

**PERANCANGAN MODEL PREDIKSI *TRIP TYPE*
PADA ULASAN *ONLINE* HOTEL MENGGUNAKAN
*FEATURE WORD VECTOR***

SKRIPSI

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

Disusun oleh:

Nama : Joel Javier Barnabas

NPM : 2017610101



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
2021**

**PERANCANGAN MODEL PREDIKSI *TRIP TYPE*
PADA ULASAN *ONLINE* HOTEL MENGGUNAKAN
*FEATURE WORD VECTOR***

SKRIPSI

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

Disusun oleh:

Nama : Joel Javier Barnabas

NPM 2017610101 Disetujui untuk sidang



Dedy Suryadi

15/08/21



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
2021**

**FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG**



Nama : Joel Javier Barnabas
NPM : 2017610101
Jurusan : Teknik Industri
Judul Skripsi : PERANCANGAN MODEL PREDIKSI *TRIP TYPE* PADA
ULASAN *ONLINE* HOTEL MENGGUNAKAN *FEATURE WORD*
VECTOR

TANDA PERSETUJUAN SKRIPSI

Bandung, Agustus 2021

Ketua Jurusan Teknik Industri


7 Sept 2021
(Dr. Ceicalia Tesavrita, S.T., M.T.)

Pembimbing Pertama


31 Agustus 2021

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

Pembimbing Kedua



(Dr. Ceicalia Tesavrita, S.T., M.T.)



Jurusan Teknik Industri
Fakultas Teknologi Industri
Universitas Katolik Parahyangan

Pernyataan Tidak Mencontek atau Melakukan Tindakan Plagiat

Saya, yang bertanda tangan di bawah ini,

Nama : Joel Javier Barnabas

NPM : 2017610101

dengan ini menyatakan bahwa skripsi dengan judul :

***“PERANCANGAN MODEL PREDIKSI TRIP TYPE
PADA ULASAN ONLINE HOTEL MENGGUNAKAN
FEATURE WORD VECTOR”***

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.

Bekasi, 12 Agustus 2021

Joel Javier Barnabas
2017610101

ABSTRAK

Hotel butuh untuk melakukan evaluasi untuk meningkatkan kualitas hotel tersebut. Ulasan online melalui Tripadvisor merupakan salah satu cara hotel melakukan evaluasi. Untuk melakukan evaluasi dibutuhkan trip type pelanggan, namun banyak pelanggan yang menuliskan ulasan dan tidak mengisi trip type pada ulasannya. Oleh karena itu dibuatlah model prediksi trip type untuk memprediksi trip type suatu ulasan. Pembuatan model prediksi menggunakan feature word vector dengan model doc2vec untuk mengubah ulasan menjadi vektor. Doc2vec dapat menggambarkan arti atau karakteristik dari sebuah dokumen menjadi vektor. Data yang diproses akan dibagi menjadi data train dan test dengan rasio 80:20. Namun dikarenakan adanya imbalanced dataset maka dilakukan over sampling dengan metode SMOTE terhadap data train. Dalam pengubahan ulasan menjadi vektor terdapat beberapa parameter yang digunakan yaitu vector size, window, dan min_count. Didapatkan nilai parameter terbaik untuk setiap parameter adalah vector size 700, window 200, dan min_count 1. Setelah itu dilakukan penentuan model klasifikasi yang terbaik antara Decision Tree, kNN, Logistic Regression, dan SVM. Didapatkan model klasifikasi terbaik adalah SVM dengan rata-rata akurasi 57%. Semua penentuan dibuktikan dengan uji Z dua populasi untuk menguji rata-rata akurasi terbaik.

ABSTRACT

Hotel needs to evaluate to enhance the hotel quality. From online reviews through Tripadvisor is one of other options for the hotel to evaluate. For a hotel to evaluate they need their customer trip types, but there is a problem lots of customer did not fill their trip type on their online review. That's why the trip type prediction model is made for predicting the review trip type. Prediction model is made using word vector feature with doc2vec model to change reviews into vectors. Doc2vec is able to describe the meaning or characteristic of a document into a vector. The data that were used are divide into data train and test with the ratio 80:20. But because of imbalanced dataset needs to do over sampling using SMOTE for the data train. In transforming reviews into vectors there is some parameters that are used that is vector size, window, and min_count. From there gotten the best parameter score for every parameter is vector size 700, window 200, and min_count 1. After that deciding the best classification model between Decision Tree, kNN, Logistic Regression and SVM. From there got the best classification model is SVM with 57% accuracy. All these are proven with Z-test for two population for testing the best mean accuracy.

KATA PENGANTAR

Puji dan Syukur kepada Tuhan Yang Maha Esa karena berkat yang sudah dilimpahkan oleh-Nya, Penulis dapat menyelesaikan penelitian skripsi ini untuk menyelesaikan pendidikan Sarjana saya dalam bidang Teknik Industri. Penelitian skripsi Saya yang berjudul “Perancangan Model Prediksi *Trip Type* Pada Ulasan *Online Hotel Menggunakan Feature Word Vector*” dapat tersusun dengan baik.

Selain itu juga selama penyusunan penelitian skripsi ini banyak pihak yang membantu, membimbing, dan juga mendukung penulis baik secara mental dan fisik. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan juga banyak terima kasih kepada:

1. Orang tua dan adik penulis Papah, Mamah, dan Dedek. Yang selalu menemani penulis dalam penyusunan skripsi ini.
2. Bapak Dedy Suryadi, S.T., M.S., Ph.D dan Dr. Ceicalia Tesavrita, S.T., M.T. sebagai dosen pembimbing skripsi penulis yang sudah membimbing, menuntun, dan memberi arahan selama proses penelitian skripsi ini.
3. Prof. Sani Susanto, Ph,D dan Bapak Fransiscus Rian Pratikto, ST., M.T., MIE yang sudah menjadi dosen penguji proposal skripsi atas catatan, komentar, dan masukkan kepada penulis dalam sidang proposal.
4. Toi, Dikun, Adot, Mp, Andre, Tope, Memet, Nyo, Samsung, Albert, Ditsur, Patrick, Randy, Frederick, Leonard, Abun, Paulus, Kinen, Jm, Dika, Sammy, Felix, Claude, Louis, Jeffrey, Jason, Yoga, Vincent Napitupulu, Colin, Primus, William. Yang sudah mengimbangi kehidupan bermain penulis selama penyusunan skripsi.
5. Vincent Nicholas dan Brandon yang sudah membantu penulis dalam mempelajari penyusunan program yang digunakan dalam penelitian skripsi ini.

Mungkin masih banyak pihak lain yang belum disebutkan satu per satu oleh penulis. Penulis ingin mengucapkan terima kasih juga kepada pihak lain yang belum disebutkan. Demikian penulis berharap penelitian skripsi ini dapat bermanfaat bagi orang banyak dan khususnya penelitian selanjutnya.

DAFTAR ISI

ABSTRAK	i
ABSTRACT	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 Rumusan dan Identifikasi Masalah	I-4
I.3 Pembatasan Masalah dan Asumsi Penelitian	I-8
I.4 Tujuan Penelitian	I-9
I.5 Manfaat Penelitian	I-9
I.6 Metodologi Penelitian	I-9
I.7 Sistematika Penulisan	I-12
BAB II TINJAUAN PUSTAKA	II-1
II.1 <i>Machine Learning</i>	II-1
II.2 <i>Text Preprocessing</i>	II-2
II.3 <i>Word Vector</i>	II-2
II.3.1 <i>Doc2Vec</i>	II-3
II.4 <i>Supervised Learning</i>	II-5
II.4.1 <i>k-Nearest Neighbors (k-NN)</i>	II-5
II.4.2 <i>Decision Tree</i>	II-6
II.4.3 <i>Logistic Regression</i>	II-10
II.4.4 <i>Support Vector Machines (SVM)</i>	II-12
II.5 Akurasi, Presisi, dan <i>Recall</i>	II-13
II.6 <i>Imbalanced Dataset</i>	II-15
II.7 Uji Z	II-16
BAB III PENELITIAN ANDA	III-1

III.1 Pengambilan Data Ulasan <i>Online</i>	III-1
III.2 <i>Text Preprocessing</i>	III-1
III.2.1 <i>Tokenizing</i>	III-1
III.2.2 <i>Stop Words Removal</i>	III-2
III.2.3 <i>Lemmatization</i>	III-3
III.3 Perancangan Algoritma Utama	III-3
III.4 Implementasi Algoritma Utama	III-6
III.4.1 Implementasi <i>Doc2Vec</i>	III-6
III.4.3 Penentuan Parameter	III-6
III.5 Kinerja Model Prediksi	III-12
BAB IV ANALISIS	IV-1
IV.1 Analisis Proses <i>Text Preprocessing</i>	IV-1
IV.2 Analisis Perancangan Algoritma Utama	IV-2
IV.3 Analisis Proses Penentuan Parameter	IV-2
IV.4 Analisis Kinerja Model	IV-3
BAB V KESIMPULAN SARAN	V-1
V.1 Kesimpulan	V-1
V.2 Saran	V-1

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR RIWAYAT HIDUP

DAFTAR TABEL

Tabel III.1	Hotel dan Ulasan Hasil <i>Web Scraping</i>	III-1
Tabel III.2	<i>Tokenizing</i>	III-2
Tabel III.3	<i>Stop Words Removal</i>	III-3
Tabel III.4	<i>Lemmatization</i>	III-3
Tabel III.5	SMOTE Pada <i>Data Train</i>	III-5
Tabel III.6	Contoh Hasil Matriks Vektor	III-6
Tabel III.7	Hasil Percobaan <i>Vector Size</i>	III-8
Tabel III.8	Hasil 30 Replikasi Akurasi Penentuan <i>Vector Size</i>	III-8
Tabel III.9	Hasil Uji Z <i>Vector Size</i>	III-9
Tabel III.10	Hasil Percobaan <i>Window</i>	III-10
Tabel III.11	Hasil 30 Replikasi Akurasi Penentuan <i>Window</i>	III-11
Tabel III.12	Hasil Uji Z <i>Window</i>	III-12
Tabel III.13	Hasil Percobaan <i>Min_Count</i>	III-12
Tabel III.14	Akurasi Setiap Model Klasifikasi	III-13
Tabel III.15	Hasil Uji Z Antar Model Klasifikasi	III-13
Tabel III.16	Contoh <i>Confusion Matrix</i>	III-15
Tabel III.17	Contoh Hasil <i>Recall, Precision, f1Score, dan Accuracy</i>	III-16

DAFTAR GAMBAR

Gambar I.1	Grafik Banyak Pengguna Tripadvisor	I-2
Gambar I.2	Contoh Ulasan pada Tripadvisor	I-2
Gambar I.3	Proporsi Pengisian <i>Trip Type</i> di Amerika	I-3
Gambar I.4	Jumlah Wisatawan Mancanegara ke Bali 2017-2020	I-3
Gambar I.5	Contoh Ulasan Tidak Memiliki <i>Trip Type</i>	I-4
Gambar I.6	Contoh Ulasan Memiliki <i>Trip Type</i>	I-5
Gambar I.7	Proporsi Pengisian <i>Trip Type</i>	I-6
Gambar I.8	Proporsi Pengisian Tipe Perjalanan Anvaya Hotel	I-6
Gambar I.9	Sistem Pergerakan Informasi Saat Ini	I-7
Gambar I.10	Sistem Pergerakan Informasi Dengan Model Prediksi	I-7
Gambar I.11	<i>Flow Chart</i> Metodologi Penelitian	I-10
Gambar II.1	<i>Machine Learning Ingredients</i>	II-1
Gambar II.2	PV-DM Model	II-4
Gambar II.3	PV-DBOW Model	II-4
Gambar II.4	Ilustrasi <i>k-Nearest Neighbors</i>	II-6
Gambar II.5	Algoritma <i>Naïve Bayes</i>	II-6
Gambar II.6	Contoh <i>Feature Tree</i>	II-7
Gambar II.7	Algoritma <i>Decision Tree</i>	II-7
Gambar II.8	<i>Impurity Metrics</i>	II-9
Gambar II.9	Rumus <i>Impurity</i>	II-9
Gambar II.10	Algoritma <i>Best Split</i>	II-9
Gambar II.11	Grafik Sigmoid	II-10
Gambar II.12	Algoritma <i>Logistic Regression</i>	II-11
Gambar II.13	<i>Decision Boundary</i>	II-12
Gambar II.13	Rumus SVM.....	II-12
Gambar II.13	Rumus <i>Soft Margin</i>	II-12
Gambar III.1	<i>Flowchart</i> Algoritma Utama	III-4
Gambar III.2	Grafik <i>Vector Size</i> Terhadap Akurasi	III-5
Gambar III.3	Grafik <i>Window</i> Terhadap Akurasi.....	III-10
Gambar III.4	Grafik <i>Min_Count</i> Terhadap Akurasi	III-13

DAFTAR LAMPIRAN

LAMPIRAN A CONTOH ULASAN <i>ONLINE</i>	A-1
LAMPIRAN B <i>SYNTAX</i> PROGRAM	B-1

BAB I

PENDAHULUAN

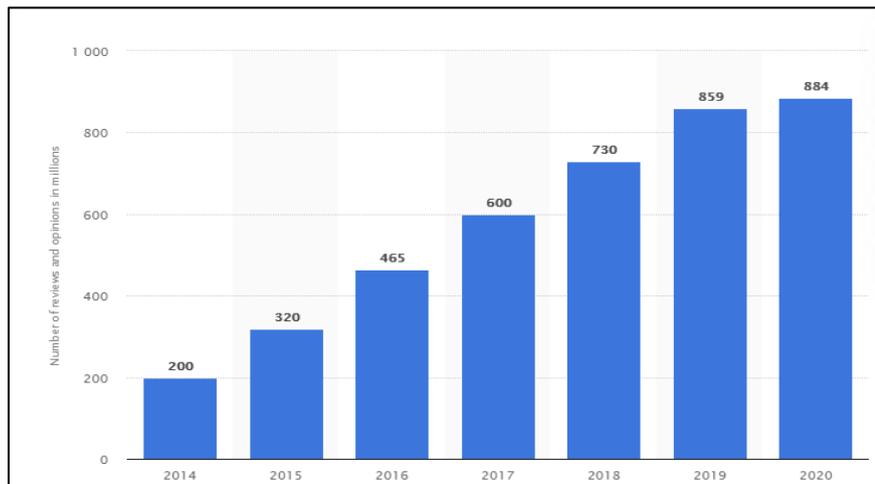
Dalam bab ini akan dijelaskan mengenai latar belakang masalah, identifikasi dan perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan yang dilakukan. Dalam penelitian mengenai perancangan model prediksi menggunakan fitur *word vector*.

I.1 Latar Belakang Masalah

Sebagian besar wisatawan mancanegara di seluruh dunia butuh tempat penginapan, salah satu alternatif yang sering digunakan adalah hotel. Banyak wisatawan yang menggunakan *web* Tripadvisor untuk mencari hotel. *Web* ini merupakan platform untuk membantu wisatawan mengetahui tempat yang akan dikunjungi atau mencari tahu tempat yang akan dikunjungi berdasarkan pengalaman wisatawan yang sebelumnya sudah pernah mengunjungi tempat tersebut. Dalam *web* ini sudah terdapat berbagai macam bahasa dan juga berbagai destinasi yang sudah cukup lengkap.

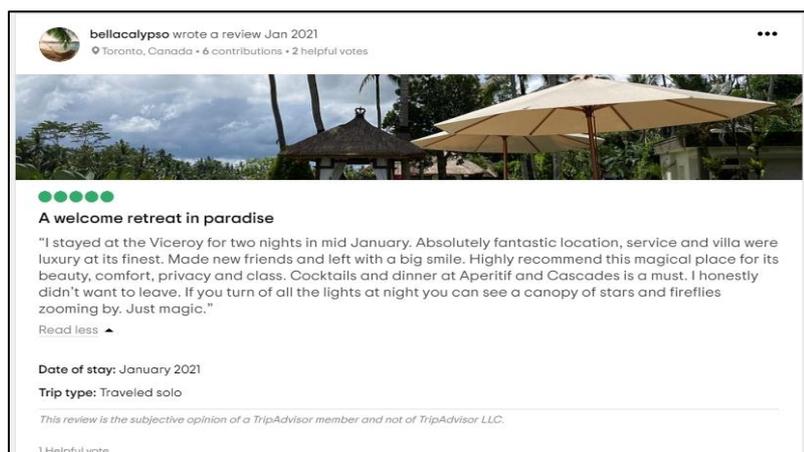
Sebuah perusahaan butuh melakukan evaluasi untuk memperbaiki dan meningkatkan kualitas perusahaan tersebut. Dalam melakukan evaluasi dibutuhkan *feedback* dari pelanggan atau pengguna. Oleh karena itu untuk tetap bertahan dan berkembang butuh *feedback* dari pelanggan mereka. Salah satu cara untuk mengetahui kebutuhan pelanggan adalah dengan membaca satu per satu ulasan atau komentar dari pelanggan.

Dengan adanya Tripadvisor, hotel yang sudah terdaftar dapat mengetahui ulasan dari pelanggan yang sudah menginap di hotel tersebut. Tripadvisor pertama ditemukan pada tahun 2000, Tripadvisor merupakan *web* yang sering digunakan oleh orang-orang untuk mencari ulasan begitu juga untuk evaluasi suatu bisnis khususnya bisnis hotel. Tripadvisor seringkali digunakan oleh wisatawan untuk mencari tahu mengenai destinasi yang akan dituju mereka. Berikut adalah grafik banyak pengguna Tripadvisor menurut Statista.com.



Gambar I.1 Grafik Banyak Pengguna Tripadvisor
(Sumber: Statista.com)

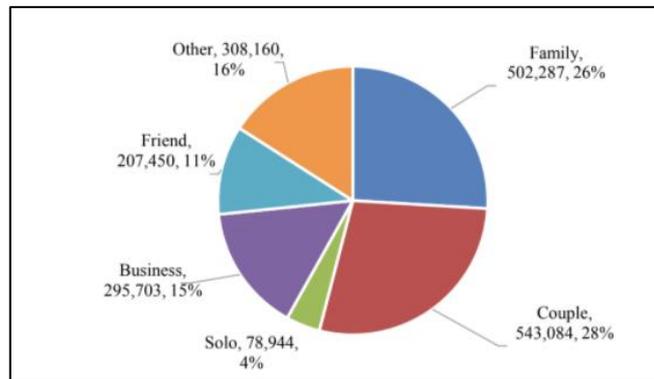
Namun para pengguna Tripadvisor pada saat mengisi ulasan atau *review* pada suatu tempat restoran ataupun tempat penginapan hotel sering kali tidak mengisi tipe perjalanan atau *trip type*. Yang merupakan tipe perjalanan contoh *solo traveller*, *family traveller*, atau *business trip*. Oleh karena itu untuk mempermudah wisatawan mengetahui hotel yang dikunjungi akan dibuat suatu program yang dapat menebak atau memprediksi tipe perjalanan dari suatu ulasan yang belum teridentifikasi. Berikut adalah contoh ulasan pada sebuah hotel yang menjadi salah satu objek penelitian ini.



Gambar I.2 Contoh Ulasan pada Tripadvisor
(Sumber: Tripadvisor.com)

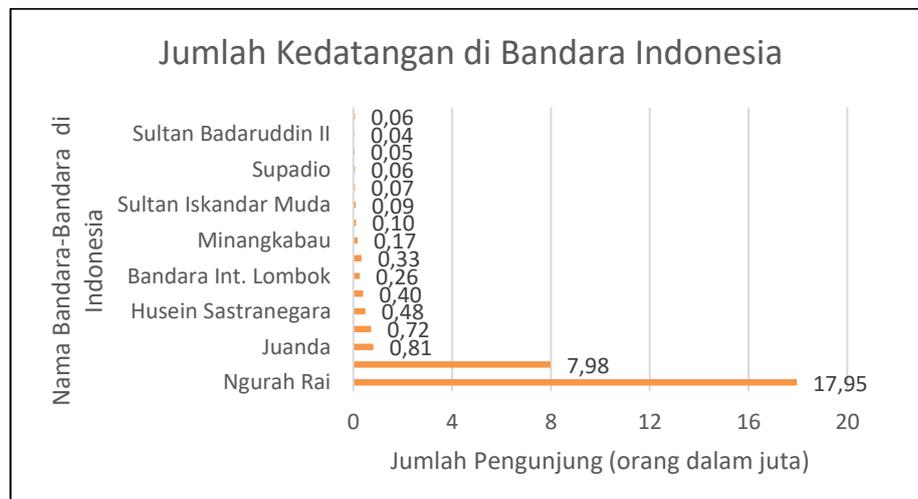
Dari contoh ulasan tersebut terlihat banyak data yang dapat digunakan untuk diolah. Metode *text mining* dapat mempermudah pemegang bisnis untuk

mengolah data, sehingga dapat meningkatkan dan memperbaiki bisnis tersebut terutama dalam penelitian ini yaitu bisnis hotel. Metode *text mining* juga bisa membantu memprediksi data yang kurang. Berdasarkan hasil penelitian yang sudah pernah dilakukan terdapat sebesar 16% dari ulasan yang terdapat di Tripadvisor hotel sekitar Amerika belum terdapat *trip type*. Berikut adalah proporsi pengisian tipe perjalanan di hotel sekitar Amerika.



Gambar I.3 Proporsi Pengisian *Trip Type* di Amerika (Sumber: Xiang, Ma, Du, & Fan, 2018)

Di Indonesia, Bali merupakan salah satu destinasi utama para wisatawan mancanegara. Sebagian besar wisatawan mancanegara datang ke Bali dibanding destinasi lain. Oleh karena itu penelitian ini menggunakan hotel-hotel yang terdapat di Bali dan terdaftar dalam Tripadvisor. Berikut adalah grafik kedatangan wisatawan mancanegara di Bandara Indonesia.



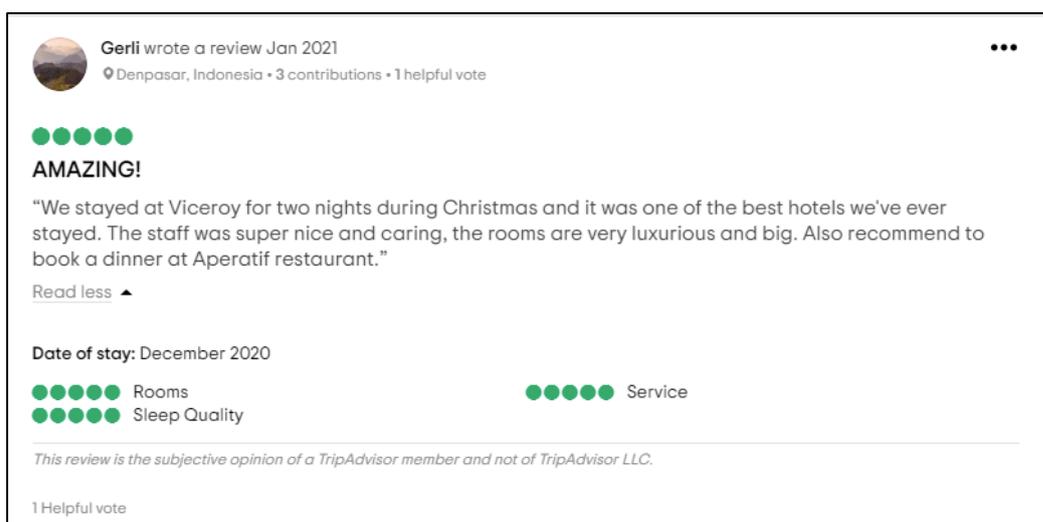
Gambar I.4 Grafik Kedatangan Wisatawan Mancanegara Di Bandara Indonesia (Sumber: Badan Pusat Statistik)

Terlihat pada grafik di atas jumlah kedatangan terbanyak adalah di Bandara Ngurah Rai Bali. Diabndingkan dengan Bandara Indonesia yang lainnya Bali memiliki jumlah pengunjung yang jauh lebih banyak. Oleh karena itu dalam pemilihan hotel yang terdapat dalam Tripadvisor adalah hotel yang terdapat di Bali.

I.2 Identifikasi dan Rumusan Masalah

Dalam subbab ini akan dijabarkan mengenai identifikasi dan rumusan masalah. Hotel yang terdapat dalam Tripadvisor memiliki masalah karena banyak penulis ulasan online Hotel yang tidak memberikan informasi mengenai tipe perjalanan mereka. Jika hotel tidak mengetahui informasi tersebut maka Hotel akan lebih kesulitan untuk mengelompokkan pelanggan mereka. Jika hotel mengetahui tipe pelanggan mereka maka hotel akan semakin mudah untuk evaluasi dan meningkatkan kinerja hotel tersebut khususnya terhadap segmen konsumen yang memang menjadi target hotel tersebut. Dengan mengelompokkan pelanggan dapat memberi keunggulan dalam peningkatan kualitas perusahaan, berdasarkan Kotler dan Keller (2009).

Dalam penelitian ini hotel yang menjadi objek penelitian adalah hotel yang terletak di Bali. Ulasan yang terdapat pada Tripadvisor tidak terstruktur hanya merupakan teks berdasarkan pengalaman pengguna. Oleh karena itu dapat berguna bagi pemilik hotel untuk mengetahui pengalaman pelanggan mereka untuk evaluasi. Berikut adalah contoh ulasan yang tidak memiliki tipe perjalanan.



Gambar I.5 Contoh Ulasan Tidak Memiliki *Trip Type*.
(Sumber: Tripadvisor.com)

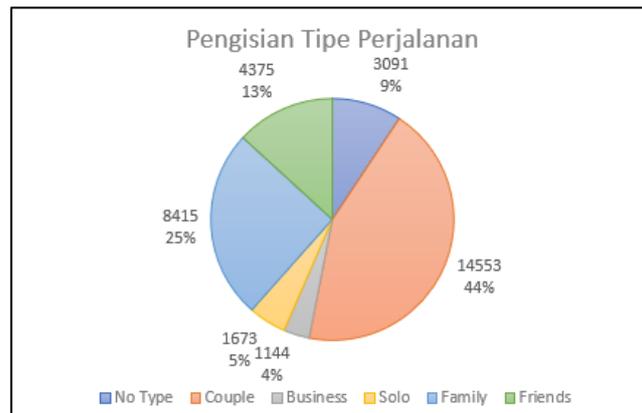
Dalam ulasan yang di atas terlihat tidak memiliki tipe perjalanan yang berarti penulis ulasan tidak mengisi informasi tersebut. Informasi mengenai tipe perjalanan sangat berguna bagi pemegang bisnis untuk melakukan evaluasi atau meningkatkan bisnis hotelnya. Dengan adanya informasi tersebut dapat membantu mempermudah mengklasifikasi perbaikan atau lain sebagainya. Berikut adalah contoh ulasan yang memiliki atau terdapat tipe perjalanan.



Gambar I.6 Contoh Ulasan Memiliki *Trip Type*
(Sumber: Tripadvisor.com)

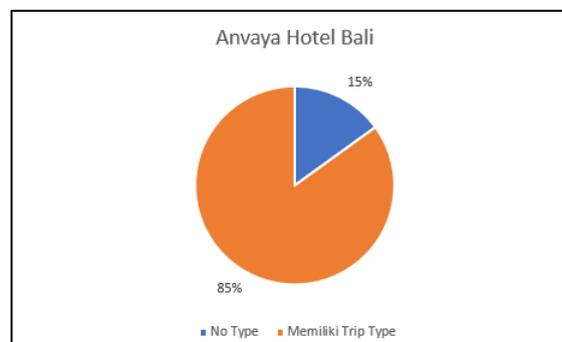
Pada contoh ulasan di atas terdapat informasi mengenai tipe perjalanan atau *trip type*. Pada contoh ulasan di atas tipe perjalanannya adalah *traveled on business* atau perjalanan bisnis. Terdapat beberapa macam tipe perjalanan yang lain yaitu *traveled solo*, *traveled as a couple*, *traveled with family*, *traveled with friends*. Oleh karena itu diperlukan prediksi tipe perjalanan. Dalam melakukan evaluasi hotel perlu mengetahui identifikasi dari ulasan tersebut. Sehingga hotel dapat melakukan evaluasi yang tepat.

Para pengguna Tripadvisor pada saat mengisi ulasan atau *review* pada suatu tempat penginapan hotel sering kali tidak mengisi tipe perjalanan atau *trip type*. Yang merupakan tipe perjalanan contoh *solo traveller*, *family traveller*, atau *business trip*. Terdapat cukup banyak ulasan yang tidak terdapat tipe perjalanan. Berikut adalah proporsi pengisian tipe perjalanan dari 20 hotel yang menjadi objek penelitian.



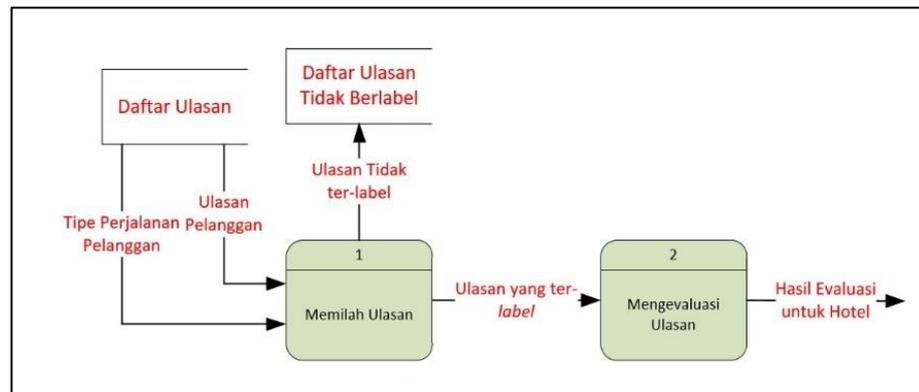
Gambar I.7 Proporsi Pengisian *Trip Type*
(Sumber: Tripadvisor.com)

Oleh karena itu untuk mempermudah mengetahui tipe perjalanan akan dibuat suatu model yang dapat memprediksi tipe perjalanan dari suatu ulasan yang belum memiliki tipe perjalanan. Pada grafik mungkin persentase yang tidak mengisi tipe perjalanan terlihat sedikit. Hal ini mungkin disebabkan karena hotel yang diambil secara acak oleh peneliti. Berikut adalah contoh hotel yang memiliki proporsi yang cukup besar.



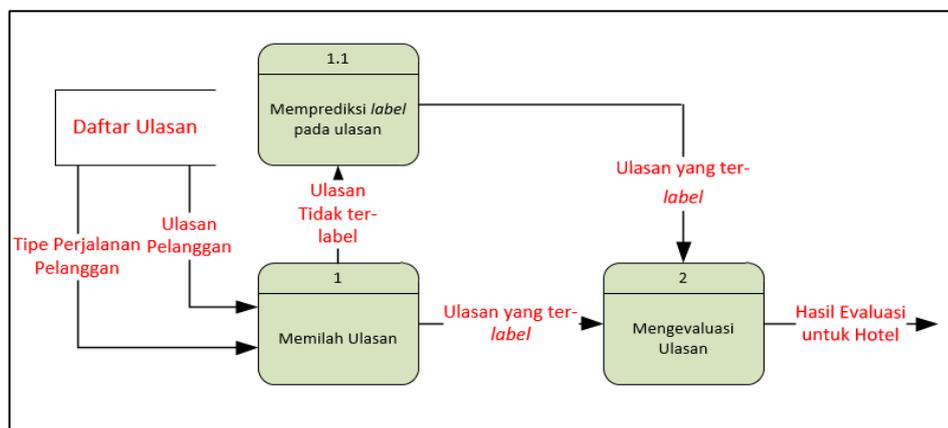
Gambar I.8 Proporsi Pengisian Tipe Perjalanan Anvaya Hotel
(Sumber: Tripadvisor.com)

Aliran informasi dalam proses pembacaan ulasan pelanggan yang digunakan hotel-hotel pada Tripadvisor saat ini masih bisa ditingkatkan. Pada aliran informasi terdapat ulasan yang tidak terlabel yang kurang dimanfaatkan. Dan diasumsikan hotel membuang ulasan yang tidak terlabel sehingga hotel kehilangan ulasan pelanggan. Berikut adalah gambaran sistem pergerakan informasi hotel saat ini.



Gambar I.9 Aliran Informasi Saat Ini

Pada Aliran informasi terdapat ulasan yang tidak terlabel jarang atau sulit untuk digunakan oleh pihak hotel. Hotel membutuhkan informasi mengenai tipe perjalanan dari sang penulis ulasan atau ulasan yang terlabel. Sehingga hotel dapat mengidentifikasi ulasan tersebut dan melakukan evaluasi secara tepat. Pada sistem pergerakan informasi saat ini juga masih membutuhkan waktu yang cukup lama untuk melakukan prediksi secara manual tanpa bantuan *machine learning*. Oleh karena itu diberikan usulan penambahan model prediksi.



Gambar I.10 Aliran Informasi Dengan Model Prediksi

Dengan model prediksi, maka daftar ulasan yang tidak terlabel dapat di prediksi sehingga menjadi ulasan yang terlabel. Sehingga tidak ada ulasan yang tidak terlabel dan dapat digunakan dengan mudah oleh pihak hotel. Pihak hotel tidak perlu melakukan prediksi secara manual yang membutuhkan waktu membaca yang lebih lama dan belum tentu dari membaca dapat mengidentifikasi *trip type* pada suatu ulasan. Pihak hotel mungkin dapat menebak secara acak dari semua tipe yang ada.

1. Bagaimana pembuatan model prediksi tipe perjalanan?
2. Bagaimana kinerja akurasi dari prediksi yang telah dilakukan oleh model prediksi?

I.3 Pembatasan Masalah dan Asumsi Penelitian

Pada subbab ini akan ditentukan batasan masalah dan asumsi yang digunakan dalam penelitian ini. Batasan masalah dan asumsi penelitian ditentukan sehingga penelitian yang dilakukan tidak membingungkan dan tersusun dengan rapih. Berikut adalah batasan masalah yang sudah ditentukan.

1. Ulasan yang digunakan terpilih secara acak dengan kata kunci pencarian "Bali".
2. Hotel yang dipilih memiliki ulasan lebih dari 1000.
3. Hotel yang dipilih untuk pengambilan data yang digunakan untuk penelitian merupakan Hotel di Bali yang memiliki minimal 1000 ulasan.

Setelah mendapatkan batasan masalah yang sudah ditetapkan berikut adalah asumsi yang digunakan dalam penelitian ini.

1. Ulasan yang digunakan adalah ulasan yang sudah dilabelkan oleh Tripadvisor berbahasa Inggris.
2. Tidak terjadi perubahan yang signifikan pada dataset yang digunakan dan diambil pada tanggal 29 maret sampai 6 april 2021, diasumsikan dapat digunakan.

I.4 Tujuan Penelitian

Dalam subbab ini akan dijabarkan mengenai tujuan penelitian yang diinginkan dari masalah yang sudah dijabarkan sebelumnya. Berdasarkan rumusan masalah dan batasan masalah yang sudah ditentukan, berikut adalah tujuan penelitian ini:

1. Dapat membuat model prediksi tipe perjalanan.
2. Mengetahui kinerja akurasi model prediksi dalam penelitian ini.

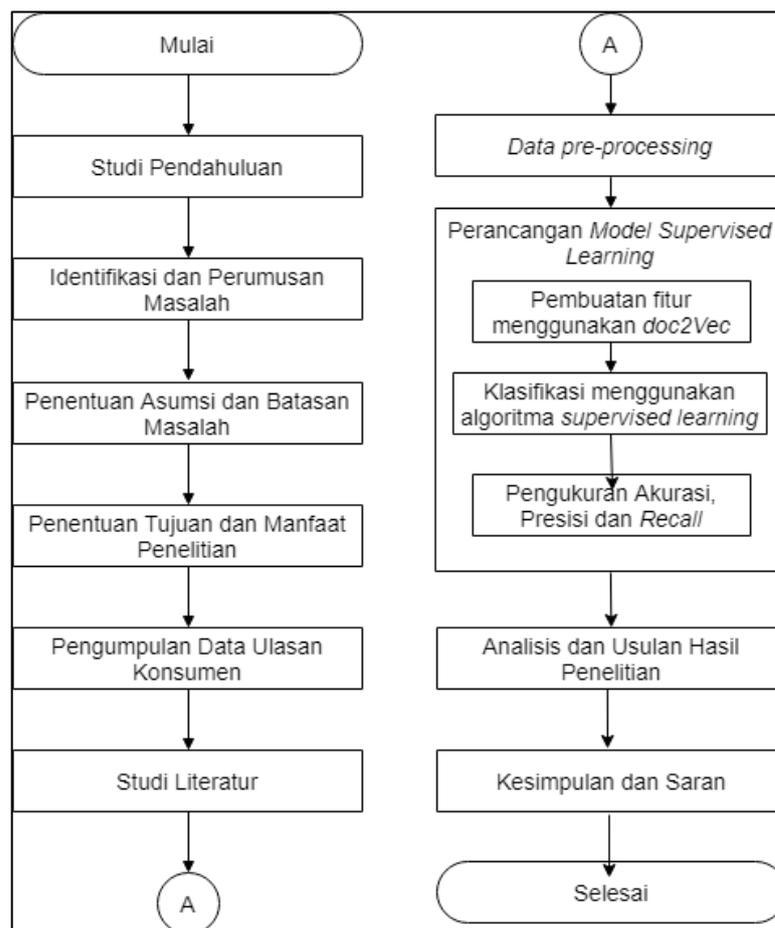
I.5 Manfaat Penelitian

Dalam penelitian yang dilakukan dapat memeberikan manfaat untuk pihak pemegang bisnis hotel, Tripadvisor, dan juga ilmu pengetahuan. Penelitian ini dapat membantu peneliti berikutnya menambah pengetahuan mengenai *text preprocessing*, *word vector*, dan *text classification*. Penelitian ini juga dapat

membantu pemegang bisnis hotel yang bersangkutan untuk lebih mengetahui pelanggan mereka. Penelitian ini juga membantu Tripadvisor untuk meningkatkan fitur yang ada pada Tripadvisor.

I.6 Metodologi Penelitian

Dalam subbab ini akan dijabarkan mengenai metodologi penelitian dalam penelitian ini. Metodologi penelitian ini digunakan sebagai tahapan yang akan dilaksanakan pada penelitian ini. Dimulai dari studi pendahuluan sampai penarikan kesimpulan dan saran dari hasil penelitian ini. Metodologi penelitian ini akan digambarkan dalam *flow chart* atau diagram alir. Terdapat *flow chart* yang memberikan gambaran mengenai langkah atau tahapan penelitian yang akan dilakukan. Maka selanjutnya akan dijabarkan untuk setiap tahapan. Berikut adalah *flow chart* metodologi penelitian ini.



Gambar I.11 Flowchart Metodologi Penelitian

Dari *flow chart* metodologi menggambarkan alur penelitian yang dilakukan. Setelah digambarkan alur penelitian ini selanjutnya akan dijelaskan untuk masing-masing bagiannya. Berikut adalah penjelasan untuk setiap bagian dari metodologi penelitian ini.

1. Studi Pendahuluan

Dalam melakukan studi literatur dilakukan studi pendahuluan juga untuk mengetahui objek penelitian. Mencari informasi mengenai ulasan pada Tripadvisor. Menemukan bahwa banyak ulasan yang tidak memiliki informasi mengenai tipe perjalanannya. Pada tahap ini juga mengumpulkan data untuk dipelajari.

2. Identifikasi dan Perumusan Masalah

Setelah melakukan studi literatur selanjutnya akan dilakukan identifikasi dan perumusan masalah. Dilakukan pencarian yang lebih mendalam berdasarkan informasi yang sudah didapat. Setelah itu dilakukan identifikasi masalah berdasarkan informasi yang sudah diteliti. Setelah itu dijadikan rumusan masalah pada penelitian ini.

3. Penentuan Asumsi dan Batasan Masalah

Dalam tahap ini akan ditentukan asumsi dan batasan masalah yang digunakan dalam penelitian ini. Asumsi dan batasan masalah yang digunakan agar penelitian ini tetap terarah.

4. Penentuan Tujuan dan Manfaat Penelitian

Pada tahap ini akan ditentukan tujuan dan manfaat dari penelitian ini. Tujuan penelitian ini untuk menjawab rumusan masalah yang sudah ada. Manfaat penelitian ini untuk pihak-pihak yang bersangkutan.

5. Pengumpulan Data Ulasan Konsumen

Dalam tahap ini dilakukan pengumpulan data dengan dilakukan *web scraping* menggunakan selenium Python. Ulasan yang digunakan adalah ulasan pada Tripadvisor hotel dibali yang memiliki setidaknya 1000 ulasan berbahasa inggris.

6. Studi Literatur

Dalam tahap ini akan dilakukan studi literatur mengenai penelitian ini. Terutama studi literatur mengenai *text mining* dan *machine learning*. Dalam tahap ini dilakukan pembelajaran mengenai *word2vec*, *text preprocessing*, *supervised learning*, dan mengenai evaluasi.

7. *Data Pre-processing*

Selanjutnya akan dilakukan data *pre-processing*, dalam tahap ini akan dilakukan pembersihan data. Data *pre-processing* digunakan agar data lebih mudah diproses. Seperti tokenisasi, menghilangkan *stop words*, *stemming*, dan *lemmatization*.

8. *Perancangan Model Supervised Learning*

Pada tahap ini dilakukan pembuatan fitur menggunakan metode *word2vec*. Setelah itu dilakukan klasifikasi dengan algoritma *supervised learning*. Algoritma yang digunakan merupakan algoritma dari *supervised learning*. Setelah itu akan dilakukan validasi dan verifikasi untuk memastikan program yang dirancang sesuai dengan algoritma yang sudah dirancang. Lalu akan dilakukan pengukuran akurasi, presisi, dan *recall* dari hasil rancangan ini.

9. *Analisis dan Usulan Hasil Penelitian*

Pada tahap ini dilakukan analisis terhadap hasil pengolahan data yang sudah dilakukan. Sehingga didapatkan usulan yang dapat diberikan.

10. *Kesimpulan dan Saran*

Pada tahap ini yang merupakan tahap terakhir yaitu penarikan kesimpulan yang sesuai dengan hasil penelitian. Dan menjawab rumusan masalah dan tujuan dari penelitian ini. Serta memberikan saran dari hasil penelitian ini dan juga saran kepada penelitian selanjutnya yang mirip.

I.7 Sistematika Penulisan

Penulisan hasil penelitian ini dibagi kedalam lima bab yang disusun sesuai sehingga dapat dimengerti dengan mudah. Dengan sistematika penulisan ini menjelaskan struktu penulisan penelitian ini. Berikut adalah sistematika penulisan penelitian yang dijadikan panduan dalam penulisan penelitian ini.

BAB I PENDAHULUAN

Dalam bab ini dijelaskan mengenai latar belakang masalah, identifikasi dan perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan yang dilakukan. Dalam penelitian mengenai perancangan model prediksi menggunakan fitur *word vector*.

BAB II TINJAUAN PUSTAKA

Dalam bab ini dijelaskan mengenai tinjauan pustaka dan studi literatur dalam penelitian ini. Pada bab ini akan dijelaskan mengenai teori-teori yang digunakan dalam penelitian ini. Teori yang akan dijelaskan terkait dengan *machine learning*, *word2vec*, *doc2vec*, *text preprocessing*, dan *supervised learning*. Teori-teori ini akan digunakan sebagai patokan dalam bab-bab selanjutnya.

BAB III PENGOLAHAN DATA

Pada bab ini dijelaskan mengenai proses pengumpulan data hingga pengolahan data. Data yang digunakan dari ulasan *online* hotel dalam aplikasi Tripadvisor. Dalam pengolahan data mentah akan diubah menjadi vektor yang digunakan sebagai fitur dalam *supervised learning* untuk menyelesaikan *task* yaitu prediksi model *trip type*.

BAB IV ANALISIS

Dalam bab ini berisi mengenai analisis hasil pengolahan data dari bab sebelumnya. Dalam bab ini akan analisis mengenai penerapan fitur *word vector* terhadap model yang telah dirancang. Dan juga analisis terhadap kinerja dari hasil perancangan model.

BAB V KESIMPULAN DAN SARAN

Dalam bab ini berisi mengenai kesimpulan dari hasil penelitian perancangan model prediksi pada ulasan *online* hotel menggunakan *feature word vector*. Dan akan dipaparkan saran untuk penelitian selanjutnya dan juga kepada pihak lain yang bersangkutan.