Lalu, terdapat saran dari peneliti, tujuannya agar penelitian kedepannya dapat lebih baik.

I.7 Sistematika Penulisan

Pada penelitian ini akan dijelaskan dalam 5 bab yaitu pendahuluan, tinjauan pustaka, pengumpulan dan pengolahan data, analisis, serta kesimpulan dan saran. Berikut ini penjelasan sistematika penulisan pada tiap bab.

BAB I PENDAHULUAN

Pada bab ini merupakan gambaran umum, dan pengenalan masalah. Terdapat latar belakang masalah, lalu penjelasan mengenai identikasi dan perumusan masalah, setelah itu adalah menentukan pembatasan masalah dan asumsi penelitian, lalu terdapat tujuan penelitian, manfaat penelitian, serta terakhir metodologi penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Pada bab ini merupakan pemaparan teori yang mendasari penelitian ini. Teori mengenai pencarian trait, analisa struktur antar kata dalam suatu kalimat menggunakan *dependency parsing*, dan teori lainnya yang mendukung penelitian ini. Sehingga penelitian ini selalu berdasarkan teori dari tinjauan pustaka.

BAB III PENGUMPULAN DAN PENGOLAHAN DATA

Pada bab ini akan membahas pengumpulan dan pengolahan data yang akan membantu dalam menjawab pertanyaan penelitian. Proses pengumpulan dan pengolahan data, diawali dengan pengumpulan data penelitian, lalu dilanjutkan dengan membahas pencarian trait, setelah itu, terdapat analisa menggunakan metode *dependency parsing* dan *dependency tree*, lalu menghitung performansi metode, dan terakhir terdapat usulan perbaikan.

BAB IV ANALISIS

Pada bagian ini akan menganalisis dari hasil pengumpulan dan pengolahan data. Bab analisis akan menjelaskan lebih dalam mengenai hasil dari proses sebelumnya. Selain itu, pada bagian analisis akan menjelaskan terkait metode yang digunakan dalam penelitian.

BAB V KESIMPULAN DAN SARAN

Pada bagian ini akan dijelaskan mengenai kesimpulan dan saran penelitian. Kesimpulan dibuat untuk mendapatkan ringkasan dari poin pertanyaan penelitian, sekaligus menjawab penelitian. Sedangkan saran bertujuan untuk memperbaiki penelitian kedepannya hingga penelitian ini menjadi lebih baik.

