

**MODEL KLASIFIKASI *FEATURE-BASED* UNTUK
MEMPREDIKSI *TRIP TYPE* PADA ULASAN *ONLINE*
HOTEL**

SKRIPSI

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

Disusun oleh:

Nama : Chrisanto Andatama

NPM : 2017610221



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

**MODEL KLASIFIKASI *FEATURE-BASED* UNTUK
MEMPREDIKSI *TRIP TYPE* PADA ULASAN *ONLINE*
HOTEL**

SKRIPSI

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

Disusun oleh:

Nama : Chrisanto Andatama

NPM : 2017610221



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

**FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG**



Nama : Chrisanto Andatama
NPM : 2017610221
Jurusan : Teknik Industri
Judul Skripsi : MODEL KLASIFIKASI *FEATURE-BASED* UNTUK
MEMPREDIKSI *TRIP TYPE* PADA ULASAN *ONLINE* HOTEL

TANDA PERSETUJUAN SKRIPSI

Bandung, Agustus 2021

Ketua Program Studi Sarjana Teknik Industri

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

Pembimbing Pertama

Pembimbing Kedua

30 Agustus 2021

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

(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 : Chrisanto Andatama

NPM : 2017610221

dengan ini menyatakan bahwa skripsi dengan judul :

**“MODEL KLASIFIKASI *FEATURE BASED* UNTUK MEMPREDIKSI *TRIP TYPE*
PADA ULASAN *ONLINE HOTEL*”**

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.

Jakarta, 9 Agustus 2021

Chrisanto Andatama
2017610221

ABSTRAK

Ulasan *online* banyak memiliki informasi kebutuhan spesifik setiap kelompok wisatawan mancanegara. Pihak hotel dapat mengetahui kebutuhan tersebut dengan adanya atribut seperti *trip type* dalam pengisian ulasan. Namun saat ini diketahui terdapat banyak wisatawan yang tidak mengisi atribut *trip type* tersebut sehingga penafsiran isi ulasan dan kebutuhan yang diinginkan oleh setiap kelompok wisatawan menjadi lebih sulit untuk diinterpretasikan. Maka dibutuhkan sebuah pendekatan agar dapat membantu melakukan prediksi *trip type* dari ulasan *online* hotel. Salah satu metode untuk memprediksi tersebut dapat menggunakan sebuah model *feature based*.

Penelitian ini bertujuan untuk melakukan sebuah pemodelan untuk dapat memprediksi *trip type* dari sebuah ulasan *online* hotel di TripAdvisor dengan menerapkan metode *feature based* seperti *tf-idf*. Algoritma *tf-idf* merupakan sebuah algoritma yang dapat menunjukkan asosiasi antar sebuah kata dalam suatu dokumen dengan sebuah nilai yang merepresentasikan seberapa relevan sebuah kata atau *feature* terhadap sebuah dokumen. Algoritma *k-Nearest Neighbors (k-NN)*, *Decision Tree* dan *Naïve Bayes* digunakan untuk mengukur performansi model prediksi *trip type*. Pada penelitian ini juga dilakukan pengujian menggunakan metode *Distinguishing Feature Selection (DFS)* dan *raw term frequency*.

Berdasarkan hasil dari pengujian menggunakan model prediksi yang dikembangkan, terdapat dua metode yang dapat menghasilkan nilai rata-rata akurasi prediksi tertinggi. Metode tersebut merupakan menggunakan metode *tf-idf* dengan nilai rata-rata akurasi sebesar 48.1908% dan metode DFS sebesar 48.3178%. Kedua nilai terbaik tersebut didapatkan menggunakan parameter 300 *features* dan *classifier decision tree*. Dengan menggunakan uji z dua populasi untuk menguji rata-rata ($\alpha = 5\%$), dapat ditarik kesimpulan bahwa tidak terdapat perbedaan rata-rata secara signifikan antar kedua rata-rata tersebut yang menjadikan kedua metode dengan parameter tersebut merupakan metode yang terbaik untuk digunakan pada model ini. Terdapat juga hasil *feature importance* dimana terdapat 59 *feature* berpengaruh dengan metode DFS dan 87 *feature* berpengaruh dengan metode TF-IDF. Hasil yang diperoleh dari *feature importance* dapat menunjukkan kebutuhan dari konsumen berdasarkan karakteristik *feature* yang diperoleh sehingga dapat digunakan untuk evaluasi layanan hotel.

ABSTRACT

There's a lot of information given on specific needs of international tourist groups in online reviews. Hotels can retrieve information of their needs through attributes like trip types. It is known that tourists nowadays rarely fill their trip types when reviewing hotels, which makes interpreting reviews and the needs of each tourist groups harder to interpret. That is why a method to help predict a trip type from online hotel reviews using a feature-based model is needed.

This research aims to create a prediction model to predict trip types based of an online hotel review from TripAdvisor by applying a feature-based method like tf-idf. The tf-idf algorithm is an algorithm that is able to show the association between terms or features in a document with a score that represents how relevant a term or feature is towards the document. K-Nearest Neighbor, Decision Tree and Naïve Bayes algorithm is also used as a measurement of the performance of the model to predict trip types. In this research, Distinguishing Feature Selector (DFS) and raw term frequency methods are also applied. These methods choose differentiating and distinguishing features in a document as a measure towards the tests applied using tf-idf.

Based on the results from the tests using the prediction model that is developed, two methods produced the highest mean accuracy scores. Those methods are tf-idf with a mean accuracy of 48.1908% and the DFS method with a mean accuracy of 48.3178%. Both scores were tested using a parameter of 300 features by using the decision tree classifier. By using the z-test for two population means ($\alpha = 5\%$), we can conclude that there is no difference between both means significantly which concludes that both methods using the same parameter is the best method to be applied in this prediction model. Tests on feature importance is also applied where 59 features were retrieved by using the DFS method and 87 features were retrieved using the tf-idf method. Results from the tests of feature importance can show the needs of hotel customers through the characteristics of the features which can be evaluated by the hotels to satisfy customer needs.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya yang menyertai dalam proses penyusunan dan penyelesaian penelitian skripsi berjudul “Model Klasifikasi *Feature-Based* untuk Memprediksi *Trip Type* pada Ulasan *Online Hotel*” sehingga penelitian skripsi ini dapat dilakukan dan diselesaikan dengan baik.

Selain itu juga, dalam penyusunan skripsi ini tentunya ada banyak pihak yang turut membantu dan mendukung penulis. Oleh karena itu, pada kesempatan ini pula penulis ingin mengungkapkan dan menghaturkan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang terlibat dan telah memberikan bantuan serta dukungan, khususnya kepada:

1. Bapak Dedy Suryadi, S.T., M.S., Ph.D dan Ibu Dr. Ceicalia Tesavrita S.T., M.T., selaku dosen pembimbing yang telah membimbing, menuntun, serta memberi arahan dan juga masukan selama proses penyusunan penelitian skripsi ini.
2. Bapak Dr. Carles Sitompul, S.T., M.T., MIM dan Bapak Fransiscus Rian Pratikto, S.T., M.T., MIE selaku dosen penguji proposal atas masukan serta kritik yang diberikan kepada penulis dalam sidang proposal.
3. Orang tua dan keluarga penulis atas segala dukungan yang diberikan tanpa henti selama ini.
4. Keluarga Beneval khususnya untuk Joel Javier B., Randy Kristian L., Matthew Gunawan, Aditya Surjawan, Arsenius Arya C., Johan Lisandy N., Daniel Djunaidi, Paulus Nicholas G., Aldi P, Gabriel B., Patrick Christian, Kenneth Kristofer dan lainnya atas dukungan, hiburan, masukan, serta menemani selama masa perkuliahan dan proses penelitian skripsi penulis.
5. Sahabat lain Steven Alvin, Gilbert Nathanael C., Ian Hazali, Malvin Tanius, Derian Adiguna, Nicholas Tambunan, Geoffrey Timothy L., Sebastian Frederick, Leonard, Steven Liu, Andre Alexander, Xavier Mark, Kevin Christopher W. dan Shandion Edric S. yang memberikan

dukungan, motivasi, serta hiburan kepada penulis selama masa perkuliahan dan penelitian skripsi penulis

Penulis juga mengucapkan terima kasih kepada pihak lain yang belum disebutkan satu per satu oleh penulis yang membantu proses penelitian ini baik secara langsung maupun tidak langsung. Akhir kata, penulis memohon maaf apabila ada kesalahan dan kekurangan dalam penulisan kata-kata pada penelitian skripsi ini. Penulis berharap agar hasil penelitian ini ini dapat berguna dan bermanfaat bagi para pembaca.

Jakarta, 9 Agustus 2021

Penulis

DAFTAR ISI

DAFTAR ISI	v
DAFTAR TABEL	iii
DAFTAR GAMBAR	iii
DAFTAR LAMPIRAN	v
BAB I PENDAHULUAN	I-1
I.1 Latar Belakang	I-1
I.2 Identifikasi dan Rumusan Masalah	I-6
I.3 Pembatasan Masalah dan Asumsi Penelitian	I-12
I.4 Tujuan Penelitian	I-13
I.5 Manfaat Penelitian.....	I-13
I.6 Metodologi Penelitian	I-13
I.7 Sistematika Penulisan	I-16
BAB II TINJAUAN PUSATAKA	II-1
II.1 <i>Machine Learning</i>	II-Error! Bookmark not defined.
II.2 <i>Text Preprocessing</i>	II-2
II.3 <i>TF-IDF</i>	II-3
II.4 <i>Distinguishing Feature Selector (DFS)</i>	II-4
II.5 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	II-6
II.6 <i>Supervised Learning</i>	II-7
II.6.1 <i>k-Nearest Neighbors (kNN)</i>	II-7
II.6.2 <i>Naïve Bayes</i>	II-8
II.6.3 <i>Decision Tree</i>	II-10
II.7 <i>Confusion Matrix</i>	II-13
II.8 Uji Z.....	II-14
BAB III PENGOLAHAN DATA	III-1
III.1 Pemilihan Data Ulasan	III-1
III.2 <i>Text Preprocessing</i>	III-2
III.2.1 Normalisasi Teks.....	III-2
III.2.2 <i>Tokenizing</i>	III-3

III.2.3	<i>Stopwords Removal</i>	III-3
III.2.4	<i>Lemmatizing</i>	III-4
III.3	Algoritma Utama Prediksi <i>Trip Type</i>	III-5
III.3.1	Algoritma TFIDFVectorizer.....	III-6
III.3.2	Algoritma CountVectorizer.....	III-7
III.3.3	Algoritma SMOTE.....	III-8
III.3.4	Algoritma DFS.....	III-9
III.4	Implementasi Algoritma Utama Prediksi <i>Trip Type</i>	III-10
III.4.1	Penentuan Parameter dalam Implementasi Algoritma Utama ...	III-10
III.4.2	Implementasi <i>TF-IDF</i>	III-10
III.4.3	Implementasi <i>Raw Term Frequency</i>	III-12
III.4.4	Implementasi DFS.....	III-13
III.4.5	Implementasi Nilai Performansi Prediksi.....	III-14
III.5	Hasil TF-IDF.....	III-17
III.5.1	Hasil TF-IDF Parameter 100 <i>Features</i>	III-21
III.5.2	Hasil TF-IDF Parameter 300 <i>Features</i>	III-23
III.5.3	Hasil TF-IDF Parameter 500 <i>Features</i>	III-25
III.6	Hasil Raw Term Frequency.....	III-26
III.6.1	Hasil <i>Raw Term Frequency</i> Parameter 100 <i>Features</i>	III-30
III.6.2	Hasil <i>Raw Term Frequency</i> Parameter 300 <i>Features</i>	III-32
III.6.3	Hasil <i>Raw Term Frequency</i> Parameter 500 <i>Features</i>	III-34
III.7	Hasil DFS.....	III-36
III.7.1	Hasil DFS Parameter 100 <i>Features</i>	III-41
III.7.2	Hasil DFS Parameter 300 <i>Features</i>	III-43
III.7.3	Hasil DFS Parameter 500 <i>Features</i>	III-45
III.8	Rekapitulasi Hasil Performansi Prediksi.....	III-46
BAB IV	ANALISIS	IV-1
IV.1	Analisis Proses <i>Text Preprocessing</i>	IV-1
IV.2	Analisis Implementasi Algoritma Utama.....	IV-2
IV.3	Analisis Pemilihan Parameter.....	IV-4
IV.4	Analisis Performansi Prediksi Model.....	IV-5
IV.5	Analisis Metode Terpilih.....	IV-6
IV.6	Analisis Metode Terpilih.....	IV-9
BAB V	KESIMPULAN DAN SARAN	V-1

V.1	Kesimpulan	V-1
V.2	Saran.....	V-2

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR RIWAYAT HIDUP

DAFTAR TABEL

Tabel III.1 Tahap Normalisasi Teks Ulasan.....	III-3
Tabel II.2 Tahap <i>Tokenizing</i>	III-3
Tabel III.3 <i>Stopwords Removal</i>	III-4
Tabel III.4 <i>Lemmatizing</i>	III-4
Tabel III.5 Contoh Perhitungan TF-IDF	III-7
Tabel III.6 Contoh Hasil Nilai DFS	III-14
Tabel III.7 Hasil Performansi Prediksi 100 <i>features TF-IDF</i>	III-17
Tabel III.8 Hasil Performansi Prediksi 300 <i>features TF-IDF</i>	III-19
Tabel III.9 Hasil Performansi Prediksi 500 <i>features TF-IDF</i>	III-20
Tabel III.10 Rekapitulasi Nilai Akurasi dan Variansi TFIDF Parameter 100 <i>Features</i>	III-21
Tabel III.11 Contoh Uji Z TFIDF Parameter 100 <i>Features</i>	III-21
Tabel III.12 Rekapitulasi TFIDF Parameter 100 <i>Features</i>	III-22
Tabel III.13 Rekapitulasi Nilai Akurasi dan Variansi TFIDF Parameter 300 <i>Features</i>	III-23
Tabel III.14 Contoh Uji Z TFIDF Parameter 300 <i>Features</i>	III-23
Tabel III.15 Rekapitulasi TFIDF Parameter 300 <i>Features</i>	III-24
Tabel III.16 Rekapitulasi Nilai Akurasi dan Variansi TFIDF Parameter 500 <i>Features</i>	III-25
Tabel III.17 Contoh Uji Z TFIDF Parameter 500 <i>Features</i>	III-25
Tabel III.18 Rekapitulasi TFIDF Parameter 500 <i>Features</i>	III-26
Tabel III.19 Hasil Performansi Prediksi 100 <i>features Raw Term</i>	III-27
Tabel III.20 Hasil Performansi Prediksi 300 <i>features Raw Term</i>	III-28
Tabel III.21 Hasil Performansi Prediksi 500 <i>features Raw Term</i>	III-29
Tabel III.22 Rekapitulasi Nilai Akurasi dan Variansi <i>Raw Term</i> Parameter 100 <i>Features</i>	III-31
Tabel III.23 Contoh Uji Z <i>Raw Term</i> Parameter 100 <i>Features</i>	III-31
Tabel III.24 Rekapitulasi <i>Raw Term</i> Parameter 100 <i>Features</i>	III-32
Tabel III.25 Rekapitulasi Nilai Akurasi dan Variansi <i>Raw Term</i> Parameter 300 <i>Features</i>	III-32

Tabel III.26 Contoh Uji Z <i>Raw Term</i> Parameter 300 <i>Features</i>	III-33
Tabel III.27 Rekapitulasi <i>Raw Term</i> Parameter 300 <i>Features</i>	III-33
Tabel III.28 Rekapitulasi Nilai Akurasi dan Variansi <i>Raw Term</i> Parameter 500 <i>Features</i>	III-34
Tabel III.29 Contoh Uji Z <i>Raw Term</i> Parameter 500 <i>Features</i>	III-35
Tabel III.30 Rekapitulasi <i>Raw Term</i> Parameter 500 <i>Features</i>	III-36
Tabel III.31 Hasil Performansi Prediksi 100 <i>features</i> DFS	III-37
Tabel III.32 Hasil Performansi Prediksi 300 <i>features</i> DFS	III-38
Tabel III.33 Hasil Performansi Prediksi 500 <i>features</i> DFS	III-40
Tabel III.34 Rekapitulasi Nilai Akurasi dan Variansi DFS Parameter 100 <i>Features</i>	III-41
Tabel III.35 Contoh Uji Z DFS Parameter 100 <i>Features</i>	III-42
Tabel III.36 Rekapitulasi DFS Parameter 100 <i>Features</i>	III-42
Tabel III.37 Rekapitulasi Nilai Akurasi dan Variansi DFS Parameter 300 <i>Features</i>	III-43
Tabel III.38 Contoh Uji Z DFS Parameter 300 <i>Features</i>	III-43
Tabel III.39 Rekapitulasi DFS Parameter 300 <i>Features</i>	III-44
Tabel III.40 Rekapitulasi Nilai Akurasi dan Variansi DFS Parameter 500 <i>Features</i>	III-45
Tabel III.41 Contoh Uji Z DFS Parameter 500 <i>Features</i>	III-45
Tabel III.42 Rekapitulasi Uji Z DFS Parameter 500 <i>Features</i>	III-46
Tabel III.43 Rekapitulasi Nilai Akurasi Terbaik Setiap Metode dan Parameter	III-47
Tabel III.44 Uji Z Parameter dan <i>Classifier</i> Terbaik TF-IDF	III-47
Tabel III.45 Uji Z Parameter dan <i>Classifier</i> Terbaik <i>Raw Term Frequency</i>	III-48
Tabel III.46 Uji Z Parameter dan <i>Classifier</i> Terbaik DFS	III-49
Tabel III.47 Uji Z Parameter dan <i>Classifier</i> Terbaik	III-49
Tabel III.48 Contoh Hasil <i>Confusion Matrix</i>	III-51
Tabel III.49 <i>Multiclass Confusion Matrix</i>	III-52
Tabel III.50 Rekapitulasi Perhitungan <i>Confusion Matrix</i>	III-52
Tabel III.51 Rekapitulasi Hasil <i>Feature Importance</i> DFS	III-53
Tabel III.52 Rekapitulasi Hasil <i>Feature Importance</i> TF-IDF	III-54

DAFTAR GAMBAR

Gambar I.1 Jumlah Pengunjung Wisatawan Mancanegara	I-1
Gambar I.2 Alternatif Transportasi yang Digunakan Wisatawan Mancanegara	I-2
Gambar I.3 Kedatangan Wisatawan Mancanegara di setiap Bandara Indonesia	I-2
Gambar I.4 Tingkat Hunian Hotel di Bali Tahun 2015-2019	I-3
Gambar I.5 Survei Kepercayaan Ulasan <i>Online</i>	I-4
Gambar I.6 Contoh Ulasan TripAdvisor	I-6
Gambar I.7 Contoh Ulasan TripAdvisor (2)	I-7
Gambar I.8 Contoh Ulasan TripAdvisor (3)	I-8
Gambar I.9 Persentase Pengisian Trip Type	I-9
Gambar I.10 Ilustrasi Sistem P Membaca Ulasan Saat Ini	I-9
Gambar I.11 Ilustrasi Sistem Membaca Ulasan Dengan Model Prediksi	I-10
Gambar I.12 Metodologi Penelitian	I-14
Gambar II.1 Komponen Utama <i>Machine Learning</i>	II-1
Gambar II.2 Contoh Perhitungan TF	II-4
Gambar II.3 Contoh Algoritma Perhitungan TF	II-4
Gambar II.4 Contoh Algoritma Perhitungan DF	II-5
Gambar II.5 Contoh Algoritma Perhitungan TF-IDF	II-5
Gambar II.6 Contoh k-NN	II-8
Gambar II.7 Contoh Algoritma Naive Bayes.....	II-10
Gambar II.8 Contoh <i>Decision Tree</i>	II-11
Gambar II.9 Algoritma <i>Decision Tree</i>	II-11
Gambar II.10 Impurity Metrics.....	II-12
Gambar II.11 Algoritma Menentukan <i>BestSplit</i>	II-12
Gambar II.12 <i>Contigency Table</i>	II-13
Gambar III.1 Flowchart Algoritma Utama	III-6
Gambar III.2 Flowchart Algoritma TFIDFVectorizer	III-6
Gambar III.3 Flowchart Algoritma CountVectorizer.....	III-7
Gambar III.4 Flowchart Algoritma SMOTE.....	III-9

Gambar III.5 Flowchart Algoritma DFS.....	III-10
Gambar III.6 Syntax Algoritma TF-IDF.....	III-11
Gambar III.7 Syntax Algoritma <i>Raw Term Frequency</i>	III-12
Gambar III.8 Syntax Algoritma DFS.....	III-13
Gambar III.9 Replikasi Data matplotlib.....	III-15
Gambar III.10 Hasil Performansi Prediksi Model.....	III-16
Gambar III.11 Rekapitulasi Akurasi Terbaik	III-50

DAFTAR LAMPIRAN

Lampiran A Contoh Ulasan Pelanggan	A-1
Lampiran B <i>Syntax Program</i>	B-1
Lampiran C <i>Syntax Model Prediksi</i>	C-1
Lampiran D <i>Feature DFS</i>	D-1
Lampiran E Replikasi <i>Oversampling</i> dan <i>Undersampling</i>	E-1

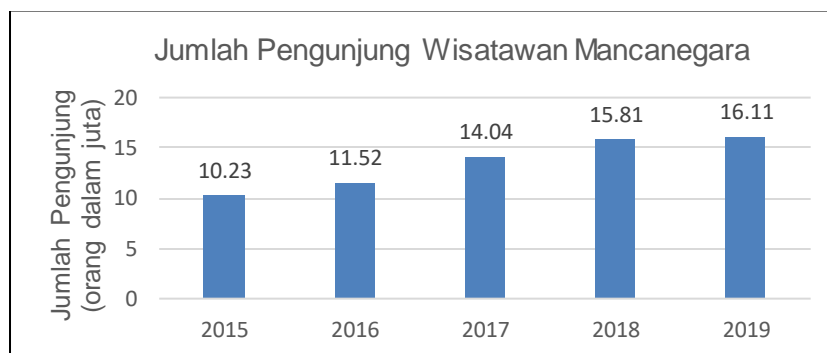
BAB I

PENDAHULUAN

Pada bab ini akan dipaparkan mengenai latar belakang masalah, identifikasi dan perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan yang dilakukan pada penelitian terkait klasifikasi *feature-based* dalam membangun model prediksi *trip type* pada ulasan *online* hotel.

I.1 Latar Belakang

Indonesia merupakan salah satu negara yang sering dikunjungi oleh wisatawan mancanegara. Jumlah wisatawan yang datang ke Indonesia mengalami peningkatan secara statistik dari tahun 2015 hingga tahun 2019. Berikut merupakan grafik dari jumlah pengunjung yang mengunjungi Indonesia dari tahun 2015 hingga 2019.

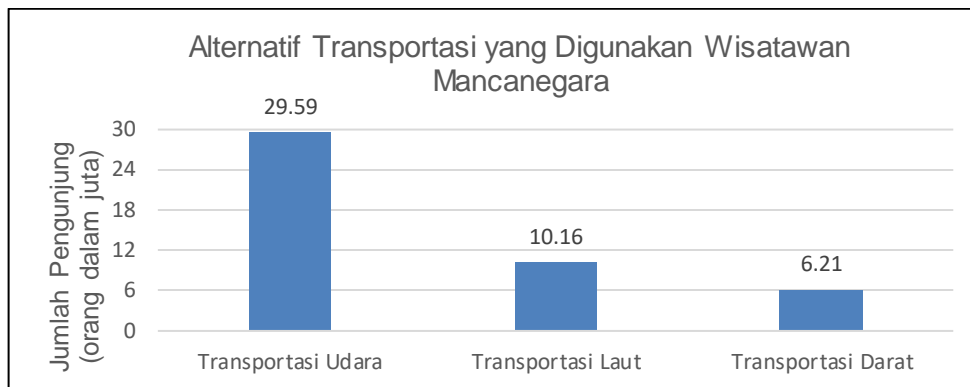


Gambar I.1 Jumlah Pengunjung Wisatawan Mancanegara
(Sumber:bps.go.id)

Pada Gambar I.1 dapat dilihat bahwa jumlah pengunjung yang datang ke Indonesia meningkat dari tahun 2015 hingga 2019. Hal ini menunjukkan bahwa industri pariwisata di Indonesia sedang bertumbuh dengan meningkatnya wisatawan mancanegara yang datang ke Indonesia.

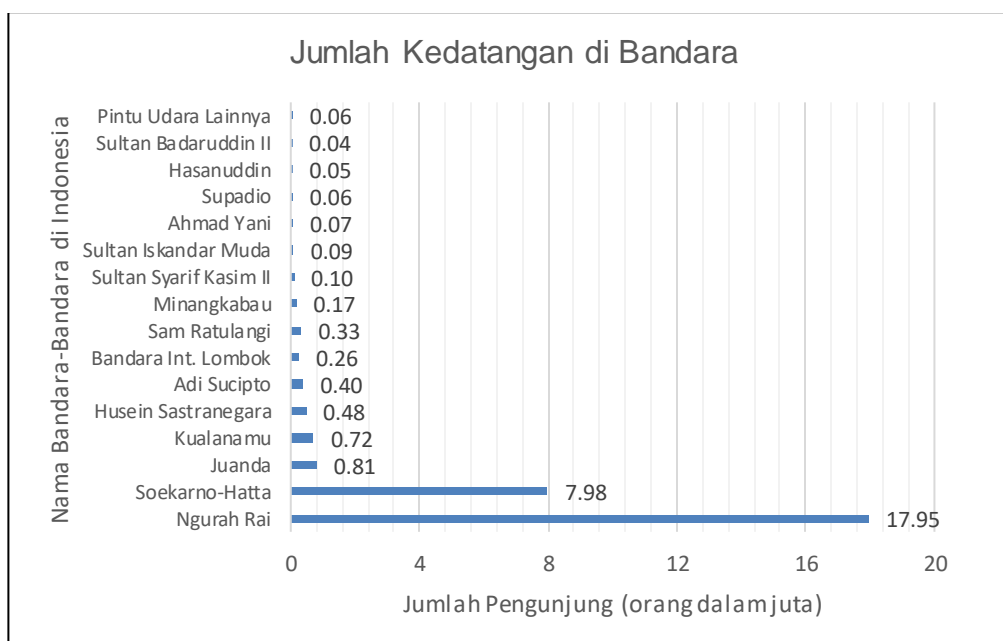
Dengan peningkatan tersebut, Indonesia memiliki potensi untuk menjadi negara destinasi pilihan favorit bagi wisatawan mancanegara, sehingga akan terdapat peningkatan wisatawan mancanegara yang datang ke Indonesia menggunakan berbagai alternatif transportasi. Menurut Badan Pusat Statistik

(BPS), terdapat alternatif transportasi yang paling sering digunakan oleh wisatawan mancanegara untuk datang ke Indonesia. Berikut merupakan alternatif transportasi yang paling sering digunakan oleh wisatawan mancanegara dari tahun 2017 hingga 2019.



Gambar I.2. Alternatif Transportasi yang Digunakan Wisatawan Mancanegara
(Sumber:bps.go.id)

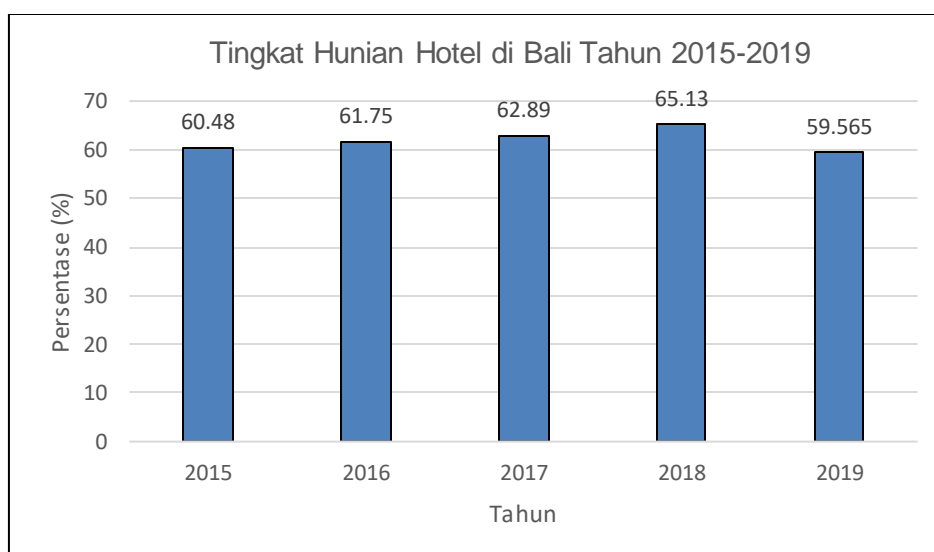
Sebanyak 29,59 juta orang wisatawan lebih memilih menggunakan transportasi udara atau pesawat untuk berpergian ke Indonesia dibandingkan menggunakan transportasi laut dan darat. Hal ini menunjukkan bahwa sebesar 64.3% wisatawan mancanegara pergi ke provinsi di Indonesia yang memiliki sebuah bandar udara. Pada Gambar 3 berikut terlihat persebaran total wisatawan mancanegara yang datang ke bandara Indonesia dari tahun 2017 hingga 2019.



Gambar I.3 Kedatangan Wisatawan Mancanegara di setiap Bandara Indonesia
(Sumber:bps.go.id)

Bandara Ngurah Rai yang berada di Provinsi Bali terdapat pengunjung wisatawan mancanegara sebesar 17.94 juta orang melakukan perjalanan ke Indonesia untuk berwisata ke provinsi Bali. Berdasarkan statistik tersebut, dapat dinyatakan bahwa provinsi Bali merupakan destinasi utama wisatawan mancanegara untuk datang ke Indonesia.

Bali sebagai destinasi utama para wisatawan mancanegara akan dikunjungi wisatawan yang tentunya membutuhkan layanan akomodasi. Dengan meningkatnya pengunjung yang datang, penyedia fasilitas akomodasi perlu menyanggah kebutuhan akomodasi para pengunjung serta meningkatkan kualitas layanan akomodasi yang disediakan. Salah satu akomodasi dasar yang sering digunakan oleh wisatawan adalah hotel. Hotel merupakan opsi akomodasi yang umumnya dipilih oleh para wisatawan. Berikut merupakan grafik dari tingkat hunian hotel di Bali dari tahun 2014 hingga 2019.



Gambar I.4. Tingkat Hunian Hotel di Bali Tahun 2015-2019
(Sumber:bps.go.id)

Tingkat penghunian kamar pada hotel berbintang di Bali cenderung memiliki tren yang meningkat dari tahun 2015 hingga 2019. Oleh karena itu, dapat disimpulkan bahwa wisatawan memiliki preferensi untuk memilih hotel sebagai fasilitas akomodasi yang umumnya digunakan dengan adanya peningkatan tersebut.

Sebelum melakukan perjalanan para wisatawan akan melakukan pemesanan hotel, pemesanan yang dilakukan umumnya dibagi sesuai dengan preferensi dan juga kebutuhan yang beragam dari kelompok-kelompok wisatawan yang ada. Maka dari itu, dibutuhkan pengetahuan mengenai informasi yang relevan terkait kebutuhan spesifik yang diinginkan oleh kelompok-kelompok wisatawan tersebut.

Salah satu cara untuk mengetahui informasi spesifik tentang hotel yang ada dapat dilihat dari ulasan *online*. Menurut sebuah survei yang dilakukan oleh statista.com, terdapat tingkat kepercayaan yang meningkat pada ulasan *online*, berikut merupakan hasil survei yang didapatkan.

	2014	2015	2016	2017	2018
Yes, always	-	8%	18%	19%	19%
Yes, if I believe the reviews are authentic	22%	31%	27%	27%	19%
Yes, for some types of businesses, no for others	34%	22%	19%	20%	15%
Yes, if there are multiple customer reviews to read	26%	19%	20%	20%	25%
No, I am often skeptical about online reviews	-	-	12%	13%	16%
No, I don't trust reviews at all	17%	20%	4%	3%	6%

Gambar I.5 Survei Kepercayaan Ulasan *Online*
(sumber: statista.com)

Pada Gambar I.5 dapat diketahui bahwa konsumen mulai memiliki tingkat kepercayaan yang lebih tinggi terhadap ulasan *online* dari tahun 2015 hingga 2018, statistik dari jawaban “Yes, *always*” menjelaskan hal tersebut. Sedangkan jawaban konsumen yang tidak memercayai ulasan *online* sama sekali mulai menurun dan terlihat pada jawaban “No, I don't trust reviews at all”. Maka dari itu, dapat disimpulkan bahwa ulasan *online* dapat menjadi sebuah referensi untuk mengetahui informasi lebih lanjut mengenai fasilitas akomodasi hotel.

Menurut geckoroutes.com, TripAdvisor merupakan *platform* untuk *travel* dan *tourism* yang paling banyak dikunjungi secara mendunia Menurut survei yang dibuat oleh statista.com, jumlah ulasan pada TripAdvisor terus meningkat dari

tahun 2014 hingga 2020. Secara keseluruhan, terdapat sekitar 884 juta ulasan yang dibuat hingga tahun 2020. Hal ini menandakan bahwa semakin banyak konsumen yang melakukan penulisan ulasan setelah mengunjungi sebuah tempat dan melakukan pemesanan pada *platform online*. Oleh karena itu, jumlah ulasan mengenai kebutuhan spesifik dan juga pengalaman kelompok-kelompok konsumen semakin meningkat. Semakin banyak ulasan yang ditulis maka akan terdapat lebih banyak pendapat yang berguna menjadi informasi setiap kelompok wisatawan.

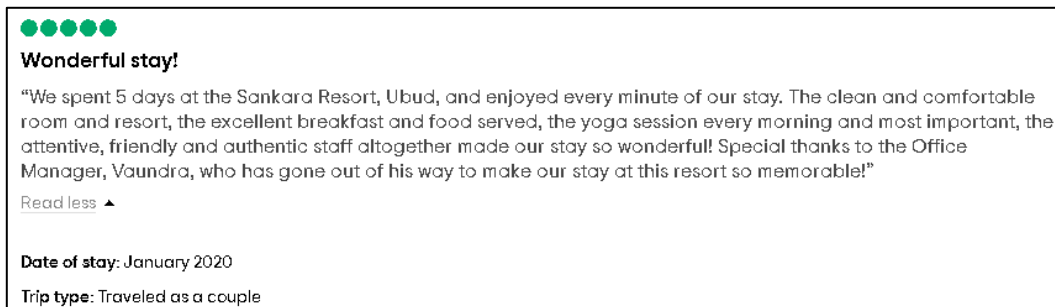
Dalam menulis ulasan hotel di Tripadvisor terdapat kolom ulasan yang dapat diisi oleh konsumen berdasarkan pengalaman mereka pada objek yang dikunjungi serta dapat memberi informasi tambahan mengenai *trip type* yang dilakukan seperti *solo traveler*, *family traveler*, *couple traveler* atau *business trip* dan penilaian untuk aspek-aspek penting hotel seperti *value*, *location*, *service*, *rooms*, *cleanliness* dan *sleep quality*.

Dunia saat ini sudah memasuki era atau zaman serba digital dimana hampir seluruh kegiatan manusia bergantung pada pemanfaatan internet, serta memungkinkan *platform tourism* dan *travel* seperti TripAdvisor mengakses data besar secara lebih mudah. Salah satu data besar yang bisa didapatkan oleh TripAdvisor merupakan ulasan *online* konsumen. Dengan peningkatan jumlah ulasan yang diberikan konsumen, informasi yang bisa diolah juga meningkat.

Dengan tersedianya informasi ulasan, *trip type* dan juga aspek-aspek penting lainnya, pihak hotel akan mampu untuk mengetahui kebutuhan dan keinginan spesifik yang dimiliki oleh setiap kelompok wisatawan. Namun dengan banyaknya ulasan yang diberikan oleh para wisatawan, banyak wisatawan yang tidak mengisi *trip type* pada ulasan, sehingga penafsiran isi ulasan serta kebutuhan yang diinginkan setiap kelompok perjalanan menjadi lebih sulit. Menurut Xiang, Du, Ma dan Fan (2017), sebanyak 16% dari konsumen yang mengisi ulasan hotel di Amerika tidak melakukan pengisian *trip type*.

Oleh karena itu untuk mempermudah pihak hotel dalam mengetahui kebutuhan yang diperlukan konsumen serta layanan yang perlu ditingkatkan hotel, diperlukan suatu metode untuk memprediksi tipe perjalanan dari suatu ulasan yang belum terlabelkan. Metode yang akan dimodelkan tersebut diperlukan agar dapat membantu pihak hotel, pihak TripAdvisor maupun pihak wisatawan agar dapat meningkatkan layanan dan juga dapat memenuhi kebutuhan dan keinginan

wisatawan dalam pelayanan akomodasinya. Berikut merupakan contoh ulasan pada salah satu hotel yang menjadi objek penelitian yaitu Hotel Sankara Ubud Pramana.



Gambar I.6. Contoh Ulasan *TripAdvisor*
(Sumber: TripAdvisor.com)

Ulasan pada hotel seperti pada contoh di atas merupakan data yang berbentuk dalam teks. Salah satu metode untuk mengolah data dalam bentuk teks adalah menggunakan metode *text mining*. *Text mining* dapat membantu fasilitator untuk memprediksi setiap kelompok wisatawan serta dapat mengetahui kebutuhan dari masing-masing kelompok wisatawan dari pola data ulasan *online* yang ditulis oleh para wisatawan.

I.2 Identifikasi dan Rumusan Masalah

Masalah yang disebutkan dalam subbab sebelumnya merupakan terdapat wisatawan yang tidak mengisi *trip type* dalam ulasan yang ditulis, sehingga hotel sulit untuk dapat menginterpretasikan kebutuhan serta keinginan spesifik yang dimiliki oleh para konsumen.

Menurut Kotler dan Keller (2009), melakukan segmentasi pasar memberi beberapa manfaat utama dibandingkan dengan pemasaran massal. Mengidentifikasi segmen pasar yang ada dan menentukan apa yang diminati dari setiap segmen pasar atau setiap kelompok wisatawan dapat memberikan jasa, harga serta penyampaian kualitas secara lebih tepat untuk setiap kelompok tersebut. Maka dari itu *problem owner* yaitu pihak hotel dapat terbantu dengan adanya segmentasi pasar yang jelas agar dapat memberikan layanan yang diinginkan oleh setiap kelompok wisatawan melalui ulasan yang diberikan.

Ulasan yang diberikan oleh para konsumen memiliki informasi penting yang dapat digunakan pihak TripAdvisor untuk mengambil informasi. Namun, jumlah ulasan yang sangat banyak dapat menjadi sebuah hal yang memakan waktu dan tidak efisien jika dibaca satu per satu. Oleh karena itu, diperlukan sebuah metode untuk dapat membantu TripAdvisor agar dapat mengambil informasi serta melakukan analisis terhadap informasi ulasan pelanggan. Analisis yang dapat mengidentifikasi hal spesifik yang diinginkan dari setiap kelompok wisatawan.

Ulasan di TripAdvisor umumnya terdapat beberapa bagian yaitu judul ulasan, tanggal ulasan, rating keseluruhan, teks ulasan, serta *trip type*. Berikut merupakan sebuah contoh ulasan yang mengisi *trip type* dari ulasan.



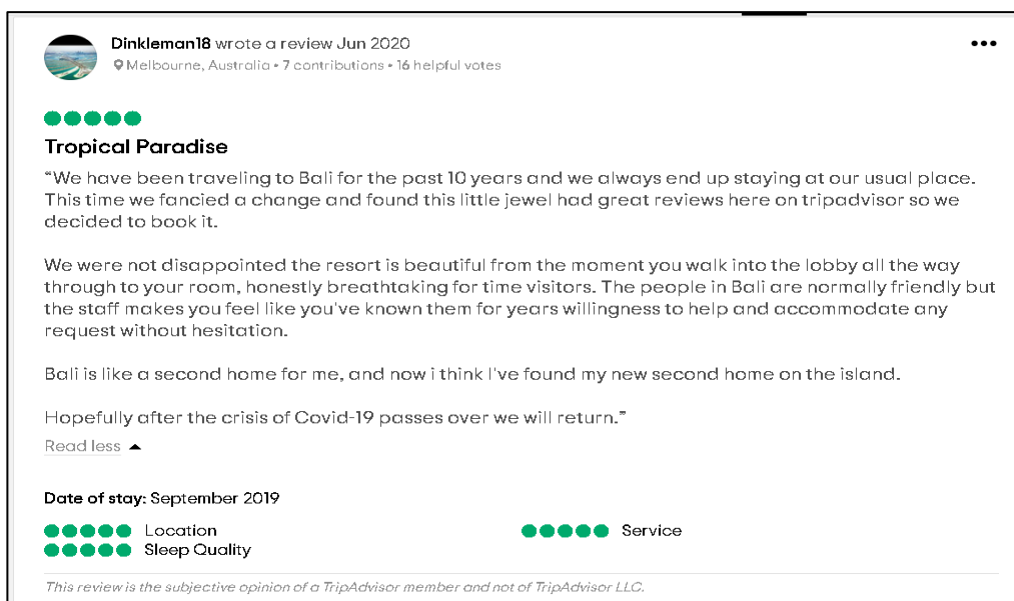
Gambar 1.7. Contoh Ulasan *TripAdvisor* (2)
(Sumber: TripAdvisor.com)

Pada ulasan tersebut dapat dilihat dengan jelas ulasan tersebut mendapatkan *feedback* positif dan dapat digunakan sebagai layanan yang perlu dipertahankan untuk pengunjung yang berkunjung bersama teman-temannya, karena telah ditunjukkan juga oleh *trip type* dari ulasan tersebut. Namun nyatanya banyak konsumen yang tidak mengisi *trip type* yang ada pada sebuah ulasan, sehingga kebutuhan ataupun keluhan yang diberikan oleh konsumen dalam setiap kelompoknya memiliki kemungkinan untuk tidak tersampaikan.

Menurut penelitian yang dilakukan oleh Korfiatis (2012), sebuah ulasan akan bermanfaat dan dapat membantu jika dikemukakan *readability formula*, yang merupakan tingkat sebuah ulasan untuk dapat dipahami atau dibaca (*readability*) dengan baik untuk dapat dilakukan Analisa, maka diperlukan atribut dalam satu

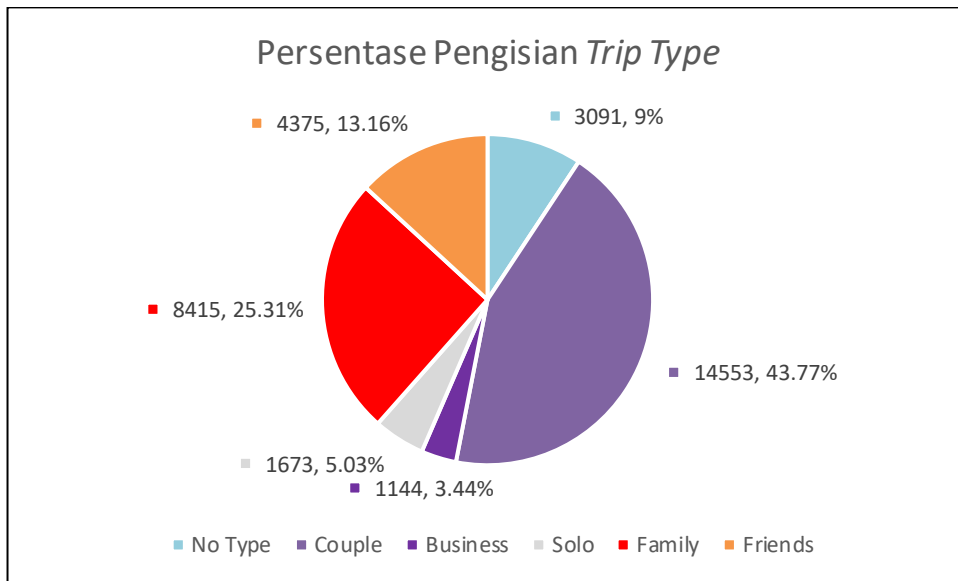
ulasan tertentu. Dalam hal ini maka atribut seperti *trip type* pada ulasan yang diambil dari TripAdvisor akan membuat sebuah ulasan dapat dianalisa dengan baik. Maka dari itu *trip type* yang merupakan pengelompokkan dari setiap kelompok wisatawan dapat bermanfaat sebagai atribut yang berfungsi untuk dapat menganalisa suatu ulasan dengan baik.

Terdapat juga ulasan yang tidak terdapat *trip type* pada *website* TripAdvisor karena dalam mengisi ulasan, pengisian *trip type* bukan sebuah kewajiban untuk mengumpulkan ulasan sehingga menjadi suatu masalah untuk para kelompok wisatawan yang ingin melihat ulasan dari suatu hotel tertentu, serta menjadi sebuah masalah bagi pihak hotel untuk mengklasifikasi kebutuhan dari setiap kelompok wisatawan yang ada. Berikut merupakan contoh ulasan yang tidak memiliki *trip type*.



Gambar I.8. Contoh Ulasan *TripAdvisor* (3)
(Sumber: TripAdvisor.com)

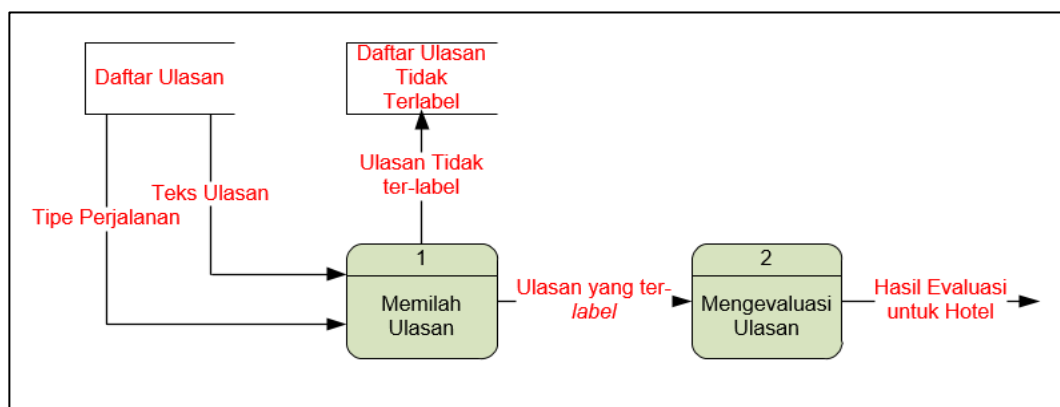
Maka dari itu untuk melakukan prediksi *trip type*, diperlukan sebuah *dataset* yang sudah terlabelkan dari masing-masing *trip type*, agar dapat dilakukan prediksi *trip type* dari ulasan yang digunakan. Dari *dataset* yang telah diambil untuk digunakan dalam penelitian, terdapat banyak ulasan dari hotel-hotel yang tidak mengisi *trip type* dari ulasan. Berikut merupakan statistik dari pengisian *trip type* yang dilakukan dari 20 hotel yang akan digunakan menjadi objek penelitian.



Gambar I.9 Persentase Pengisian *Trip Type*

Pada grafik di atas didapatkan bahwa sebesar 9% dari konsumen yang mengisi hotel dalam 20 ulasan hotel yang digunakan sebagai data awal. Angka tersebut dapat digolongkan kecil jika dilihat dari persentasenya, namun 9% ulasan tersebut mengandung 3091 ulasan yang memiliki karakteristiknya masing-masing serta kebutuhan setiap kelompok wisatawan yang belum dapat dianalisa. Maka dari itu dibutuhkan sebuah metode untuk dapat memprediksi fitur yang ada dalam kelompok ulasan yang belum memiliki *trip type* sehingga kebutuhan setiap kelompok wisatawan dapat dianalisa dengan akurasi yang lebih tepat.

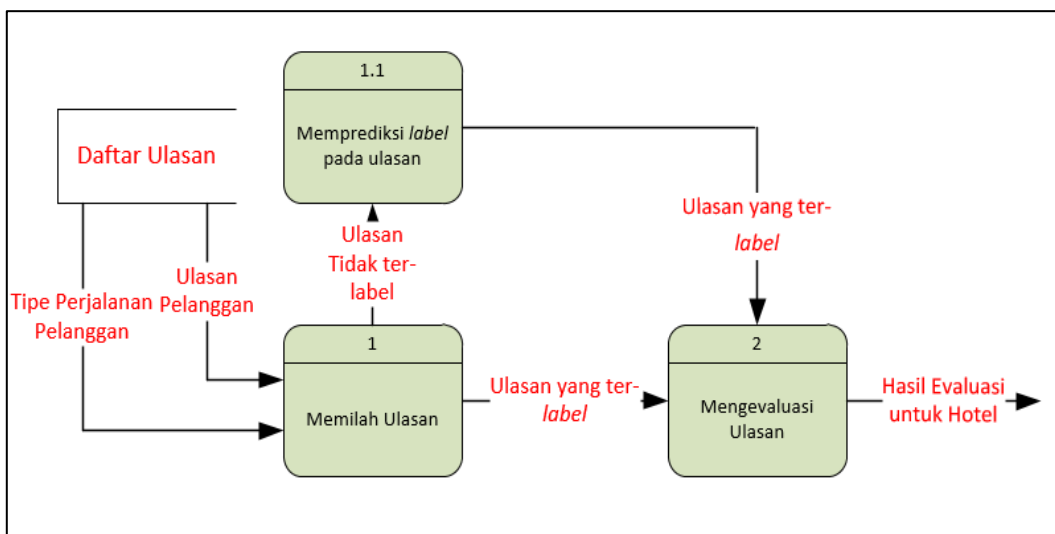
Sistem membaca ulasan dalam hotel pada situs *TripAdvisor* saat ini digambarkan dengan memiliki beberapa tahapan dalam membaca ulasan yang diberikan oleh konsumen kepada hotel. Berikut pada Gambar I.9 merupakan ilustrasi sistem pergerakan informasi terkait proses membaca ulasan saat ini.



Gambar I.10. Ilustrasi Sistem Membaca Ulasan Saat Ini

Ilustrasi proses yang dimiliki yang dilakukan dalam membaca ulasan saat ini memiliki proses dimana ulasan yang ditulis pada TripAdvisor akan dipilah oleh pihak hotel. Ulasan yang didapatkan terdapat informasi ulasan serta tipe perjalanan (*trip type*) yang dimiliki oleh penulis ulasan. Lalu ulasan yang ditulis akan dipilah lagi oleh pihak hotel berdasarkan *trip type* nya agar mengetahui kebutuhan yang diperlukan oleh setiap kelompok wisatawan serta mengetahui kekurangan yang perlu diperbaiki oleh pihak hotel dalam setiap segmen pasar yang dimiliki oleh pihak hotel. Jika sistem membaca ulasan yang dimiliki hotel saat ini seperti yang digambarkan pada ilustrasi tersebut, maka ulasan akan dibaca secara manual dan membutuhkan waktu yang lama jika ulasan memiliki jumlah yang sangat besar dan akan sangat membuang sumber daya seperti waktu, uang dan manusia. Ulasan yang tidak memiliki label (*trip type*) juga akan diprediksi secara manual dan pada akhirnya akan dievaluasi oleh pihak hotel dan digunakan untuk mengevaluasi performansi yang dilakukan hotel terhadap layanan yang diberikan kepada konsumen.

Sistem tersebut dapat diperbaiki jika sebuah model prediksi yang terotomasi dan tidak membutuhkan sumber daya yang besar dapat melakukan pekerjaan memprediksi ulasan. Hal ini dapat menguntungkan bagi pihak hotel dalam mempercepat untuk mendapatkan informasi yang dibutuhkan untuk memperbaiki layanan yang dapat diberikan kepada pelanggan hotel. Berikut pada Gambar I.10 merupakan sistem pergerakan informasi yang dapat digunakan jika telah terdapat sebuah model prediksi *trip type*.



Gambar I.11. Ilustrasi Sistem Membaca Ulasan Dengan Model Prediksi

Dengan gambaran sistem membaca ulasan yang dimiliki saat ini serta diketahuinya terdapat wisatawan tidak mengisi *trip* type yang merupakan sebuah masalah karena terdapat informasi terkait kebutuhan konsumen yang belum dapat dianalisa, maka dibutuhkan sebuah model untuk dapat memprediksi ulasan yang belum memiliki label atau *trip type* tersebut. Dalam ulasan tersebut terdapat *feature-feature* yang dapat merepresentasikan isi sebuah ulasan. *Feature* tersebut sulit untuk dapat mengkarakteristikan ulasan secara langsung, maka sebuah metode *feature-based* untuk digunakan dalam model prediksi merupakan hal yang tepat untuk dapat digunakan dalam model agar prediksi dapat dilakukan dengan baik.

Menurut Beel, Gipp, Langer dan Breitinger (2016), sebanyak 55% dari *paper recommender systems* yang digunakan untuk membantu konsumen atau *user* untuk mendapatkan pelayanan atau kualitas produk yang diinginkan dari setiap konsumen menggunakan *content-based filtering* (CBF). CBF merupakan sebuah teknik dalam *machine learning* yang menggunakan kesamaan dari *features* dalam menentukan sebuah keputusan. Hanya sebanyak 18% menggunakan *collaborative filtering* dan 82% dari sistem rekomendasi menggunakan pendekatan *feature based* yang termasuk dalam CBF. Maka dari itu metode *feature based* seperti *tf-idf* yang menghitung relevansi sebuah kata dalam sebuah dokumen, *Distinguishing Feature Selector* (DFS) yang menentukan *features* yang membedakan dalam sebuah dokumen serta *raw term frequency* yang menghitung jumlah *features* pada sebuah dokumen dapat digunakan. Metode-metode tersebut dapat digunakan karena berbasis *feature* yang merupakan atribut penting dalam melakukan prediksi.

Dalam melakukan hal tersebut, diperlukan juga sebuah *learning algorithm* seperti *k-Nearest Neighbor* atau Naïve Bayes yang dapat melatih sebuah *dataset* sehingga didapatkan output yang diinginkan. Performansi dari model akan diuji menggunakan *classifier-classifier* tersebut. Performansi dari model perlu diuji agar dapat mengetahui seberapa baik kemampuan yang dimiliki oleh model untuk dapat memprediksi sehingga didapatkan gambaran awal model prediksi tersebut saat diaplikasikan dalam masalah nyata untuk memprediksi sebuah ulasan. Dapat diketahui bahwa model tersebut dapat melakukan tugas memprediksi dengan baik atau tidak.

Salah satu output yang penting untuk didapatkan dalam melakukan *training dataset* tersebut merupakan kata kunci, karena kata kunci merupakan fitur sangat penting yang dapat menjadi acuan untuk dapat melakukan klasifikasi teks sehingga dapat diketahui karakteristik dari sebuah data. Dengan menggunakan metode ini, kata-kata kunci yang menjadi fitur yang paling berpengaruh dalam model ini dapat diprediksi untuk setiap *trip type* dari ulasan konsumen. Berdasarkan latar belakang dan identifikasi masalah yang sudah dipaparkan, berikut merupakan rumusan masalah yang didapatkan.

1. Bagaimana model klasifikasi yang menggunakan *feature* berbasis perhitungan *tf-idf*, *raw term frequency* dan DFS?
2. Fitur apa saja yang paling berpengaruh dari ulasan pelanggan terhadap model prediksi yang dikembangkan?
3. Bagaimana performansi akurasi dari model prediksi yang telah dilakukan?

I.3 Pembatasan Masalah dan Asumsi Penelitian

Berdasarkan identifikasi dan rumusan masalah di atas, ditentukan beberapa batasan masalah dan asumsi penelitian agar penelitian lebih terarah dan memiliki fokus yang jelas. Berikut merupakan batasan yang telah ditentukan untuk penelitian yang dilakukan.

1. Ulasan yang digunakan untuk penelitian merupakan ulasan dari 20 hotel yang diambil secara acak dan terletak di kota Bali dan ulasan berasal dari situs TripAdvisor.
2. Hotel yang dipilih untuk digunakan dalam penelitian merupakan hotel yang memiliki jumlah ulasan minimal 1000 ulasan dan berbahasa Inggris.
3. Usulan yang diberikan tidak mencapai tahap implementasi.

Setelah menetapkan batasan masalah penelitian, berikut merupakan asumsi yang digunakan selama penelitian dilakukan.

1. Setiap ulasan yang telah dilabelkan oleh TripAdvisor "*English*" memiliki isi ulasan yang berbahasa Inggris dan tidak terdapat kesalahan eja atau ketik.
2. Tidak terdapat perubahan signifikan pada *dataset* yang digunakan untuk pemodelan dan diasumsikan dapat digunakan

I.4 Tujuan Penelitian

Tujuan penelitian ini merupakan untuk penyelesaian dari masalah yang sudah dipaparkan pada rumusan masalah. Berdasarkan rumusan masalah yang ditentukan, berikut merupakan tujuan dari penelitian yang dilakukan.

1. Membuat model prediksi *trip type* dari ulasan pelanggan dengan menerapkan metode *tf-idf*, DFS dan *raw term frequency*.
2. Mengetahui *feature* yang berpengaruh dari ulasan pelanggan terhadap prediksi yang dilakukan.
3. Mengetahui performansi akurasi dari model prediksi yang telah dikembangkan untuk memprediksi *trip type*.

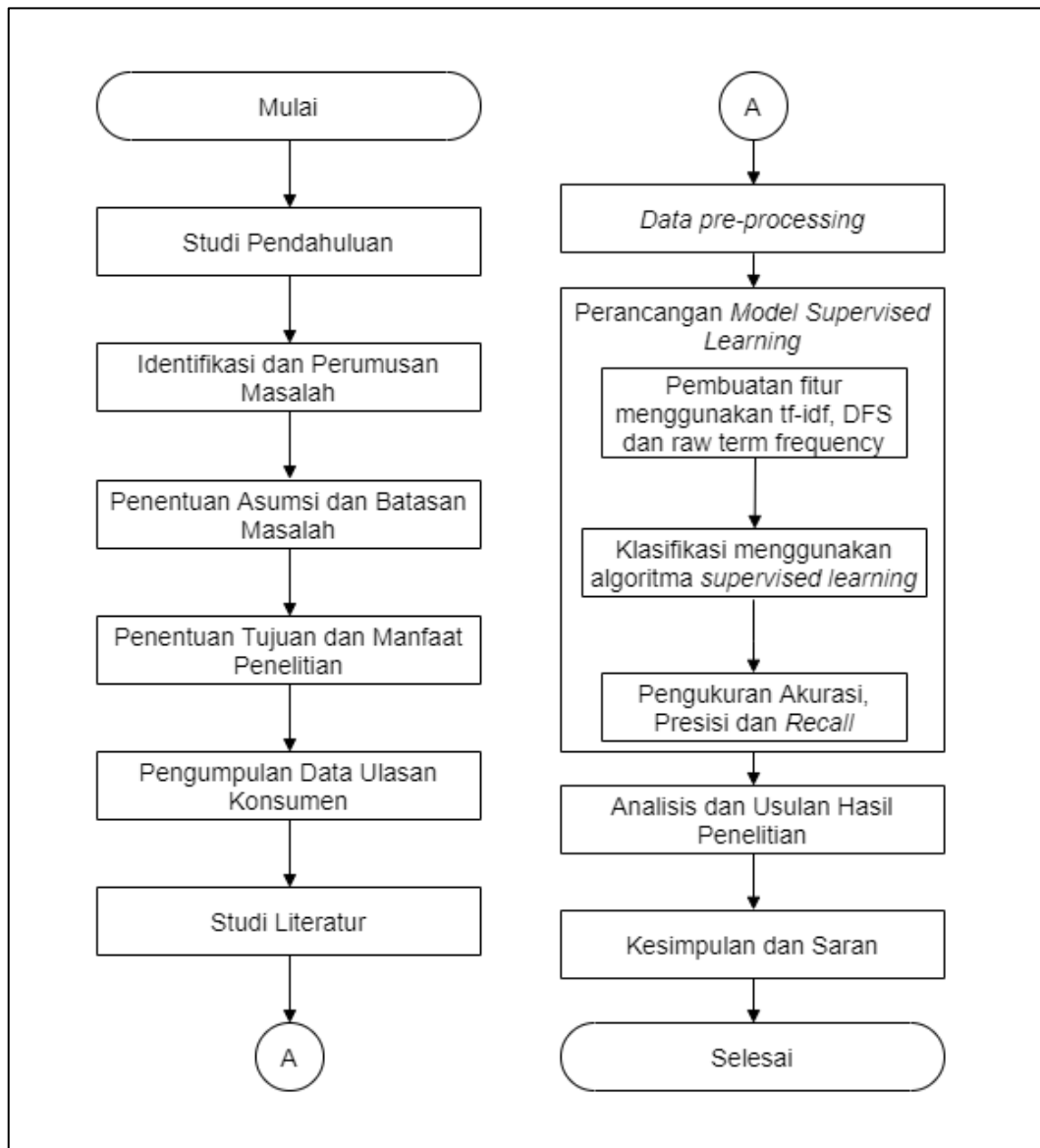
I.5 Manfaat Penelitian

Berdasarkan penelitian yang dilakukan, diharapkan hasil dari penelitian ini mampu memberikan manfaat yang nyata untuk beberapa pihak. Manfaat bagi beberapa pihak tersebut adalah:

1. Manfaat bagi pihak hotel dan konsumen.
Manfaat yang diharapkan merupakan prediksi yang dilakukan mampu membantu untuk mengelompokkan ulasan dengan lebih baik pada setiap kelompok *trip type* agar keinginan dan kebutuhan setiap kelompok wisatawan dapat tersampaikan serta layanan yang diberikan oleh TripAdvisor kualitasnya dapat meningkat.
2. Manfaat bagi pembaca
Pembaca dapat menjadikan hasil penelitian ini sebagai referensi bagi penelitian selanjutnya yang memiliki tujuan atau topik penelitian yang serupa.

I.6 Metodologi Penelitian

Setelah membahas studi literatur yang akan digunakan dalam penelitian, hal berikutnya yang akan dipaparkan merupakan metodologi dari penelitian. Metodologi penelitian digunakan sebagai dasar dalam langkah-langkah pengerjaan yang akan dilakukan selama penelitian. Berikut merupakan *flowchart* dari metodologi penelitian model klasifikasi *feature-based* untuk memprediksi *trip type* pada ulasan *online* hotel di Bali.



Gambar 1.12 Metodologi Penelitian

Flowchart metodologi penelitian memberikan sebuah gambaran visual terkait langkah penelitian yang akan dilakukan. Berikut merupakan penjelasan dari masing-masing bagian metodologi penelitian.

1. Studi Literatur

Pada tahap ini, dilakukan studi awal terhadap objek, dilakukan juga pemahaman dari konsep yang diperlukan dalam penelitian serta pemilihan metode pemodelan yang digunakan untuk melakukan prediksi *trip type*. Dalam tahap ini juga dilakukan pengumpulan referensi yang

dapat membantu penelitian. Literatur yang digunakan berasal dari buku, jurnal penelitian terkait dan internet.

2. Identifikasi dan Perumusan Masalah

Setelah melakukan studi literatur, masalah yang ditinjau dalam penelitian akan diuraikan secara lebih dalam berdasarkan informasi dari studi yang dilakukan. Identifikasi dilakukan pada masalah terkait *trip type* yang akan dilakukan prediksi. Masalah tersebut lalu dirumuskan dalam rumusan masalah.

3. Penentuan Asumsi dan Batasan Masalah

Lalu akan ditentukan asumsi yang diperlukan dalam penelitian. Setelah menentukan asumsi penelitian, batasan penelitian juga ditentukan. Asumsi dan batasan ditentukan agar penelitian lebih memiliki fokus yang jelas sehingga penelitian tidak memiliki lingkup yang terlalu luas.

4. Penentuan Tujuan dan Manfaat Penelitian

Tahap penentuan tujuan dan manfaat penelitian dilakukan agar penelitian yang dilakukan lebih terarah dan dapat menjawab masalah yang dirumuskan, sehingga mampu untuk memberi manfaat bagi pihak TripAdvisor, pihak hotel, setiap kelompok wisatawan yang menggunakan layanan TripAdvisor serta pembaca dari penelitian ini.

5. Pengumpulan Data Ulasan Konsumen

Pada tahap ini akan dilakukan pengumpulan data ulasan *online* konsumen dari situs TripAdvisor. Ulasan *online* yang diambil merupakan ulasan hotel di Bali. Hotel yang dipilih merupakan hotel yang ada di Bali dengan jumlah ulasan minimal 1000 dan ulasan berbahasa Inggris. Data dikumpulkan menggunakan program *PyCharm Edu* dengan *package library* Selenium yang menggunakan bahasa pemrograman *Python*.

6. Data *pre-processing*

Pada tahap ini, dilakukan proses pembersihan data yang telah diambil dari ulasan *online*. Tahap ini dilakukan agar data yang digunakan menjadi lebih terstruktur, menghilangkan *noise* dan juga dapat dinormalisasi sehingga data dapat dipakai sepenuhnya untuk tahap penelitian berikutnya. Tahap ini terbagi menjadi beberapa tahap penting seperti *tokenization*, menghilangkan *stop words*, dan *lemmatization*.

7. Perancangan *Model Supervised Learning*

Lalu perancangan model *supervised learning* untuk melakukan prediksi *trip type* dilakukan. Dalam tahap ini terdapat beberapa sub tahapn yaitu diawali dengan pembobotan term menggunakan *tf-idf*, DFS, dan *raw term frequency*, lalu melakukan klasifikasi dengan algoritma *feature-based* seperti *Naïve Bayes*, *Decision Tree* atau *k-Nearest Neighbor*. Setelah itu dilakukan juga verifikasi dan validasi Pada tahap ini, dilakukan verifikasi dan validasi terhadap program yang dirancang untuk memastikan bahwa program yang dirancang telah sesuai dengan algoritma yang telah dirancang. Sub-tahap terakhir melakukan pengukuran akurasi, presisi dan *recall* dari model yang telah dilakukan.

8. Usulan dan Analisis Hasil Penelitian

Setelah melakukan perancangan dan menjalankan model, dilakukan analisis terhadap hasil dari penelitian klasifikasi *feature-based* untuk memprediksi *trip type* dari ulasan *online*. Analisis ini dilakukan dengan menggunakan nilai akurasi dari klasifikasi yang dilakukan dengan menggunakan algoritma *Naïve Bayes*.

9. Kesimpulan dan Saran

Tahap ini merupakan tahap akhir dari penelitian yang telah dilakukan. Kesimpulan yang diberikan berisi rangkuman dari hasil penelitian yang telah dilakukan. Selain kesimpulan, diberikan juga saran untuk setiap kelompok wisatawan, pihak hotel, pihak *TripAdvisor*, dan penelitian selanjutnya.

I.7 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini akan menjadi panduan dalam melakukan penulisan penelitian ini. Selain itu, sistematika penulisan juga dapat mempermudah pemahaman keseluruhan dari penelitian ini. Berikut merupakan sistematika penulisan dari penelitian ini

BAB I PENDAHULUAN

Pada Bab I Pendahuluan ini akan dijelaskan mengenai latar belakang yang mendorong pelaksanaan penelitian, identifikasi dan rumusan masalah,

batasan dan asumsi penelitian, tujuan penelitian, manfaat penelitian, serta metodologi sistematis yang akan digunakan dalam penelitian ini

BAB II TINJAUAN PUSTAKA

Pada bab ini akan dibahas mengenai tinjauan Pustaka, teori serta studi literatur yang akan digunakan terkait dengan penelitian yang dilakukan. Mulai dari teori mengenai *machine learning*, *supervised learning*, *text preprocessing*, *tf-idf*, *Naïve Bayes* dan teori relevan lainnya. Seluruh pengertian dan teori tersebut akan digunakan untuk merancang model prediksi dalam bab selanjutnya.

BAB III PENGOLAHAN DATA

Pada Bab III ini akan berisi mengenai proses yang dimulai dari pengambilan data ulasan mentah dari TripAdvisor, lalu dilakukannya *text preprocessing* hingga pemilihan *feature* yang akan digunakan melalui metode-metode seperti *tf-idf* serta DFS hingga tahap penyelesaian model prediksi dengan melihat performansi dari model yang telah dibuat melalui *classifier-classifier* seperti *Naïve Bayes*, *Decision Tree* dan *k-Nearest Neighbors* sebagai dasar dari model prediksi yang akan dikembangkan. Seluruh pemrograman dalam menggunakan bahasa pemrograman *python* yang dilakukan di dalam lingkungan *pycharm Edu* akan dipaparkan dalam bab ini.

BAB IV ANALISIS

Pada Bab IV ini akan dilakukan analisis terhadap proses penelitian yang dilakukan. Analisis yang dilakukan meliputi hasil dari penerapan metode *tf-idf*, *raw term frequency* serta DFS dalam membangun model, analisis fitur yang terpilih serta analisis performansi model prediksi *trip type*.

BAB V KESIMPULAN DAN SARAN

Pada Bab V ini akan dipaparkan kesimpulan penelitian yang akan menjawab tujuan dari penelitian ini. Akan disampaikan juga saran untuk penelitian selanjutnya yang memiliki topik serupa.

