

**SISTEM REKOMENDASI HOTEL DENGAN PENDEKATAN
*CONTENT-BASED FILTERING***

(Studi Kasus: Hotel di Yogyakarta pada *Website Nusatrip.com*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program Studi
Statistika



Disusun Oleh:

Cheryl Ayu Melyani

18611001

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2022**

HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR

Judul : Sistem Rekomendasi Hotel dengan Pendekatan
Content-Based Filtering
(Studi Kasus: Hotel di Yogyakarta pada Website
Nusatrip.com)

Nama Mahasiswa : Cheryl Ayu Melyani

NIM : 18611001

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN

Yogyakarta, 21 Januari 2022

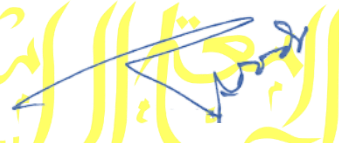
Mengetahui,

Pembimbing

Ketua Prodi Statistika



(Ayundyah Kesumawarti, S.Si., M.Si)



(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN

TUGAS AKHIR

SISTEM REKOMENDASI HOTEL DENGAN PENDEKATAN CONTENT- BASED FILTERING

(Studi Kasus : Hotel di Yogyakarta pada Website Nusatrip.com)

Nama Mahasiswa : Cheryl Ayu Melyani

NIM : 18611001

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 18 FEBRUARI 2022:

Nama Penguji

Tanda Tangan

1. Raden Bagus Fajriya Hakim, Dr., S.Si., M.Si.

2. Arum Handini Primandari, S.Pd.Si., M.Sc.

3. Ayundyah Kesumawati, S.Si., M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu 'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah SWT, Tuhan semesta alam atas rahmat, taufik, dan hidayah-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan laporan tugas akhir ini. Tidak lupa sholawat serta salam tercurahkan kepada Baginda Nabi Muhammad SAW. Laporan Tugas Akhir yang berjudul “*Sistem Rekomendasi Hotel dengan Pendekatan Content-Based Filtering*” ini disusun untuk memenuhi persyaratan dalam menyelesaikan jenjang Strata Satu (S1) di Program Studi Statistika, Universitas Islam Indonesia.

Adapun dalam proses penyusunan Tugas Akhir ini tidak terlepas dari berbagai pihak yang memberikan bantuan, bimbingan, semangat, dan dukungan. Maka dari itu, pada kesempatan ini penulis ingin menyampaikan rasa terimakasih kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Bapak Dr. Edy Widodo, M.Si., selaku Ketua Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia beserta jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing Tugas Akhir yang memberikan bimbingan serta dukungan penuh dalam proses penyusunan Tugas Akhir ini dari awal hingga selesai.
4. Bapak Achmad Fauzan, S.Pd., M.Si. selaku Dosen Pembimbing Akademik yang memberikan dukungan penuh dari awal masuk kuliah hingga selesai.
5. Bapak Heriyanto dan Triutami selaku orang tua dan Cheka Putri Maharani kakak saya tercinta yang tak henti-hentinya selalu mendukung, dan memberikan semangat dalam keadaan apapun.

6. Sahabat serta teman-teman seperjuangan, Bella Destia, Ayu Pramesti, Dhea Indah Syafira Said, Firda Rahmatul Ummah, dan Andini Nur Amalina yang selalu memberikan semangat untuk penulis menyelesaikan Tugas Akhir ini.
7. Sahabat dan teman-teman, Galih Eka Pramudya, Archie, Alhaq, dan anggota dari Keluarga Bebek serta Suditara yang senantiasa memberikan semangat dan dukungan dalam menyelesaikan Tugas Akhir.
8. Teman-teman seperjuangan Statistika Angkatan 2018 yang senantiasa berjuang untuk memperoleh gelar S.Stat.
9. Semua pihak yang tidak bisa disebutkan satu persatu, terima kasih banyak untuk segala dukungannya

Penulis menyadari Tugas Akhir ini masih terdapat kekurangan di dalamnya dan jauh dari kata sempurna. Dengan demikian, penulis mengharapkan kritik dan saran yang membangun agar Tugas Akhir ini dapat bermanfaat bagi penulis maupun pembaca. Apabila pada Tugas Akhir ini memiliki banyak kesalahan, penulis meminta maaf yang sebesar-besarnya.

Wassalamualaikum Warahmatullahi Wabarakatuh

Yogyakarta, 21 Januari 2022



Cheryl Ayu Melyani

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
PERNYATAAN	xii
INTISARI.....	xiii
ABSTRACT	xiv
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB 2 TINJAUAN PUSTAKA.....	6
BAB 3 LANDASAN TEORI	12
3.1 <i>Online Travel Agent</i>	12
3.2 Nusatrip	13
3.3 Hotel.....	14
3.4 <i>Text Mining</i>	16
3.4 <i>Term Frequency-Invers Document Frequency (TF-IDF)</i>	17
3.4.1 Perhitungan <i>Term-Frequency (TF)</i>	18
3.4.2 Perhitungan <i>Invers Document Frequency (IDF)</i>	19

3.4.3	Perhitungan <i>Term Frequency-Invers Document Frequency</i> (TF-IDF)	19
3.5	<i>Cosine Similarity</i>	20
3.6	<i>Recommendation System</i>	21
3.7	Teknik <i>Recommendation System</i>	22
3.7.1	<i>Personalized Recommendation System</i>	23
3.7.2	<i>Non-Personalized Recommendation System</i>	27
3.8	<i>Flask</i>	27
3.9	Heroku	28
BAB 4	METODOLOGI PENELITIAN	29
4.1	Populasi dan Sampel Penelitian	29
4.2	Jenis dan Sumber Data	29
4.3	Variabel Penelitian	29
4.4	Metode Analisis Data	30
4.5	Tahapan Analisis Penelitian	31
BAB 5	HASIL DAN PEMBAHASAN	32
5.1	Analisis Deskriptif	32
5.2	<i>Preprocessing Data</i>	33
5.2.1	<i>Case Folding</i>	34
5.2.2	<i>Remove Punctuation and Remove Whitespace</i>	36
5.2.3	<i>Remove Stopword</i>	38
5.2.4	<i>Lemmatization</i>	39
5.2.5	<i>Tokenizing</i>	41
5.3	Pembobotan Kata	44
5.4	Hasil Rekomendasi Menggunakan <i>Cosine Similarity</i>	47
5.5	Kaji Ulang Rekomendasi Hotel	51
5.6	Pengaplikasian Sistem Rekomendasi pada <i>Website</i>	55
BAB 6	PENUTUP	57
6.1	Kesimpulan	57

6.2	Saran	57
	DAFTAR PUSTAKA	59
	LAMPIRAN	63

DAFTAR TABEL

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu	9
Tabel 4.1 Definisi Operasional Variabel	29
Tabel 5.1 <i>Dokumen Awal</i>	33
Tabel 5.2 Hasil <i>Case Folding</i>	34
Tabel 5.3 Hasil <i>Remove Punctuation</i> dan <i>Remove Whitespace</i>	36
Tabel 5.4 Hasil <i>Remove Stopword</i>	38
Tabel 5.5 Hasil <i>Lemmatization</i>	40
Tabel 5.6 Hasil <i>Tokenizing</i>	41
Tabel 5.7 Hasil Perhitungan TF pada Sampel Pilihan	44
Tabel 5.8 Hasil Perhitungan TF-IDF pada Sampel Pilihan	45
Tabel 5.9 10 Dokumen yang Mirip dengan Dokumen ke-132	47
Tabel 5.10 Hasil Rekomendasi Hotel.....	48
Tabel 5.11 Keseluruhan Deskripsi dari Dokumen ke-132 dan Dokumen ke-170 Setelah <i>Preprocessing Data</i>	49
Tabel 5.12 Perbandingan deskripsi Good Karma Yogyakarta- Hostel dan Happy Buddha Yogyakarta – Hostel	50
Tabel 5.13 Hasil Rekomendasi Hotel Eastparc berdasarkan Deskripsi	51
Tabel 5.14 Hasil Rekomendasi Hotel Eastparc dan Nilai <i>Cosine Similarity</i> berdasarkan Fasilitas serta <i>Nearby attraction</i>	52
Tabel 5.14 Hasil Rekomendasi Hotel Eastparc dan Info Kemiripannya.....	54

DAFTAR GAMBAR

Gambar 1.1 Grafik Perkembangan Minat <i>Staycation</i>	1
Gambar 3.1 Logo <i>Nusatrip.com</i>	14
Gambar 3.2 Ilustrasi <i>Recommendation System</i>	22
Gambar 3.3 Tipe-tipe <i>Recommendation System</i>	23
Gambar 3.4 Ilustrasi <i>Content-Based Filtering</i>	24
Gambar 4.1 Tahapan Analisis Penelitian	31
Gambar 5.1 Jumlah Hotel di Yogyakarta berdasarkan Bintang	32
Gambar 5.2 Frekuensi 10 Kata Tertinggi	43
Gambar 5.3 Diagram Alir Algoritma Tambahan	54
Gambar 5.4 Halaman Pencarian.....	55
Gambar 5.5 Halaman Hasil Rekomendasi.....	56

DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i>	63
Lampiran 2 Perhitungan Manual TF-IDF dan <i>Cosine Similarity</i>	63
Lampiran 3 <i>Script Python</i> Sistem Rekomendasi	63
Lampiran 4 <i>Script Flask</i> dan <i>HTML</i>	63

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 21 Januari 2022



Cheryl Ayu Melyani

INTISARI

SISTEM REKOMENDASI HOTEL DENGAN PENDEKATAN CONTENT-BASED FILTERING

(Studi Kasus: Hotel di Yogyakarta pada *Website Nusatrip.com*)

Cheryl Ayu Melyani

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Meningkatnya pandemi Covid-19 membuat aktivitas masyarakat menjadi terhambat seringkali memunculkan stress jika harus berada dirumah secara terus menerus. Hal ini menyebabkan meningkatnya tren *staycation* atau kegiatan berlibur di kota sendiri dengan menyewa sebuah hotel. Teknologi penyewaan hotel mulai dialihkan dengan adanya OTA (*Online Travel Agent*). Adanya berbagai macam hotel dengan berbagai macam fasilitas yang disuguhkan membuat masyarakat sering merasa kebingungan dalam memilih hotel yang akan ditematinya. Untuk membantu mengatasi hal tersebut, peneliti mencoba membuat sebuah *recommendation system* untuk membantu calon penghuni hotel dalam memilih hotel sesuai dengan pilihannya. Selain itu juga dapat membantu perusahaan dalam meningkatkan pemesanan kamar hotel melalui *website*-nya. Dalam penelitian ini, peneliti akan membangun sebuah *recommendation system* hotel di Yogyakarta pada salah satu OTA di Indonesia menggunakan *Content-Based Filtering Methods*, pembobotan data teks menggunakan *Term Frequency-Invers Document Frequency (TF-IDF) Methods*, dan mengukur kemiripan dokumen menggunakan *Cosine Similarity Methods*. Berdasarkan hasil rekomendasi hotel Good Karma Yogyakarta sebagai contoh pengujian, didapatkan 10 hotel yang mirip yaitu Happy Buddha Yogyakarta – Hostel, Nextdoor Homestay, Hotel Puspita, OYO 426 Hotel Gading Resto, Omah Jegog Homestay, Prawirotaman Homestay, RedDoorz near Prawirotaman, Ayodhya Garden Hostel Yogyakarta by HOM, Bringin House Yogyakarta, dan House 24 Yogyakarta dengan nilai *cosine similarity* secara berturut-turut sebesar 0.956666513, 0.946570717, 0.917459394, 0.912534877, 0.886439718, 0.88221982, 0.881775275, 0.875845794, 0.872030219, dan 0.871514859.

Kata Kunci: *Staycation, Hotel, Recommendation System, Content-Based Filtering.*

ABSTRACT

HOTEL RECOMMENDATION SYSTEM WITH CONTENT-BASED FILTERING APPROCH

(Case Study: Hotel in Yogyakarta on Nusatrip.com Website)

Cheryl Ayu Melyani

Department of Statistics, Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia

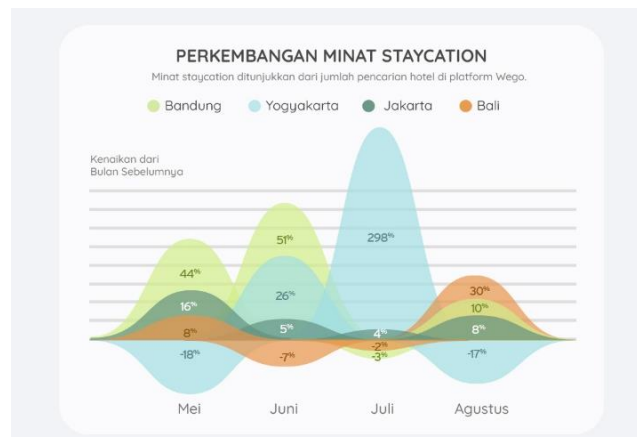
The increasing of Covid-19 pandemic has hampered people's activities, often causing stress they are only stay at home continuously. This has led to an increasing trend of staycations or holiday activities in the city itself by renting a hotel. Hotel rental technology has begun to be transferred with the existence of OTA (Online Travel Agent). The existence of various kinds of hotels with various kinds of facilities that makes people feel confused in choosing which hotel to occupy. To help overcome this, the researchers tried to create a recommendation system to help prospective hotel residents choose the hotel according to their choice. In addition, it can also assist companies in increasing hotel room reservations through its website. In this study, researchers will build a hotel recommendation system in Yogyakarta at one of the OTAs in Indonesia using Content-Based Filtering Methods, weighting text data using Term Frequency-Inverse Document Frequency (TF-IDF) Methods and measuring document similarity using Cosine Similarity Methods. Based on the results of the Good Karma Yogyakarta hotel recommendations as a test example, 10 similar hotels were obtained, namely Happy Buddha Yogyakarta – Hostel, Nextdoor Homestay, Hotel Puspita, OYO 426 Hotel Gading Resto, Omah Jegog Homestay, Prawirotaman Homestay, RedDoorz near Prawirotaman, Ayodhya Garden Hostel Yogyakarta by HOM, Bringin House Yogyakarta, and House 24 Yogyakarta with cosine similarity values 0.956666513, 0.946570717, 0.917459394, 0.912534877, 0.886439718, 0.88221982, 0.881775275, 0.875845794, 0.872030219, and 0.871514859.

Keywords: *Staycation, Hotel, Recommendation System, Content-Based Filtering.*

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Pandemi Covid-19 yang telah berlangsung kurang lebih selama 2 tahun ini memberikan dampak ke berbagai sektor. Aktivitas yang biasa masyarakat lakukan menjadi terhambat. Satu sisi, masyarakat tidak ingin terpapar virus *Corona* dan memutuskan untuk melakukan segala aktivitas di rumah masing-masing. Pada sisi yang lain, mereka sadar bahwa jika terus menerus di rumah saja maka akan memunculkan stress. Berdasarkan data *Google Trends*, grafik dari tren *staycation* cenderung naik dari awal tahun 2020 hingga sekarang. Menurut *HowStuffWorks*, *staycation* merupakan kombinasi antara berlibur dan tetap ada di rumah (Layton, 2009). Namun bagi sebagian orang, *staycation* adalah kegiatan berlibur di kota sendiri dengan menyewa sebuah hotel untuk menghilangkan kejenuhan akibat terlalu banyak kegiatan di rumah. Survey yang dilakukan oleh *Wego.co.id* mengenai tren *staycation* menunjukkan bahwa Yogyakarta menjadi salah satu destinasi yang diminati sebagai kota tujuan untuk *staycation*. Hal ini ditunjukkan dari jumlah pencarian hotel pada Mei-Agustus 2020 di *platform Wego* (Ahsan, 2020).



Gambar 1.1 Grafik Perkembangan Minat *Staycation*

Sumber: <https://travel.wego.com/berita/staycation-kala-pandemi-bandung-dan-yogyakarta-jadi-incaran/>

Berdasarkan Peraturan Pemerintah Republik Indonesia No.65 Tahun 2021, hotel adalah bangunan yang secara khusus diperuntukkan untuk menginap ataupun beristirahat dan memperoleh pelayanan serta fasilitas dengan dipungut biaya. Hotel merupakan salah satu hunian sementara yang sangat berkembang pesat di Yogyakarta. Keberadaan hotel di Yogyakarta sudah tersebar hingga pinggiran kota. Menurut Badan Pusat Statistik, pada tahun 2020 terdapat 2.123 hotel di Yogyakarta yang terdiri dari hotel bintang dan hotel non bintang (BPS, 2020). Namun saat ini muncul juga berbagai macam *homestay* dan *kost exclusive* yang disewakan menjadi hunian sementara.

Di era digital ini, perkembangan teknologi informasi telah berkembang pesat dimana muncul berbagai inovasi penawaran suatu produk atau jasa. Salah satu inovasi yang mengalami evolusi dari konvensional menjadi moderen ada pada sektor pariwisata (Sandi Wachyuni & Wiweka, 2020). Munculnya *Online Travel Agent* (OTA) yang merupakan inovasi pada sektor pariwisata, mampu menggantikan kegiatan pemesanan hotel dimana yang sebelumnya harus memesan hotel secara langsung, kini dapat dilakukan secara online menggunakan OTA yang berbasis *website* (Hendriyati, 2019). Kecenderungan masyarakat jaman sekarang, dimana hampir semua kegiatan dilakukan dengan mengakses internet membuat bisnis OTA ini semakin pesat. Berbagai macam OTA di Indonesia antara lain *Traveloka*, *Booking.com*, *PegiPegi*, *Tiket.com*, *Agoda*, *Nusatrip.com*, *Mister Aladin*, *Hotelmurah.com* dan masih banyak lagi.

Dengan adanya berbagai macam hotel berbintang maupun non bintang yang ada di Yogyakarta, tentunya masyarakat yang ingin melakukan *staycation* akan kebingungan dalam memilih hotel sesuai dengan kriteria maupun fasilitas yang diminati. OTA menyediakan macam-macam pilihan hotel yang dapat disesuaikan dengan kebutuhan dan keinginan masyarakat. Segala informasi mengenai lokasi,

fasilitas, dan keunggulan hotel yang meliputi penjelasan mengenai kamar, restoran, kolam renang, area parkir, dan lain-lain dijelaskan dalam kolom deskripsi pada suatu *website* OTA. Untuk membantu masyarakat dalam pemilihan hotel, peneliti berupaya untuk membuat sebuah *recommendation system* pemilihan hotel bagi masyarakat yang akan melakukan *staycation* sehingga memberikan kemudahan dalam menentukan hotel pilihannya. Adanya *recommendation system* dapat meningkatkan penjualan suatu produk dan bisa menumbuhkan loyalitas pembeli melalui *website* dari perusahaan tersebut (C. S. D. Prasetya, 2017). OTA pemesanan hotel berbasis *website* yang digunakan pada penelitian ini adalah *Nusatrip.com*. Nusatrip merupakan agen *travel online* pertama di Indonesia yang menawarkan berbagai akomodasi, mulai dari tiket maskapai domestik maupun internasional serta menawarkan berbagai macam hotel di seluruh dunia. Pada *website* Nusatrip belum memiliki sistem rekomendasi, sehingga penelitian ini diharapkan dapat menawarkan *recommendation system* untuk *website Nusatrip.com* itu sendiri.

Recommendation system merupakan perangkat lunak dan suatu teknik yang memberikan saran sebuah *item* (produk atau jasa) yang menarik bagi pengguna dan ditujukan untuk mendukung penggunaannya dalam berbagai proses pengambilan keputusan. Sistem ini menjadi alat penemuan informasi paling kuat dan populer di suatu *website* (Ricci et al., 2015). Beberapa tipe dari klasifikasi *recommendation system* antara lain adalah *content-based*, *collaborative-based*, dan *hybrid-based*.

Content Based Filtering merupakan tipe dari sistem rekomendasi dengan konten atau deskripsi sebagai dasar dari algoritma ini. Metode ini akan memperoleh konten dalam *item* kemudian membandingkannya dengan konten *item* yang sebelumnya dicari, selanjutnya akan merekomendasikan *item* yang memiliki konten atau deskripsi serupa dengan apa yang *user* cari (Larasati & Februriyanti, 2021). Biasanya metode tersebut dipergunakan untuk merekomendasikan berita, artikel, ataupun situs *website*. Dalam penelitian ini, peneliti akan membangun sebuah *recommendation system* hotel di Yogyakarta pada salah satu OTA di Indonesia

menggunakan *Content-Based Filtering Methods*, pembobotan data teks menggunakan *Term Frequency-Invers Document Frequency (TF-IDF) Methods*, dan mengukur kemiripan dokumen menggunakan *Cosine Similarity Methods*. Adanya *recommendation system* yang terbentuk, diharapkan dapat mempermudah seseorang dalam pemilihan hotel sesuai dengan hotel pilihannya.

1.2 Rumusan Masalah

Berdasarkan pejabaran dari latar belakang, berikut merupakan identifikasi beberapa rumusan masalah pada penelitian ini:

1. Bagaimana hasil kemiripan masing-masing dokumen menggunakan *Cosine Similarity Methods*?
2. Bagaimana alur memberikan rekomendasi hotel di Yogyakarta dengan *Cosine Similarity Methods* dan *Content Based Filtering Methods*?
3. Bagaimana hasil rekomendasi hotel di Yogyakarta menggunakan *Content-Based Filtering Methods*?
4. Bagaimana langkah membangun aplikasi *website* dari sistem rekomendasi berdasarkan model yang diperoleh.

1.3 Batasan Masalah

Agar menghindari perluasan pokok bahasan, berikut merupakan beberapa batasan masalah yang digunakan dalam penelitian ini:

1. Data yang digunakan adalah data deskripsi hotel, fasilitas hotel, dan *nearby attraction* hotel di Yogyakarta dan diperoleh dari *website* Nusatrip.com pada tanggal 16 November 2021.
2. Alat yang digunakan untuk mengambil data adalah *WebScrapper*. *Microsoft Excel 2016* digunakan untuk menyimpan data hasil *scrapping*. *Anaconda versi 4.8.3* digunakan untuk menganalisis data menggunakan *TF-IDF*

Methods dan *Cosine Similarity Methods*. *Flask Python* dan *Heroku* digunakan untuk membuat *interface* dari sistem rekomendasi yang dihasilkan.

3. Rekomendasi yang diberikan berupa Top 10 hotel yang paling mirip dengan hotel yang dipilih.

1.4 Tujuan Penelitian

Berdasarkan pejabaran dari rumusan masalah, berikut merupakan tujuan dari penelitian ini:

1. Mengetahui hasil kemiripan masing-masing dokumen menggunakan *Cosine Similarity Methods*.
2. Mengetahui alur memberikan rekomendasi hotel di Yogyakarta dengan *Cosine Similarity Methods* dan *Content Based Filtering Methods*.
3. Mengetahui hasil rekomendasi hotel di Yogyakarta menggunakan *Content-Based Filtering Methods*.
4. Mengetahui langkah membangun aplikasi *website* dari sistem rekomendasi berdasarkan model yang diperoleh.

1.5 Manfaat Penelitian

Berikut merupakan manfaat yang dapat diambil dari penelitian ini:

1. Menambah wawasan dan pengetahuan akan *Recommendation System*.
2. Hasil penelitian ini dapat membantu wisatawan di Yogyakarta dalam memilih hotel yang akan dihuni.
3. *Recommendation System* yang dihasilkan dapat membantu perusahaan dalam meningkatkan pemesanan kamar hotel melalui *website*-nya.
4. Diharapkan dapat menjadi referensi untuk penelitian selanjutnya maupun penelitian yang berhubungan dengan *Recommendation System*.

BAB 2 TINJAUAN PUSTAKA

Penelitian terdahulu dipakai sebagai tolak ukur serta referensi peneliti dalam menuliskan penelitian serta untuk memperdalam teori yang akan digunakan dalam penelitian ini. Selain itu, penelitian terdahulu digunakan untuk menjadi bahan perbandingan. Berikut adalah penelitian terkait *recommendation system* yang peneliti pakai untuk menjadi acuan dalam penelitian ini.

Penelitian mengenai *recommendation system* yang pertama dengan judul “*Research Supervisor Recommendation System Based on Topic Conformity*”. Tujuan dari penelitian ini adalah memudahkan mahasiswa dalam menentukan dosen pembimbing sesuai dengan topik penelitian yang dilakukan. Dalam penelitian ini dihasilkan suatu sistem yang memberikan rekomendasi dosen pembimbing berdasarkan kesamaan dengan jurnal penelitian dosen pembimbing yang bersangkutan. Data yang digunakan adalah judul dan abstrak penelitian. Dalam menganalisis data, digunakan aplikasi dari *text mining*. Penelitian ini menerapkan pembobotan kata dengan metode TF-IDF dan perhitungan kemiripan antara dua objek dengan metode *Cosine Similarity*. Berdasarkan hasil analisis, didapatkan nilai akurasi perbandingan hasil rekomendasi dari sistem dengan data aktual sebesar 75% (Rismanto et al., 2020).

Penelitian tentang *recommendation system* yang kedua yang berjudul “Sistem Rekomendasi Buku Menggunakan *Weighted Tree Similarity* dan *Content Based Filtering*”. Penelitian ini menghasilkan sistem rekomendasi buku dengan melihat kemiripan sebuah item dengan item yang lain. Metode yang digunakan dalam pembuatan sistem ini adalah *Content Based Filtering*. Untuk mencari nilai kemiripan antar item, penelitian ini menggunakan metode *Weighted Tree Similarity*. Tujuan dari penelitian ini adalah memudahkan pengunjung Perpustakaan Daerah Provinsi Kalimantan Selatan dalam mencari buku yang memiliki persamaan dengan buku yang

sebelumnya dipilih ketika buku-buku tersebut tidak ada atau telah dipinjam orang lain. Berdasarkan sistem rekomendasi dari pengujian yang dilakukan, didapatkan nilai *precision* sebesar 88% (Alkaff et al., 2020).

Penelitian tentang *recommendation system* yang berikutnya dengan judul “*Recommendation System with Content-Based Filtering Method for Culinary Tourism in Mangan Application*”. Penelitian ini dilakukan karena informasi mengenai kuliner belum tentu cukup membantu wisatawan dalam menentukan kuliner yang mereka inginkan. Sistem rekomendasi yang dihasilkan memuat gambar, nama, dan jarak dari masing-masing restoran sehingga dapat memudahkan pengguna dalam memilih restoran tertentu. Metode yang diterapkan dalam penelitian ini adalah metode *Content-Based Filtering*. Dalam menghitung kesamaan objek, digunakan metode *Cosine Similarity* berdasarkan item profil yang terbentuk dari fitur konten restoran. Dari hasil penelitian menunjukkan bahwa mempunyai rata-rata *precision* sebesar 0,8915 dan rata-rata akurasi sebesar 0,58118. Rendahnya nilai akurasi ini disebabkan oleh kesalahan sistematik karena pengaruh jumlah data yang diujikan terlalu sedikit sehingga pengujian menjadi kurang maksimal dan kurang akurat (Mondi et al., 2019).

Penelitian tentang *recommendation system* selanjutnya berjudul “*Recommendation System for Thesis Topic Using Content-based Filtering*”. Munculnya kendala yang sering dialami oleh mahasiswa yaitu kesulitan dalam menentuka topik tugas akhir atau skripsi mereka, sehingga penelitian ini dilakukan untuk menghasilkan sistem rekomendasi topik tugas akhir. Tujuannya agar memudahkan para mahasiswa saat menentukan topik untuk skripsi mereka yang diukur melalui persentase nilai *relevance*, *novelty*, *serendipity*, dan *diversity*. Dalam penelitian ini, untuk pembobotan teks digunakan metode TF-IDF sedangkan untuk mengukur kemiripan antar dokumen, pada penelitian ini menggunakan metode *Euclidean Distance*. Berdasarkan hasil analisis, didapatkan persentase rata-rata dari pencapaian tujuan sistem rekomendasi yaitu 4 aspek diatas sebesar 83%. Selain itu,

sistem rekomendasi yang dibangun mempunyai rata-rata running time sebesar 7,46 detik (Kusuma & Musdholifah, 2021).

Penelitian yang dilakukan oleh Yusuf & Cherid (2020) yang berjudul “Implementasi Algoritma *Cosine Similarity* dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar” memiliki tujuan untuk memudahkan dalam menemukan seminar yang cocok dengan keinginan user tanpa harus melihat daftar yang terdiri atas banyak seminar. Metode yang digunakan dalam membangun sistem rekomendasi pada penelitian ini adalah *Content-Based Filtering* dan salah satu algoritma dari *text mining* yaitu *Cosine Similarity*. Berdasarkan sistem rekomedasi yang telah dibangun, sistem mampu berjalan dan sesuai dengan prediksi dimana dibandingkan langsung dengan data yang ada pada *database*.

Penelitian selanjutnya dilakukan oleh Larasati & Februariyanti (2021) yang berjudul “Sistem Rekomendasi *Product Emina Cosmetics* dengan Menggunakan Metode *Content-Based Filtering*”. Penelitian ini merancang suatu sistem rekomendasi yang memberikan anjuran kepada pelanggan dalam memilih produk-produk kosmetik Emina berdasarkan produk yang dicari sebelumnya. Sistem rekomendasi ini dikembangkan karena para pelanggan terkadang merasa kesusahan dalam pemilihan produk yang akan dibeli dan digunakan. *Content-based Filtering* dipilih sebagai metode dalam penelitian ini, dimana data yang digunakan adalah isi konten dari produk-produk emina. Dalam perhitungan nilai kemiripan masing-masing deskripsi produk, digunakan metode *Cosine Similarity* sedangkan untuk pembobotan antar kata menggunakan metode TF-IDF. Dari sistem rekomendasi yang dibangun, dihasilkan 10 produk kosmetik emina sebagai rekomendasi dengan mengurutkan hasil dari perhitungan kemiripan teks mulai dari yang tertinggi ke yang terendah. Nilai perhitungan *cosine similarity* yang tertinggi sebesar 0,7195.

Penelitian yang dilakukan oleh Fiarni et al. (2019) dengan judul “*Product Recommendation System Design Using Cosine Similarity dan Content-based Filtering Methods*” membangun sistem rekomendasi yang dapat mengelompokkan produk dan pelanggan sesuai dengan karakteristiknya, kemudian memberikan rekomendasi

alternatif produk kepada pelanggan berdasarkan karakteristik dan riwayat transaksi. Sistem tersebut dapat membantu perusahaan dalam mempromosikan produk yang sesuai dengan karakteristik pelanggan mereka. Penelitian ini menggunakan metode *Content-based Filtering*. Dalam menghitung skor kesamaan produk, dibandingkan tiga metode untuk mendapatkan skor terbaik yaitu dengan metode *Cosine Similarity*, *Profile Matching*, dan *K-Means Algorithm*. Dari ketiga model yang diusulkan, didapatkan metode yang terbaik yaitu *Cosine Similarity*. Berdasarkan hasil analisis dan perhitungan, nilai *precision* dan *recall* dari skor kesamaan produk dan rekomendasi pelanggan sebesar 100% dan 93,47%. Di sisi lain, metode yang paling tepat dalam memberikan rekomendasi kepada pelanggan adalah *Content-based Filtering* karena metode ini menghitung skor rekomendasi pelanggan yang dapat membantu perusahaan dalam pemasaran produknya sehingga pemasaran dapat lebih efektif.

Berdasarkan beberapa penelitian terdahulu yang penulis gunakan sebagai bahan acuan dan referensi, maka didapatkan rangkuman dan perbedaan dari penelitian sebelumnya. Berikut merupakan rangkuman dan perbedaan penelitian yang peneliti lakukan saat ini dengan penelitian terdahulu yang disajikan dalam Tabel 2.1.

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
1	Ridwan Rismanto, Arie Rachmad Syulistyo, & Bebbly Pramudya Citra Agusta (2020)	<i>Research Supervisor Recommendation System Based on Topic Conformity</i>	<i>Cosine Similarity</i> dan Metode TF-IDF	Sama-sama menggunakan <i>Cosine Similarity</i> dan Metode TF-IDF	Studi kasus penelitian sekarang yaitu deskripsi hotel di Yogyakarta
2	Muhammad Alkaff, Husnul Khatimi, & Andi Erlady (2020)	Sistem Rekomendasi Buku Menggunakan <i>Weighted Tree</i>	<i>Weighted Tree Similarity</i> dan <i>Content Based</i>	Sama-sama menggunakan <i>Content Based Fitering</i>	Penelitian sekarang tidak menggunakan <i>Weighted Tree Similarity</i> , tetapi

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
		<i>Similarity dan Content Based Filtering</i>	<i>Filtering</i>		menggunakan metode <i>Cosine Similarity</i>
3	Resha Havilah Mondi, Ardhi Wijayanto, & Winarno (2019)	<i>Recommendation System with Content-Based Filtering Method for Culinary Tourism in Mangan Application</i>	<i>Cosine Similarity dan Content Based Filtering</i>	Sama-sama menggunakan <i>Cosine Similarity dan Content Based Filtering</i>	Dalam proses <i>preprocessing</i> data, penelitian sekarang tidak menggunakan Eliminasi, <i>Flitering</i> , dan <i>Steaming</i> . Selain itu, pada penelitian sekarang dilakukan pembobotan kata menggunakan metode <i>TF-IDF</i> .
4	Hans Satria Kusuma & Aina Musdholifah (2021)	<i>Recommendation System for Thesis Topics Using Content-Based Filtering</i>	<i>Euclidean Distance dan Content Based Filtering</i>	Sama-sama menggunakan <i>Content Based Fitering</i>	Penelitian sekarang tidak menggunakan <i>Euclidean Distance</i> , tetapi menggunakan metode <i>Cosine Similarity</i> .
5	Muhtar Yusuf & Anis Cherid (2020)	Implementasi Algoritma <i>Cosine Similarity</i> dan Metode TF-IDF Berbasis PHP	Algoritma <i>Cosine Similarity</i> dan Metode TF-IDF	Sama-sama menggunakan <i>Cosine Similarity dan Content Based</i>	Dalam proses <i>preprocessing</i> data, penelitian sekarang menggunakan <i>Lemmatization</i>

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
		Untuk Menghasilkan Rekomendasi Seminar		<i>Fitering</i>	sedangkan penelitian terdahulu tidak ada proses tersebut.
6	Fatoni Batari Agung Larasati & Henry Februariyanti (2021)	Sistem Rekomendasi <i>Product Emina Cosmetics</i> dengan Menggunakan Metode <i>Content-Based Filtering</i>	<i>Cosine Similarity</i> dan <i>Content Based Filtering</i>	Sama-sama menggunakan <i>Cosine Similarity</i> dan <i>Content Based Fitering</i>	Studi kasus penelitian sekarang yaitu deskripsi hotel di Yogyakarta
7	Cut Fiarni, Herastia Maharani, & Nathania Calista (2019)	<i>Product Recommendation System Design Using Cosine Similarity dan content-based Filtering Methods</i>	<i>Cosine Similarity, Profile Matching, K-Means Alorithm,</i> dan <i>Content Based Filtering</i>	Sama-sama menggunakan <i>Cosine Similarity</i> dan <i>Content Based Fitering</i>	Pada penelitian sekarang hanya menggunakan <i>Cosine Similarity</i>

BAB 3 LANDASAN TEORI

3.1 *Online Travel Agent*

Online Travel Agent (OTA) merupakan biro perjalanan yang memiliki peran sebagai wadah penjualan dan penawaran akomodasi secara *online* melalui suatu *website*. Tidak jauh berbeda dengan agen travel biasanya, segala kegiatan pencarian informasi, pemesanan, dan transaksi pembayaran pada OTA dilakukan dengan menggunakan internet atau secara *online*. Menurut Malla (2016) dalam (Hendriyati, 2019) OTA menawarkan berbagai macam akomodasi dan pelayanan meliputi reservasi maskapai penerbangan, hotel, paket liburan, isi ulang pulsa, pembayaran tagihan, dan lain-lain. *Online travel agent* dibagi menjadi dua jenis yaitu (Hendriyati, 2019) :

1. *Quarantee Payment*

Online travel agent jenis *quarantee payment* adalah agen travel online dimana dalam pemesanan hotel maupun akomodasi lainnya, pelanggan harus menyelesaikan transaksi dengan membayar penuh kepada pihak agen travel online sebelum mendapatkan *voucher* akomodasi tersebut. Dengan demikian pembayaran akomodasi adalah tanggung jawab dari pihak OTA. Jika suatu saat terjadi pembatalan pesanan, pihak hotel tidak akan rugi karena masih dapat menarik pembayaran dari pihak OTA sesuai dengan perjanjian.

2. *Non Quarantee Payment*.

Online travel agent jenis *non quarantee payment* adalah agen travel online dimana dalam pemesanan hotel maupun akomodasi lainnya tidak mengharuskan pelanggan untuk menyelesaikan transaksi pembayaran, dengan demikian pelanggan yang sudah melakukan registrasi bisa langsung memperoleh *voucher* akomodasi tersebut. Oleh karena itu, pembayaran

akomodasi bukan merupakan tanggung jawab pihak OTA. Jenis OTA *non guarantee payment* memiliki kekurangan yaitu apabila suatu saat terjadi pembatalan pesanan, pihak hotel bisa dirugikan karena tidak bisa menarik pembayarannya dari pihak OTA maupun dari pelanggan.

3.2 Nusatrip

Nusatrip merupakan salah satu agen *travel online* yang menawarkan pelayanan pemesanan online maskapai penerbangan dan hotel domestik maupun internasional. Nusatrip didirikan oleh Hans Ebenhahn pada April tahun 2013 dan berkedudukan di Jakarta. Nusatrip merupakan agen travel online yang berlisensi. Agen travel online ini memiliki akreditasi IATA (*International Air Transport Association*) dan merupakan anggota dari ASITA (*Association of the Indonesia Tour and Travel Agencies*). Selain itu, Nusatrip merupakan bagian dari grup yang memiliki bermacam-macam jenis perusahaan yang memusatkan diri pada penanaman modal industri di Indonesia yaitu GLB Group (PT. Gema Lintas Benua). Agen ini memberikan berbagai macam informasi harga, ketersediaan tempat, dan informasi terkait hotel di Indonesia maupun mancanegara. Nusatrip menyatukan teknologi dan informasi travel masa kini yang memudahkan pelanggan dalam memilih ribuan tiket maskapai penerbangan dengan harga yang murah. Terdapat 460 maskapai penerbangan domestik maupun internasional yang berangkat dari Indonesia dan siap melayani pemesanan 100.000 hotel di 10.000 destinasi di Indonesia ataupun seluruh dunia. Pemesanan tiket akomodasi menggunakan portal ini dapat diakses pada *website* Nusatrip.com.



Gambar 3.1 Logo *Nusatrip.com*

3.3 Hotel

Berdasarkan Peraturan Pemerintah Republik Indonesia No.65 Tahun 2021, hotel adalah bangunan yang secara khusus diperuntukkan untuk menginap ataupun beristirahat dan memperoleh pelayanan serta fasilitas dengan dipungut biaya. Peraturan mengenai perhotelan dan klasifikasi hotel diatur dalam surat keputusan menteri perhubungan No.PM. 10/P.V.301/PHT/77 dalam (Pasaribu, 2019) mengklasifikasikan hotel berdasarkan bintang yaitu:

1. Hotel Bintang Satu

Hotel bintang satu dikelola langsung oleh pemiliknya sehingga tergolong kecil dengan kriteria:

- Minimal berjumlah 15 kamar
- Luas standar 20 m²
- Salah satu fasilitas yaitu kolam renang

2. Hotel Bintang Dua

Kriteria hotel bintang dua adalah:

- Minimal berjumlah 20 kamar termasuk satu kamar *suite*
- Luas kamar standar sebesar 24 m²
- Luas kamar *suite* sebesar 44 m²
- Salah satu fasilitas ada sarana olahraga

3. Hotel Bintang Tiga

Kriteria hotel bintang tiga adalah:

- Minimal berjumlah 30 kamar termasuk 2 kamar *suite*
- Luas kamar standar sebesar 24 m²
- Luas kamar *suite* sebesar 48 m²
- Memiliki fasilitas meliputi sarana olahraga, kolam renang, dan restoran yang menawarkan hidangan sarapan, makan siang, dan makan malam.

4. Hotel Bintang Empat

Kriteria hotel bintang empat adalah:

- Minimal berjumlah 50 kamar termasuk 3 kamar *suite*
- Luas kamar standar sebesar 24 m²
- Luas kamar *suite* sebesar 48 m²
- Memiliki *lobby* hotel dengan luas minimum 100 m²
- Kamarmandi dilengkapi dengan air panas/dingin
- Memiliki fasilitas meliputi bar, sarana rekreasi, sarana olahraga, dan fasilitas yang disediakan minimal diatas hotel bintang tiga.

5. Hotel Bintang Lima

Kriteria hotel bintang lima adalah:

- Minimal berjumlah 100 kamar termasuk 4 kamar *suite*
- Pelayanan multi bahasa
- Luas kamar standar minimal 26 m²
- Luas kamar *suite* minimal 52 m²
- Perabotan dan tempat tidur menggunakan kualitas tertinggi.
- Terdapat restoran dengan layanan 24 jam.
- Terdapat fasilitas sarana olahraga termasuk lapangan *tennis*, *fitness center*, *sauna*, dan *bowling center*.

3.4 Text Mining

Salah satu bagian dari data *mining* adalah *text mining*. *Text mining* atau pencarian pengetahuan pada basis data teks merupakan suatu tahapan untuk melakukan pencarian pengetahuan yang berpusat pada data teks atau dokumen, memiliki tujuan untuk mendapatkan suatu informasi yang berguna (Rivki & Bachtiar, 2017). *Text mining* merupakan ekstraksi menarik dari sebuah teks bebas atau teks yang tidak terstruktur yang mencakup pengambilan suatu informasi, klasifikasi teks, pengelompokan teks, hingga entitas, relasi, dan ekstraksi peristiwa lainnya (Kao & Poteet, 2007).

Sebelum *text mining* dilakukan, langkah awal yang harus dilalui adalah mempersiapkan data yang akan dianalisis. Tahap *text preprocessing* merupakan tahapan dimana pengolahan data yang sebelumnya merupakan data tidak berstruktur menjadi data yang berstruktur (Nugraha & Sebastian, 2018). Sebuah teks biasanya memiliki struktur data yang masih berantakan, sehingga tidak langsung dapat diolah. Pada tahap inilah teks akan dibersihkan dari bagian yang tidak perlu hingga disiapkan menjadi teks yang terstruktur dan siap diolah lebih lanjut. Berikut adalah tahapan *text preprocessing* yang digunakan dalam penelitian ini:

1. *Case Folding*

Case Folding merupakan tahapan merubah huruf kapital yang terkandung pada sebuah teks menjadi huruf kecil (Wahyuni et al., 2017). Hal tersebut bertujuan agar suatu kata yang sama tetapi mengandung huruf kapital tidak terdeteksi berbeda dengan kata yang tidak mengandung huruf kapital.

2. *Remove Punctuation*

Tahapan *remove punctuation* dilakukan untuk menghilangkan karakter tanda baca seperti koma, titik, tanda tanya, tanda seru dan karakter selain angka dan huruf (Y. N. Prasetya et al., 2021).

3. *Remove Whitespace*

Remove whitespace adalah proses menghilangkan spasi berlebih yang ada dalam teks ataupun dokumen.

4. *Remove Stopword*

Remove Stopword merupakan tahapan untuk menghilangkan atau menghapus kata yang tidak memiliki arti dan tidak bermakna sehingga didapatkan teks atau dokumen dengan kata yang mengandung makna saja (Fitri et al., 2020). Kata-kata yang tidak mengandung makna dan tidak memiliki arti disebut *stopword*. Beberapa contoh *stopword* dalam bahasa inggris adalah *and, are, as, does, you, has, for, is, this*, dan sebagainya.

5. *Lemmatization*

Lemmatization merupakan suatu tahapan untuk menemukan bentuk dasar dari sebuah kata (lemma) dengan memperhatikan analisis morfologi dan kamus (Supriyati & Iqbal, 2018). Contohnya adalah kata bahasa inggris “*see*”, “*seeing*”, “*seen*” yang setelah dilakukan *lemmatization* akan dikembalikan ke bentuk kata dasarnya sesuai dengan kamus yaitu “*see*”.

6. *Tokenizing*

Tokenizing atau yang sering disebut tokenisasi adalah proses pemenggalan atau pemisahan kata pada sebuah teks atau dokumen berdasarkan tanda penghubung (-) dan tanda spasi (Husni et al., 2015). Tokenisasi dipakai untuk mengubah kalimat menjadi token-token atau bagian-bagian tertentu (Supriyati & Iqbal, 2018). Sebagai contoh, pada proses tokenisasi kalimat “*this property located near malioboro street*” akan menghasilkan 6 token yaitu “*this*”, “*property*”, “*located*”, “*near*”, “*malioboro*”, dan “*street*”

3.4 *Term Frequency-Invers Document Frequency (TF-IDF)*

Metode pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF) adalah sebuah cara yang digunakan untuk memberikan bobot hubungan suatu term atau kata terhadap sebuah dokumen dengan memadukan dua konsep untuk perhitungan bobot, yaitu *Term Frequency* (TF) yang merupakan frekuensi

kemunculan kata dalam suatu teks atau dokumen tertentu dan *Invers Document Frequency* (IDF) yang merupakan inverse frekuensi teks atau dokumen yang memuat kata tersebut (Nurjannah & Fitri Astuti, 2013). Menurut Robertson (2006:45) pada jurnal (Larasati & Februariyanti, 2021), TF-IDF merupakan metode yang efisien, mudah, dan menghasilkan hasil yang akurat. Dalam menentukan kesamaan antara dua teks atau dokumen, pembobotan kata menjadi penting dan berpengaruh,

3.4.1 Perhitungan *Term-Frequency* (TF)

TF merupakan frekuensi munculnya kata dalam suatu teks atau dokumen tertentu. Semakin besar frekuensi munculnya kata pada suatu teks atau dokumen maka semakin besar pula nilai bobot untuk kata itu sendiri (Apriani et al., 2021). Berikut merupakan beberapa macam dari *term frequency* (Siregar et al., 2017) :

1. *Raw TF*

Metode yang digunakan untuk mendapatkan nilai TF sebuah kata yang didapatkan berdasarkan kemunculan kata dalam sebuah teks atau dokumen. Misalkan suatu kata muncul dalam dokumen sebanyak 2 kali maka nilai dari TF adalah 2.

2. *Logarithmic TF*

Cara yang digunakan untuk mendapatkan nilai TF menggunakan fungsi matematika yaitu logaritmik. Berikut merupakan rumus dari Logarithmic TF:

$$TF = 1 + \log(TF) \quad (3.1)$$

3. *Binnary TF*

Dalam metode ini, munculnya kata atau *term* pada dokumen tidak memiliki pengaruh karena TF akan bernilai 0 jika sebuah kata tidak ada pada dokumen tersebut, sedangkan jika sebuah kata atau *term* ada pada dokumen tersebut TF akan bernilai 1.

4. *Augmented TF*

Augmented TF digunakan untuk pembobotan kata. Berikut merupakan rumus dari *Augmented TF*:

$$TF = 0,5 + 0,5 \times \frac{TF}{\max (TF)} \quad (3.2)$$

Dimana nilai max (TF) merupakan jumlah munculnya term terbanyak pada sebuah dokumen yang sama.

3.4.2 Perhitungan *Invers Document Frequency (IDF)*

Proses IDF berkebalikan dengan TF, dimana semakin besar frekuensi munculnya kata atau term maka nilai bobotnya akan semakin kecil nilai bobot kata itu sendiri (Apriani et al., 2021). Berikut persamaan dari IDF:

$$IDF_j = \ln\left(\frac{D}{df_j}\right) \quad (3.3)$$

Dimana:

IDF_j : *Invers document frequency* dari term j

D : Total dokumen

df_j : Total dokumen dari term j

3.4.3 Perhitungan *Term Frequency-Invers Document Frequency (TF-IDF)*

Perhitungan TF-IDF menggunakan persamaan (Nisrina, 2020):

$$W_{ij} = TF_{ij} \times IDF_j \quad (3.4)$$

$$W_{ij} = TF_{ij} \times \ln\left(\frac{D}{df_j}\right) \quad (3.5)$$

Dimana

W_{ij} : Bobot dari term j terhadap dokumen i

TF_{ij} : Frekuensi munculnya term j di dalam dokumen i

IDF_j : *Invers document frequency* dari term j

D : Total dokumen

df_j : Total dokumen dari term j

Berdasarkan persamaan 3.5, terlihat bahwa berapapun nilai dari TF_{ij} akan bernilai 0 jika nilai $D = df_j$, maka ditambahkan nilai 1 pada sisi IDF agar

menghindari nilai W_{ij} tidak menjadi 0 (Apriani et al., 2021). Persamaan dari TF-IDF akan menjadi seperti berikut:

$$W_{ij} = TF_{ij} \times \left(\ln \left(\frac{D}{df_j} \right) + 1 \right) \quad (3.6)$$

3.5 *Cosine Similarity*

Cosine similarity merupakan metode untuk mengukur kemiripan antara dua vektor n-dimensi yang biasanya digunakan pada bidang pencarian informasi untuk membandingkan dua teks atau dokumen (Jannach et al., 2011). Kata kunci atau keyword dalam sebuah dokumen dinyatakan sebagai ukuran untuk menghitung kemiripan antar dua dokumen. *Cosine similarity* adalah salah satu metode untuk mengukur kemiripan dokumen yang populer, dimana digunakan untuk menghitung nilai cosinus sudut antara dua vektor (Mondi et al., 2019). Nilai dari *cosine similarity* berada pada rentang 0 hingga 1. Pada saat kedua dokumen itu sama atau identik, maka sudutnya adalah 0 derajat sehingga nilai kesamaannya adalah 1. Sebaliknya, jika dua dokumen itu tidak sama atau tidak identik, maka sudutnya adalah 90 derajat dimana nilai kesamaannya adalah 0 (Yasni et al., 2018). Dengan demikian, jika dua buah teks atau dokumen semakin mirip maka nilai *cosine similarity* akan mendekati 1, sedangkan jika nilai *cosine similarity* mendekati 0 maka dua teks atau dokumen tersebut semakin tidak mirip. Berikut merupakan persamaan dari *cosine similarity* berdasarkan Ye (2014) pada jurnal (Wahyuni et al., 2017):

$$\text{Cos} \propto = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3.7)$$

Dimana:

- A : Vektor A, yang akan dibandingkan kesamaannya
- B : Vektor B, yang akan dibandingkan kesamaannya
- A_i : *Term* i yang ada pada dokumen A
- B_i : *Term* i yang ada pada dokumen B
- $A \cdot B$: *Dot product* antara vektor A dan vektor B

$|A|$: Panjang vektor A

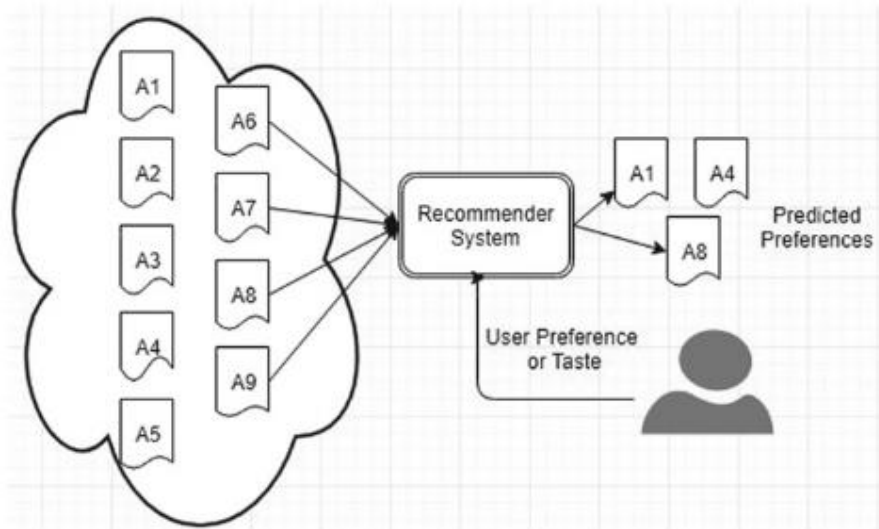
$|B|$: Panjang vektor B

$|A||B|$: *Cross product* antara $|A|$ dan $|B|$

3.6 *Recommendation System*

Recommendation system merupakan perangkat lunak dan suatu teknik yang memberikan saran sebuah *item* yang menarik bagi pengguna dan ditunjukkan untuk mendukung penggunaannya dalam berbagai proses pengambilan keputusan (Larasati & Februariyanti, 2021). Misalnya pada proses pengambilan keputusan seperti barang apa yang akan pengguna beli, musik apa yang akan didengar, film yang akan ditonton, hotel apa yang akan dipesan, berita apa yang akan dibaca dan masih banyak lagi. Sistem rekomendasi menjadi salah satu alat penemuan informasi paling kuat dan populer di suatu *website* untuk mengatasi informasi yang berlebih dan dapat membantu mereka dalam menentukan pilihan yang baik (Ricci et al., 2015). Hal tersebut sejalan pada era sekarang yaitu era digital, dimana terdapat banyak data yang tersedia pada suatu *website*. Dengan demikian *website* akan membutuhkan suatu sistem untuk penyaringan informasi. Adanya penyaringan informasi membuat seseorang dapat memprioritaskan informasi tersebut secara efisien dan tentunya akan terhindar dari masalah kelebihan informasi (Shah et al., 2016).

Dalam cara kerjanya, sistem rekomendasi tidak hanya akan memberikan rekomendasi *item* yang paling disukai oleh pengguna melainkan juga memberikan rekomendasi secara personal kepada setiap pengguna yang mana akan menyuguhkan rekomendasi *item* yang sesuai dengan ketertarikan pengguna (Mondi et al., 2019).

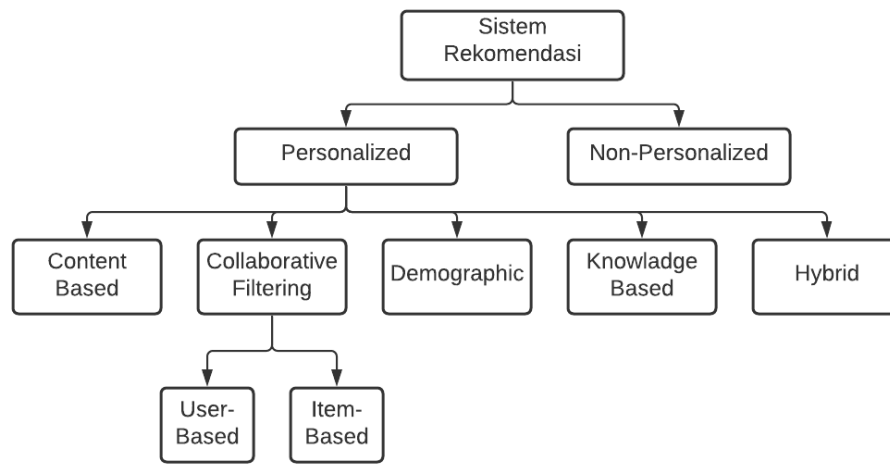


Gambar 3.2 Ilustrasi *Recommendation System* (Sumber: Raghuwanshi & Pateriya, 2019)

Ilustrasi dari sistem rekomendasi digambarkan seperti Gambar 3.2. Dalam suatu website atau aplikasi pasti memiliki *database* dari *item-item* yang berada di dalamnya. Sesuai dengan kegunaan sistem rekomendasi yaitu akan merekomendasikan atau memberikan saran *item-item* kepada pengguna berdasarkan jejak, *rating*, ataupun preferensi atau selera dari penggunanya. Dari berbagai *item* yang terdapat dalam suatu *database* pada *website* atau aplikasi, sistem yang berisi algoritma rekomendasi yang ditetapkan oleh pemberi rekomendasi akan mengolah jejak, *rating*, atau preferensi dari pengguna sehingga akan memberikan saran atau rekomendasi *item* sesuai dengan apa yang mereka inginkan, jejak yang mereka punya, atau bisa juga berdasarkan *rating* dari pengguna lainnya (Raghuwanshi & Pateriya, 2019).

3.7 Teknik *Recommendation System*

Terdapat dua tipe dari sistem rekomendasi yaitu *Personalized Recommendation System* dan *Non-Personalized Recommendation System* dimana terdapat berbagai teknik dari sistem rekomendasi yang ditampilkan pada Gambar 3.3.



Gambar 3.3 Tipe-tipe *Recommendation System* (Sumber: Arekar et al., 2015)

3.7.1 *Personalized Recommendation System*

Sistem rekomendasi tipe *personalized* bertujuan untuk memberikan saran atau rekomendasi *item* yang diinginkan pengguna berdasarkan perilaku dan preferensi dari masing-masing pengguna (Arekar et al., 2015)

3.7.1.1 *Content-Based Filtering*

Metode *Content-Based Filtering* merupakan metode sistem rekomendasi yang berbasis konten atau fitur dari *item* kemudian dibandingkan dengan *item* yang pengguna sukai sebelumnya (Raghuwanshi & Pateriya, 2019) Sistem rekomendasi menggunakan metode *Content-Based Filtering* ini hanya berdasarkan *item* yang dicari atau disukai oleh pengguna itu sendiri dan tidak melibatkan pengguna lain dalam penentuan rekomendasi. Dengan demikian, jika pengguna berubah, teknik dengan *Content-Based Filtering* masih memungkinkan untuk menyesuaikan rekomendasi atau saran *item* yang bersesuaian dalam waktu yang singkat (Purnaramadhan, 2021). Algoritma ini melihat kemiripan konten dari sebuah *item* yang kemudian akan direkomendasikan kepada pengguna (Larasati & Februariyanti, 2021)

Adapun kelebihan dan kekurangan dari Metode *Content-Based Filtering* yaitu (Aamir & Bhusry, 2015):

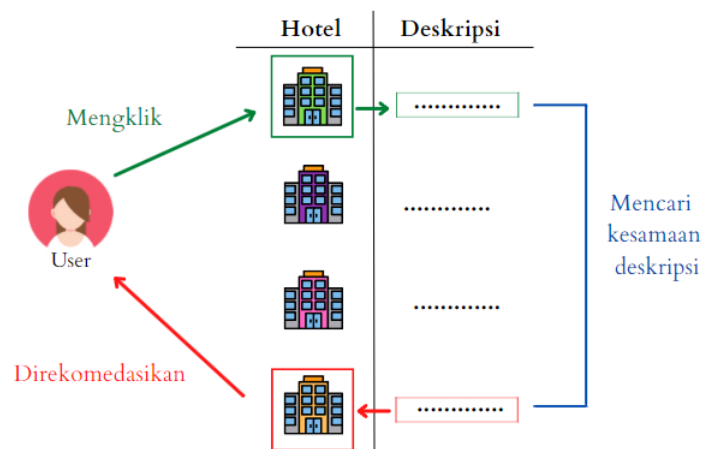
Kelebihan

- Metode *Content-Based Filtering* hanya memerlukan konten dari *item* dan profil pengguna itu sendiri untuk rekomendasi.
- Metode *Content-Based Filtering* dapat menjelaskan tentang fitur dari *item* yang menjadi dasar rekomendasi kepada pengguna.
- Item* yang baru dapat direkomendasikan kepada pengguna walaupun belum memiliki *rating* dari pengguna lain karena berdasarkan konten dari *itemnya*.

Kekurangan

- Jika konten pada *item* tidak menyertakan informasi yang lengkap dan memadai untuk membedakan dengan *item* yang lainnya secara akurat, rekomendasi akan menjadi kurang tepat.
- Kendala serendipity (kejadian yang tidak terduga), dimana sistem dengan metode ini akan sulit dalam memberikan rekomendasi atau saran yang tidak terduga *item* yang dipilih hanya berdasarkan konten.

Berikut merupakan ilustrasi dari *Content-Based Filtering* yang disajikan dalam Gambar 3.4.



Gambar 3.4 Ilustrasi *Content-Based Filtering* (Sumber: Dokumen Pribadi)

Berdasarkan Gambar 3.4, apabila pengguna memilih atau mengklik hotel hijau, maka sistem rekomendasi menggunakan *Content-Based Filtering* akan memindai konten atau deskripsi dari hotel tersebut kemudian akan mencari hotel lain yang ada pada *database* dengan konten yang bersesuaian atau memiliki konten yang mirip dengan hotel hijau. Pada ilustrasi, ternyata hotel orange merupakan hotel yang kontennya mirip dengan hotel hijau sehingga sistem rekomendasi akan menyarankan atau merekomendasikan hotel orange kepada pengguna.

3.7.1.2 Collaborative-Based Filtering

Metode *Collaborative-Based Filtering* merupakan metode rekomendasi dengan pemberian *rating* oleh pengguna dalam suatu *item* menjadi dasar rekomendasinya (Larasati & Februariyanti, 2021). Sederhananya, metode ini dapat juga diritikan sebagai metode yang memberikan saran atau rekomendasi berdasarkan suatu hal yang memiliki kemiripan preferensi antar satu pengguna dengan pengguna yang lain (Devi & Tonora, 2015). Perbedaan mendasar dengan metode *Content-Based Filtering* yang tidak melibatkan pengguna lain dalam penentuan rekomendasi, dalam metode ini justru melibatkan pengguna yang lain. Tujuan dari sistem rekomendasi *Collaborative-Based Filtering* adalah untuk memberikan saran atau rekomendasi *item* baru kepada pengguna berdasarkan kesukaan pengguna sebelumnya atau *rating* yang pengguna sebelumnya berikan terhadap suatu *item* (Aamir & Bhusry, 2015). Ada dua teknik dalam metode ini yaitu:

1. *User-Based*

Pendekatan *Collaborative-Based Filtering* menggunakan *User-Based* adalah berdasarkan minat pengguna yang memiliki selera atau preferensi yang sama dengan memberikan peringkat atau *rating* pada sebuah *item*.

2. *Item-Based*

Pendekatan menggunakan *Item-Based* didasarkan pada *item* karena *item* yang dinilai oleh pengguna memiliki kemiripan.

Adapun kelebihan dan kekurangan dari Metode *Collaborative-Based Filtering* yaitu (Devi & Tonora, 2015):

Kelebihan:

1. Metode *Collaborative-Based Filtering* masih tetap dapat bekerja walaupun konten pada suatu *item* atau pengguna sangat minim bahkan tidak ada. Hal tersebut dikarenakan metode ini tidak menggunakan konten *item* maupun profil pengguna sebagai dasar rekomendasinya.
2. Dapat menyelesaikan kesulitan penggunaan metode *Content-Based Filtering* yaitu pada analisis gambar, video, maupun musik karena pada *Content-Based Filtering* didasarkan oleh *history* yang ada sebelumnya.

Kekurangan:

1. Seringkali mengalami masalah *cold start users* dan *cold start item*, dimana pada saat ada *item* atau pengguna baru, *item* tersebut tidak langsung bisa dijadikan rekomendasi dikarenakan harus menunggu *rating* yang diberikan oleh pengguna lain terlebih dulu.

3.7.1.3 Demographic

Teknik metode demografis mendasarkan rekomendasinya menggunakan informasi dari pengguna (Arekar et al., 2015). Metode demografis adalah metode yang memanfaatkan atribut dari pengguna yang dikategorikan sebagai data demografis seperti usia, jenis kelamin, Pendidikan, dan lain-lain untuk menghasilkan rekomendasi (Aamir & Bhusry, 2015).

3.7.1.4 Knowledge-Based

Menurut Burke (2000) pada penelitian Arekar et al. (2015), pengetahuan eksplisit mengenai *item* klasifikasi, minat pengguna, dan standar dari rekomendasi *item* yang harus direkomendasikan adalah basis dari sistem rekomendasi *knowledge-based*. Dalam kasus seperti ingin membeli mobil, rumah dan lain-lain dimana kegiatan tersebut tidak sering dilakukan, peringkat mengenai *item* tidak berperan besar sehingga dalam kasus tersebut sistem rekomendasi *knowledge-based* cocok

digunakan karena dalam algoritma ini mempertimbangkan pengetahuan mengenai *item* dan fiturnya, preferensi pengguna, dan kriteria rekomendasi sebelum memberikan rekomendasi (Das et al., 2017).

3.7.1.5 Hybrid Filtering

Sistem rekomendasi *Hybrid Filtering* merupakan gabungan dari beberapa metode rekomendasi yang berbeda untuk membangun sistem yang lebih kuat, karena dengan menggabungkan beberapa metode tersebut dapat mengurangi kekurangan dari satu metode dengan adanya metode lain yang dikombinasikan (Das et al., 2017). Pendekatan sistem rekomendasi menggunakan metode *hybrid* biasanya dasarnya adalah menggabungkan *Collaborative Filtering* dan *Content-Based Filtering* untuk meningkatkan akurasi dari suatu sistem rekomendasi (B.Thorat et al., 2015).

3.7.2 Non-Personalized Recommendation System

Sistem rekomendasi *Non-Personalized* merupakan sistem rekomendasi yang merekomendasikan suatu *item* kepada pengguna berdasarkan apa yang pengguna lain katakan mengenai *item* tersebut secara rata-rata, hal tersebut berarti rekomendasi bersifat independen dari pelanggan sehingga semua pelanggan akan mendapatkan rekomendasi *item* yang sama (Arekar et al., 2015).

3.8 Flask

Flask adalah *web framework* berbahasa *Python*. Kegunaan dari *Flask* yaitu untuk kerangka kerja sebuah aplikasi dan *interface* yang terstruktur dari sebuah *website*. *Flask* merupakan *web framework* yang termasuk dalam jenis *microframework* yang bertujuan untuk membangun *core* dari sebuah aplikasi atau *website* dengan semudah mungkin. Hal tersebut membuat *Flask* disebut memiliki fleksibilitas dan skalabilitas yang tinggi dibandingkan *framework* lainnya (Irsyad, 2018). Beberapa keunggulan dari *Flask* adalah sebagai berikut.

1. Memiliki *core* sederhana sehingga mudah dan ringan dijalankan.

2. Memiliki kemampuan dengan mudah dalam memproses fungsi *HTTP request*.
3. Mempunyai API yang koheren.
4. Dokumentasi yang banyak dapat terstruktur dengan baik dan dapat langsung digunakan.
5. Mudah dalam men-*deploy* untuk produksi.
6. Memiliki fleksibilitas yang tinggi.

3.9 Heroku

Berdasarkan Heroku.com, Heroku merupakan *cloud platform* yang berguna sebagai tempat penyimpanan dan merupakan *web hosting* berbasis *cloud* yang digunakan untuk mengempangkan aplikasi *website* dengan berbagai bahasa pemrograman. Menurut Wikipedia, Heroku telah dikembangkan sejak tahun 2007 oleh James Lindenbaum, Adam Wiggins, dan Orion Henry di San Francisco, California. Saat ini ada berbagai macam bahasa pemrograman yang dapat dijalankan di platform ini diantaranya adalah *Java*, *Node.js*, *Scala*, *Clojure*, *Python*, *PHP*, dan *Go*.

BAB 4 METODOLOGI PENELITIAN

4.1 Populasi dan Sampel Penelitian

Populasi pada penelitian ini adalah semua informasi hotel pada *website online travel agent Nusatrip.com*, sedangkan sampel yang digunakan dalam penelitian ini adalah data nama, deskripsi dan bintang hotel di Yogyakarta pada *website Nusatrip.com* yang diambil pada tanggal 16 November 2021 dengan jumlah data yang diambil sebesar 705 data.

4.2 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut diperoleh dari hasil *scrapping* pada *website OTA Nusatrip.com* yang diambil pada tanggal 16 November 2021.

4.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah Nama Hotel, Deskripsi Hotel, dan Bintang Hotel. Pejelasan dan definisi operasional dari variabel-variabel tersebut disajikan dalam Tabel 4.1.

Tabel 4.1 Definisi Operasional Variabel

No	Variabel	Definisi Operasioal Variabel
1	Nama Hotel	Identitas yang digunakan untuk mengenali suatu hotel.
2	Deskripsi Hotel	Deskripsi hotel merupakan informasi terkait lokasi dan fasilitas dari hotel
3	Bintang Hotel	Jenis hotel berdasarkan penilaian fasilitas dan pelayanan dari sebuah hotel dimana semakin banyak bintang maka pelayanan dan fasilitas

No	Variabel	Definisi Operasioal Variabel
		hotel semakin baik dan lengkap. 0: Hotel non bintang 1: Hotel bintang 1 2: Hotel bintang 2 3: Hotel bintang 3 4: Hotel bintang 4 5: Hotel bintang 5
3	Fasilitas Hotel	Fasilitas hotel merupakan informasi mengenai fasilitas dari hotel.
4	<i>Nearby attraction</i> Hotel	<i>Nearby attraction</i> Hotel merupakan informasi jarak antara hotel ke tempat-tempat wisata

4.4 Metode Analisis Data

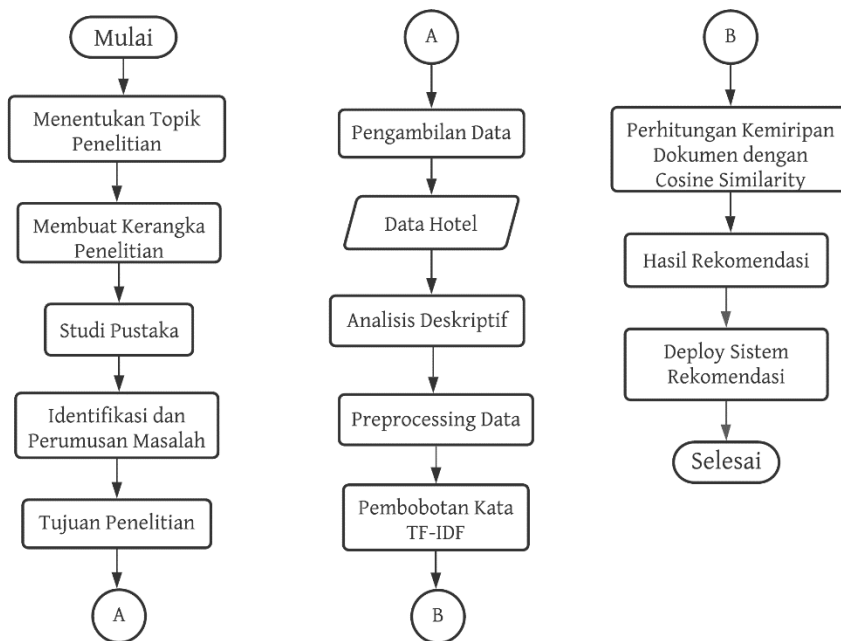
Alat dan *software* yang digunakan dalam penelitian ini adalah *Web Scraper* yang merupakan alat untuk melakukan *scrapping* pada *website*, *Microsoft Excel 2016* yang digunakan untuk menyimpan data serta pengolahan diagram, *Anaconda 4.8.3* untuk menganalisis data, serta *Flask Python* dan *Heroku* untuk *men-deploy* model. Berikut metode analisis data yang digunakan dalam penelitian ini:

1. Analisis deskriptif yang digunakan untuk mengetahui gambaran umum dari data hotel di Yogyakarta pada *website Nusatrip.com* terhitung sampai waktu pengambilan data yaitu 16 November 2021.
2. *Preprocessing* data yang digunakan untuk membersihkan data hingga siap diolah ke pengolahan berikutnya.
3. Pembobotan *term* atau kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

4. Sistem rekomendasi menggunakan metode *Content-Based Filtering* dengan melakukan perhitungan kemiripan antar teks atau dokumen menggunakan metode *Cosine Similarity* untuk menghasilkan rekomendasi hotel.

4.5 Tahapan Analisis Penelitian

Tahapan analisis dari penelitian ini dijelaskan dari penentuan topik penelitian hingga tahap analisis disajikan dalam Gambar 4.1.

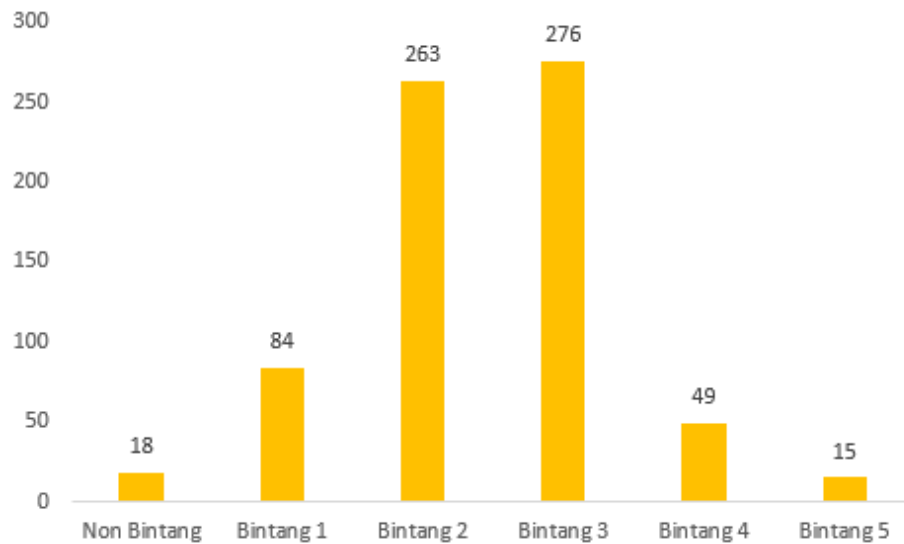


Gambar 4.1 Tahapan Analisis Penelitian

BAB 5 HASIL DAN PEMBAHASAN

5.1 Analisis Deskriptif

Analisis deskriptif ini berguna untuk melihat gambaran dari data yang digunakan yaitu data hotel di Yogyakarta yang didapatkan dari *website Nusatrip.com* pada tanggal 16 November 2021.



Gambar 5.1 Jumlah Hotel di Yogyakarta berdasarkan Bintang

Gambar 5.1 merupakan grafik dari jenis hotel berdasarkan bintang. Grafik tersebut menunjukkan jumlah hotel di Yogyakarta pada *website Nusatrip.com* yang diklasifikasikan berdasarkan bintang. Dari grafik yang terbentuk, terlihat bahwa sebaran hotel terbanyak ada pada hotel bintang dua dan bintang tiga. Salah satu faktor mengapa hotel bintang dua dan bintang tiga memiliki frekuensi yang banyak adalah menyesuaikan dengan pola kecenderungan konsumen. Hotel bintang dua dan bintang tiga ini memiliki fasilitas yang cukup lengkap dan juga harga yang tidak terlalu mahal. Hal tersebut sejalan dengan pendapat yang disampaikan *Co-Founder and Chief Marketing Office Tiket.com*, Gaery Undarsa pada artikel *republika.co.id* dimana kecenderungan konsumen seringkali memilih akomodasi dengan harga murah,

memiliki diskon, serta banyak promosi dimana mereka memilih akomodasi yang terjangkau. Hal tersebut melihat kecenderungan konsumen dalam memilih akomodasi hotel (Asrianti & Azizah, 2020).

5.2 *Preprocessing Data*

Tahapan pertama dalam proses analisis adalah *preprocessing data*. Data mentah seringkali memiliki struktur yang tidak beraturan dan mempunyai bagian-bagian yang tidak penting, maka tahapan ini perlu dilakukan agar data menjadi terstruktur sehingga dapat diolah dan dianalisis lebih lanjut. *Preprocessing data* yang dilakukan dalam penelitian ini adalah *Case Folding*, *Remove Punctuation*, *Remove Whitespace*, *Stopword Removal*, *Lemmatization*, dan *Tokenizing Text*. Berikut ini merupakan contoh data yang digunakan pada tahap *preprocessing data* yaitu menggunakan dokumen ke-132 yang disajikan dalam Tabel 5.1.

Tabel 5.1 *Dokumen Awal*

Deskripsi Hotel
<p>This property located near Malioboro Street. This smoke-free hostel features a 24-hour front desk, tour/ticket assistance, and a garden. WiFi in public areas is free. Rooms offer free WiFi and showers.</p> <p>Take in the views from a terrace and a garden and make use of amenities such as complimentary wireless Internet access.</p> <p>There is no front desk at this property. To make arrangements for check-in please contact the property ahead of time using the information on the booking confirmation. Make yourself at home in one of the 5 guestrooms. Complimentary wireless Internet access is available to keep you connected. Bathrooms with showers are provided.</p> <p><i>Nearby attractions</i> Batik Plentong 0.5 km / 0.3 mi Lana Gallery 1 km / 0.6 mi Masjid Jogokariyan 1.3 km / 0.8 mi Alun Alun Kidul 1.4 km / 0.9 mi Sasono Hinggil 1.6 km / 1 mi Water Castle (Tamansari) 2.1 km / 1.3 mi Taman Pintar Science Park 2.1 km / 1.3 mi Gumuling Well 2.2 km / 1.3 mi Kampoeng Cyber 2.2 km / 1.4 mi Malioboro Street 2.3 km / 1.4 mi De Mata Trick Eye 3D Museum 2.3 km / 1.4 mi Sonobudoyo Museum 2.3 km / 1.4 mi</p>

Deskripsi Hotel
Senopati 2.4 km / 1.5 mi Palace of Yogyakarta 2.4 km / 1.5 mi Fort Vredeborg Museum 2.5 km / 1.6 mi The nearest major airport is Yogyakarta (JOG-Adisucipto Intl.) - 11.9 km / 7.4 mi

5.2.1 Case Folding

Case Folding merupakan tahapan awal *preprocessing* data. Proses ini merupakan proses mengubah huruf kapital di dalam sebuah dokumen teks menjadi huruf kecil. Hal ini bertujuan agar kata yang mengandung huruf kapital tidak akan terdeteksi berbeda dengan kata yang tidak mengandung huruf kapital padahal kata tersebut adalah sama. Hasil dari proses *Case Folding* ditampilkan dalam Tabel 5.2 berikut.

Tabel 5.2 Hasil *Case Folding*

Dokumen ke-	Sebelum	Sesudah
132	<p>This property located near Malioboro Street. This smoke-free hostel features a 24-hour front desk, tour/ticket assistance, and a garden. WiFi in public areas is free. Rooms offer free WiFi and showers.</p> <p>Take in the views from a terrace and a garden and make use of amenities such as complimentary wireless Internet access.</p> <p>There is no front desk at this property. To make arrangements for check-in</p>	<p>this property located near malioboro street. this smoke-free hostel features a 24-hour front desk, tour/ticket assistance, and a garden. wifi in public areas is free. rooms offer free wifi and showers.</p> <p>take in the views from a terrace and a garden and make use of amenities such as complimentary wireless internet access.</p> <p>there is no front desk at this property. to make arrangements for</p>

Dokumen ke-	Sebelum	Sesudah
	<p>please contact the property ahead of time using the information on the booking confirmation.</p> <p>Make yourself at home in one of the 5 guestrooms. Complimentary wireless Internet access is available to keep you connected. Bathrooms with showers are provided.</p> <p><i>Nearby attractions</i></p> <p>Batik Plentong 0.5 km / 0.3 mi Lana Gallery 1 km / 0.6 mi Masjid Jogokariyan 1.3 km / 0.8 mi Alun Alun Kidul 1.4 km / 0.9 mi Sasono Hinggil 1.6 km / 1 mi Water Castle (Tamansari) 2.1 km / 1.3 mi Taman Pintar Science Park 2.1 km / 1.3 mi Gumuling Well 2.2 km / 1.3 mi Kampoeng Cyber 2.2 km / 1.4 mi Malioboro Street 2.3 km / 1.4 mi De Mata Trick Eye 3D Museum 2.3 km / 1.4 mi Sonobudoyo Museum 2.3 km / 1.4 mi Senopati 2.4 km / 1.5 mi Palace of Yogyakarta 2.4 km / 1.5 mi Fort Vredeburg Museum 2.5 km / 1.6 mi The nearest major airport is Yogyakarta (JOG-Adisucipto Intl.) - 11.9 km / 7.4 mi</p>	<p>check-in please contact the property ahead of time using the information on the booking confirmation.</p> <p>make yourself at home in one of the 5 guestrooms. complimentary wireless internet access is available to keep you connected. bathrooms with showers are provided.</p> <p><i>nearby attractions</i></p> <p>batik plentong 0.5 km / 0.3 mi lana gallery 1 km / 0.6 mi masjid jogokariyan 1.3 km / 0.8 mi alun alun kidul 1.4 km / 0.9 mi sasono hinggil 1.6 km / 1 mi water castle (tamansari) 2.1 km / 1.3 mi taman pintar science park 2.1 km / 1.3 mi gumuling well 2.2 km / 1.3 mi kampoeng cyber 2.2 km / 1.4 mi malioboro street 2.3 km / 1.4 mi de mata trick eye 3d museum 2.3 km / 1.4 mi sonobudoyo museum 2.3 km / 1.4 mi senopati 2.4 km / 1.5 mi palace of yogyakarta 2.4 km / 1.5 mi fort vredeburg museum 2.5 km / 1.6 mi the nearest major airport is yogyakarta (jog-adisucipto intl.) - 11.9 km / 7.4 mi</p>

5.2.2 *Remove Punctuation and Remove Whitespace*

Di dalam sebuah teks atau dokumen pasti mengandung karakter dan tanda baca, bisa berupa tanda titik, tanda koma, tanda tanya, tanda seru, dan lain-lain. Karakter dan tanda baca tersebut tidak diperlukan karena tidak memiliki makna dalam sebuah analisis data teks, sehingga harus dihapuskan atau dihilangkan. Untuk menghilangkan karakter dan tanda baca diperlukan adanya tahapan *Remove Punctuation*. Selain itu, jika teks atau dokumen mengandung paragraf maka akan memiliki spasi yang berlebih. Dalam mengatasi spasi yang berlebih, dilakukan tahapan *Remove Whitespace*. Hasil dari proses *Remove Punctuation* dan *Remove Whitespace* disajikan dalam Tabel 5.3.

Tabel 5.3 Hasil *Remove Punctuation* dan *Remove Whitespace*

Dokumen ke-	Sebelum	Sesudah
132	<p>this property located near malioboro street. this smoke-free hostel features a 24-hour front desk, tour/ticket assistance, and a garden. wifi in public areas is free. rooms offer free wifi and showers.</p> <p>take in the views from a terrace and a garden and make use of amenities such as complimentary wireless internet access.</p> <p>there is no front desk at this property. to make arrangements for check-in</p>	<p>this property located near malioboro street this smoke free hostel features a 24 hour front desk tourticket assistance and a garden wifi in public areas is free rooms offer free wifi and showers take in the views from a terrace and a garden and make use of amenities such as complimentary wireless internet access there is no front desk at this property to make arrangements for check in please contact the property ahead of time using the information on the booking confirmation make</p>

Dokumen ke-	Sebelum	Sesudah
	<p>please contact the property ahead of time using the information on the booking confirmation.</p> <p>make yourself at home in one of the 5 guestrooms. complimentary wireless internet access is available to keep you connected. bathrooms with showers are provided.</p> <p><i>nearby attractions</i></p> <p>batik plentong 0.5 km / 0.3 mi lana gallery 1 km / 0.6 mi masjid jogokariyan 1.3 km / 0.8 mi alun alun kidul 1.4 km / 0.9 mi sasono hinggil 1.6 km / 1 mi water castle (tamansari) 2.1 km / 1.3 mi taman pintar science park 2.1 km / 1.3 mi gumuling well 2.2 km / 1.3 mi kampoeng cyber 2.2 km / 1.4 mi malioboro street 2.3 km / 1.4 mi de mata trick eye 3d museum 2.3 km / 1.4 mi sonobudoyo museum 2.3 km / 1.4 mi senopati 2.4 km / 1.5 mi palace of yogyakarta 2.4 km / 1.5 mi fort vredeburg museum 2.5 km / 1.6 mi the nearest major airport is yogyakarta (jog-adisucipto intl.) - 11.9 km / 7.4 mi</p>	<p>yourself at home in one of the 5 guestrooms complimentary wireless internet access is available to keep you connected bathrooms with showers are provided <i>nearby attractions</i></p> <p>batik plentong 0.5 km / 0.3 mi lana gallery 1 km / 0.6 mi masjid jogokariyan 1.3 km / 0.8 mi alun alun kidul 1.4 km / 0.9 mi sasono hinggil 1.6 km / 1 mi water castle tamansari 2.1 km / 1.3 mi taman pintar science park 2.1 km / 1.3 mi gumuling well 2.2 km / 1.3 mi kampoeng cyber 2.2 km / 1.4 mi malioboro street 2.3 km / 1.4 mi de mata trick eye 3d museum 2.3 km / 1.4 mi sonobudoyo museum 2.3 km / 1.4 mi senopati 2.4 km / 1.5 mi palace of yogyakarta 2.4 km / 1.5 mi fort vredeburg museum 2.5 km / 1.6 mi the nearest major airport is yogyakarta jog adisucipto intl 11.9 km / 7.4 mi</p>

5.2.3 Remove Stopword

Remove Stopword atau *Stopword removal* merupakan tahapan untuk menghilangkan atau menghapus kata yang tidak memiliki arti dan tidak bermakna sehingga didapatkan teks atau dokumen dengan kata yang mengandung makna saja (Fitri et al., 2020). Dalam analisis teks, tahapan ini sangat penting untuk mengurangi jumlah kata yang harus diproses didalam sebuah teks atau dokumen. Contoh *stopword* dalam bahasa inggris adalah *and, are, as, does, you, has, for, is, this*, dan sebagainya. Hasil dari tahapan *Remove Stopword* ditampilkan dalam Tabel 5.4.

Tabel 5.4 Hasil *Remove Stopword*

Dokumen ke-	Sebelum	Sesudah
132	<p><i>this</i> property located near malioboro street <i>this</i> smoke free hostel features a 24 hour front desk tourticket assistance <i>and a</i> garden wifi <i>in</i> public areas <i>is</i> free rooms offer free wifi <i>and</i> showers <i>take in the</i> views <i>from a</i> terrace <i>and a</i> garden and make use of amenities such as complimentary wireless internet access there is no front desk at this property to make arrangements for check in please contact the property ahead of time using the information on the booking confirmation make yourself at home in one of the 5 guestrooms complimentary wireless internet access is available to keep you connected bathrooms with showers are provided <i>nearby attractions</i> batik plentong 0 5 km 0 3 mi lana gallery 1</p>	<p>property located near malioboro street smoke free hostel features 24 hour front desk tourticket assistance garden wifi public areas free rooms offer free wifi showers views terrace garden use amenities complimentary wireless internet access front desk property arrangements check contact property ahead time using information booking confirmation home one 5 guestrooms complimentary wireless internet access available keep connected bathrooms showers provided <i>nearby attractions</i> batik plentong 0 5 km 0 3 mi lana gallery 1 km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi</p>

Dokumen ke-	Sebelum	Sesudah
	km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4 mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace of yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi the nearest major airport is yogyakarta jog adisucipto intl 11 9 km 7 4 mi	sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4 mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi nearest major airport yogyakarta jog adisucipto intl 11 9 km 7 4 mi

5.2.4 Lemmatization

Sebuah teks atau dokumen pastinya mengandung kalimat dengan kata-kata yang bervariasi. Terkadang teks atau dokumen khususnya teks berbahasa Inggris sering kali menggunakan kalimat dengan kata kerja dalam bentuk lampau, aktif, maupun pasif. Dimana kata-kata tersebut tentunya berasal dari kata yang sama. Adanya variasi kata tersebut akan mempersulit proses pengolahan data karena kata yang sebenarnya memiliki makna atau arti yang sama akan menjadi berbeda, sehingga diperlukan proses *Lemmatization* atau mengekstrak akar kata. *Lemmatization* merupakan suatu tahapan untuk menemukan bentuk dasar dari sebuah kata (lemma) dengan memperhatikan analisis morfologi dan kamus (Supriyati & Iqbal, 2018). Contohnya adalah kata bahasa Inggris “see”, “seeing”, “seen” yang setelah dilakukan *lemmatization* akan dikembalikan ke bentuk kata dasarnya sesuai

dengan kamus yaitu “*see*”. Hasil dari proses *Lemmatization* ditampilkan pada Tabel 5.5 berikut.

Tabel 5.5 Hasil *Lemmatization*

Dokumen ke-	Sebelum	Sesudah
132	<p>property located near malioboro street smoke free hostel features 24 hour front desk tourticket assistance garden wifi public areas free rooms offer free wifi showers views terrace garden use amenities complimentary wireless internet access front desk property arrangements check contact property ahead time using information booking confirmation home one 5 guestrooms complimentary wireless internet access available keep connected bathrooms showers provided <i>nearby attractions</i> batik plentong 0 5 km 0 3 mi lana gallery 1 km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi</p>	<p>property located near malioboro street smoke free hostel feature 24 hour front desk tourticket assistance garden wifi public area free room offer free wifi shower view terrace garden use amenity complimentary wireless internet access front desk property arrangement check contact property ahead time using information booking confirmation home one 5 guestroom complimentary wireless internet access available keep connected bathroom shower provided <i>nearby attraction</i> batik plentong 0 5 km 0 3 mi lana gallery 1 km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4</p>

Dokumen ke-	Sebelum	Sesudah
	de mata trick eye 3d museum 2 3 km 1 4 mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi nearest major airport yogyakarta jog adisucipto intl 11 9 km 7 4 mi	mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi nearest major airport yogyakarta jog adisucipto intl 11 9 km 7 4 mi

5.2.5 Tokenizing

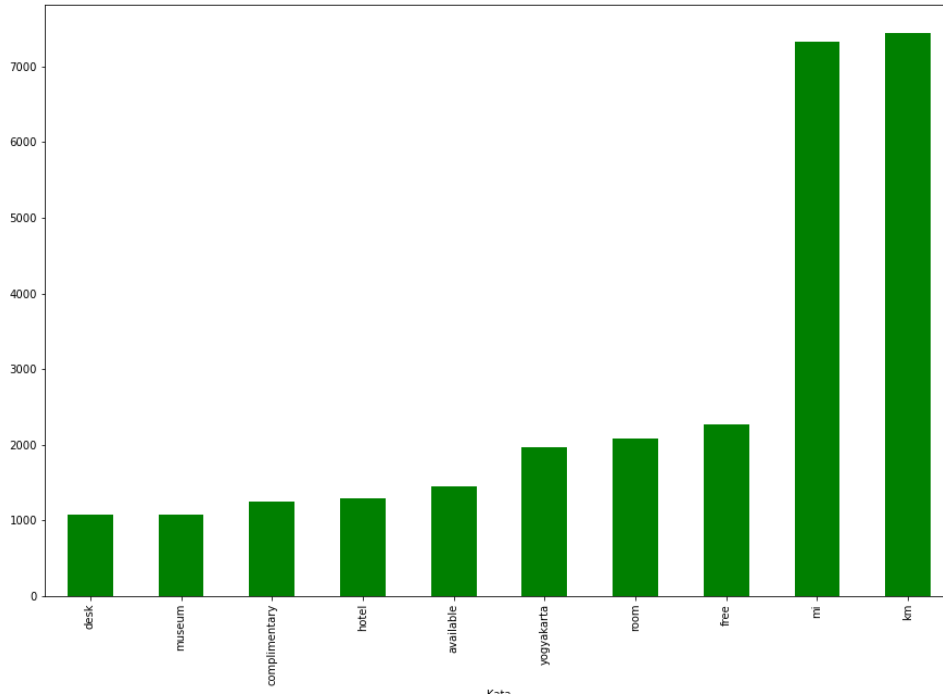
Proses *Tokenizing* digunakan untuk memisah kata pada sebuah teks atau dokumen berdasarkan tanda penghubung (-) dan tanda spasi sehingga mengubah kalimat menjadi token-token atau bagian-bagian tertentu. Hasil dari proses *Tokenizing* disajikan dalam Tabel 5.6

Tabel 5.6 Hasil *Tokenizing*

Dokumen ke-	Sebelum	Hasil Tokenizing
131	property located near malioboro street smoke free hostel feature 24 hour front desk tourticket assistance garden wifi public area free room offer free wifi shower view terrace garden use amenity complimentary wireless internet access front desk property arrangement check contact property ahead time using information booking confirmation home one 5 guestroom complimentary wireless internet access available keep connected	'property', 'located', 'near', 'malioboro', 'street', 'smoke', 'free', 'hostel', 'feature', '24', 'hour', 'front', 'desk', 'tourticket', 'assistance', 'garden', 'wifi', 'public', 'area', 'free', 'room', 'offer', 'free', 'wifi', 'shower', 'view', 'terrace', 'garden', 'use', 'amenity', 'complimentary', 'wireless', 'internet', 'access', 'front', 'desk', 'property', 'arrangement', 'check', 'contact', 'property', 'ahead', 'time', 'using', 'information', 'booking', 'confirmation', 'home', 'one', '5', 'guestroom', 'complimentary', 'wireless', 'internet', 'access', 'available', 'keep', 'connected'

Dokumen ke-	Sebelum	Hasil Tokenizing
	bathroom shower provided <i>nearby</i> <i>attraction</i> batik plentong 0 5 km 0 3 mi lana gallery 1 km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4 mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi nearest major airport yogyakarta jog adisucipto intl 11 9 km 7 4 mi	'complimentary', 'wireless', 'internet', 'access', 'available', 'keep', 'connected', 'bathroom', 'shower', 'provided', 'nearby', 'attraction', 'batik', 'plentong', '0', '5', 'km', '0', '3', 'mi', 'lana', 'gallery', '1', 'km', '0', '6', 'mi', 'masjid', 'jogokariyan', '1', '3', 'km', '0', '8', 'mi', 'alun', 'alun', 'kidul', '1', '4', 'km', '0', '9', 'mi', 'sasono', 'hinggil', '1', '6', 'km', '1', 'mi', 'water', 'castle', 'tamansari', '2', '1', 'km', '1', '3', 'mi', 'taman', 'pintar', 'science', 'park', '2', '1', 'km', '1', '3', 'mi', 'gumuling', 'well', '2', '2', 'km', '1', '3', 'mi', 'kampoeng', 'cyber', '2', '2', 'km', '1', '4', 'mi', 'malioboro', 'street', '2', '3', 'km', '1', '4', 'mi', 'de', 'mata', 'trick', 'eye', '3d', 'museum', '2', '3', 'km', '1', '4', 'mi', 'sonobudoyo', 'museum', '2', '3', 'km', '1', '4', 'mi', 'senopati', '2', '4', 'km', '1', '5', 'mi', 'palace', 'yogyakarta', '2', '4', 'km', '1', '5', 'mi', 'fort', 'vredeburg', 'museum', '2', '5', 'km', '1', '6', 'mi', 'nearest', 'major', 'airport', 'yogyakarta', 'jog', 'adisucipto', 'intl', '11', '9', 'km', '7', '4', 'mi'

Dari hasil tokenisasi kata pada data hotel didapatkan matriks datanya sebesar 705 baris \times 3654 kolom, berarti ada 705 dokumen dengan jumlah kata yang terbentuk sebesar 3654. Berikut merupakan kata-kata dengan frekuensi tertinggi yang ditampilkan pada Gambar 5.2.



Gambar 5.2 Frekuensi 10 Kata Tertinggi

Berdasarkan Gambar 5.2 yang menampilkan 10 kata tertinggi atau paling sering muncul dalam dokumen. Kata yang paling sering muncul adalah “*km*” dan “*mi*” dimana kata tersebut merupakan ukuran jarak hotel ke tempat-tempat wisata. Terdapat juga kata yang menunjukkan fasilitas hotel seperti “*room*” dan “*desk*”, serta kata sifat yang menggambarkan fasilitas seperti “*available*”, “*complimentary*”, dan “*free*”

Setelah proses melewati tahapan pertama yaitu *preprocessing data* dimana data yang tadinya belum terstruktur kemudian diubah menjadi data yang lebih terstruktur, dilanjutkan proses selanjutnya yaitu proses pembobotan kata.

5.3 Pembobotan Kata

Dalam memberikan bobot kata yang ada di dalam dokumen, digunakan metode pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF). Metode ini memadukan dua konsep untuk perhitungan bobot, yaitu *Term Frequency* (TF) yang merupakan frekuensi kemunculan kata dalam suatu teks atau dokumen tertentu dan *Invers Document Frequency* (IDF) yang merupakan *inverse* frekuensi teks atau dokumen yang memuat kata tersebut (Nurjannah & Fitri Astuti, 2013). Peneliti menggunakan beberapa sampel kata yang terdapat dalam dokumen yaitu “*airport*”, “*housekeeping*”, “*laundry*”, “*nearby*”, dan “*wifi*”.

Langkah pertama yang harus dilakukan adalah menghitung nilai *Term Frequency* (TF). Berikut adalah hasil perhitungan TF ditampilkan pada Tabel 5.7.

Tabel 5.7 Hasil Perhitungan TF pada Sampel Pilihan

Dokumen ke-	airport	housekeeping	laundry	nearby	...	wifi
1	1	0	1	0	...	0
2	0	0	1	0	...	0
3	1	2	0	1	...	1
4	3	1	1	2	...	2
5	0	0	0	0	...	0
6	1	1	0	1	...	2
7	1	0	0	1	...	2
8	0	0	1	0	...	2
9	0	0	0	0	...	0
10	2	0	0	1	...	1
11	0	0	1	0	...	1
12	0	0	1	1	...	0
13	1	0	0	1	...	0
14	1	0	0	0	...	1
15	2	0	0	1	...	2
16	1	0	0	1	...	0
17	1	1	0	1	...	2
18	3	1	1	1	...	2
19	2	0	0	1	...	2
20	1	0	1	0	...	0
⋮	⋮	⋮	⋮	⋮	...	⋮

Dokumen ke-	airport	housekeeping	laundry	nearby	...	wifi
705	0	0	0	0	...	0
df	519	222	164	461	...	456
IDF	0.306294	1.15552	1.458331	0.4248	...	0.435705

Contoh yang digunakan pada perhitungan manual penelitian ini adalah kata “*laundry*” yang berada pada dokumen ke-4. Nilai 0 pada tabel TF memiliki arti bahwa pada suatu dokumen tertentu, sebuah kata tertentu tidak muncul dalam dokumen tersebut, sedangkan nilai 1 berarti bahwa pada suatu dokumen tertentu, sebuah kata tertentu muncul sebanyak 1 kali pada dokumen tersebut. Misalkan kata “*laundry*” dalam dokumen 4. Berdasarkan Tabel 5.7, nilai TF pada kata “*laundry*” dalam dokumen ke-4 adalah 1, artinya kata “*laundry*” muncul sebanyak 1 kali pada dokumen ke-4. Selanjutnya mencari nilai *df* yang merupakan nilai *document frequency* dimana merupakan banyaknya dokumen yang dari suatu *term* yang muncul. Misalkan pada kata “*laundry*”, didapatkan sebanyak 164 dokumen yang mengandung kata tersebut. Kemudian dilanjutkan menghitung nilai *Invers Document Frequency* (IDF) menggunakan persamaan 3.3.

$$IDF_j = \ln\left(\frac{D}{df_j}\right)$$

$$IDF_{laundry} = \ln\left(\frac{705}{164}\right) = 1.458331$$

D merupakan total dokumen atau banyaknya dokumen pada penelitian ini yaitu sebanyak 705 dokumen sedangkan *df_j* ini merupakan nilai *document frequency* kata “*laundry*” yaitu sebanyak 164 dokumen. Setelah mendapatkan nilai IDF, Langkah berikutnya adalah mencari bobot dari kata yaitu nilai dari *Term Frequency-Invers Document Frequency* (TF-IDF). Berikut adalah hasil perhitungan TF-IDF ditampilkan pada Tabel 5.8

Tabel 5.8 Hasil Perhitungan TF-IDF pada Sampel Pilihan

Dokumen ke-	airport	housekeeping	laundry	nearby	...	wifi
1	1.306294	0	2.458331	0	...	0

Dokumen ke-	airport	housekeeping	laundry	nearby	...	wifi
2	0	0	2.458331	0	...	0
3	1.306294	4.311041	0	1.4248	...	1.435705
4	3.918882	2.15552	2.458331	2.8496	...	2.87141
5	0	0	0	0	...	0
6	1.306294	2.15552	0	1.4248	...	2.87141
7	1.306294	0	0	1.4248	...	2.87141
8	0	0	2.458331	0	...	2.87141
9	0	0	0	0	...	0
10	2.612588	0	0	1.4248	...	1.435705
11	0	0	2.458331	0	...	1.435705
12	0	0	2.458331	1.4248	...	0
13	1.306294	0	0	1.4248	...	0
14	1.306294	0	0	0	...	1.435705
15	2.612588	0	0	1.4248	...	2.87141
16	1.306294	0	0	1.4248	...	0
17	1.306294	2.15552	0	1.4248	...	2.87141
18	3.918882	2.15552	2.458331	1.4248	...	2.87141
19	2.612588	0	0	1.4248	...	2.87141
20	1.306294	0	2.458331	0	...	0
⋮	⋮	⋮	⋮	⋮	...	⋮
705	0	0	0	0	...	0

Berikut ini merupakan perhitungan manual nilai TF-IDF menggunakan persamaan 3.6.

$$W_{ij} = TF_{ij} \times \left(\ln \left(\frac{D}{df_j} \right) + 1 \right)$$

$$W_{4;laundry} = 1 \times \left(\ln \left(\frac{705}{164} \right) + 1 \right)$$

$$W_{4;laundry} = 1 \times (1,458331 + 1) = 2,458831$$

Berdasarkan hasil pembobotan yang telah dicari, maka bobot dari kata “*laundry*” pada dokumen ke-4 memiliki bobot kata sebesar 2,458831.

Setelah dilakukan perhitungan nilai TF-IDF secara manual, hasil tersebut dibandingkan dengan perhitungan menggunakan *software Anaconda 4.8.3 Python* dan

didapatkan hasil yang sama. Hasil lengkap dari perhitungan TF-IDF secara manual dan menggunakan *software* dapat dilihat pada lampiran. Berikutnya akan dilakukan perhitungan kemiripan dokumen menggunakan metode *Cosine Similarity*.

5.4 Hasil Rekomendasi Menggunakan *Cosine Similarity*

Sistem rekomendasi pada penelitian kali ini adalah dengan membandingkan kemiripan teks atau dokumen yang berisi deskripsi dari masing-masing hotel dengan menggunakan metode *Cosine Similarity*. Sebelumnya telah dihitung nilai dari pembobotan kata menggunakan TF-IDF yang selanjutnya akan dihitung kemiripan antar dokumennya. Kemudian akan didapatkan nilai kemiripan dari suatu dokumen dengan dokumen yang lain, selanjutnya akan diurutkan dari yang terbesar dan diambil 10 dokumen teratas dengan nilai kemiripan tertinggi sebagai hasil rekomendasi. Sebagai contoh peneliti menggunakan dokumen ke-132 sebagai sampel perhitungan manual. Berikut merupakan 10 dokumen teratas yang dihasilkan sistem rekomendasi ini menggunakan perhitungan nilai *Cosine Similarity* yang disajikan dalam Tabel 5.9.

Tabel 5.9 10 Dokumen yang Mirip dengan Dokumen ke-132

No	Dokumen ke-	Cosine Similarity
1	170	0.956666513
2	303	0.946570717
3	210	0.917459394
4	360	0.912534877
5	315	0.886439718
6	422	0.88221982
7	501	0.881775275
8	45	0.875845794
9	63	0.872030219
10	223	0.871514859

Berikut ini merupakan perhitungan manual nilai *Cosine Similarity* antara dokumen ke-132 dan dokumen ke-170 menggunakan persamaan 3.7.

$$\text{Cos } \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$\text{Cos } \alpha = \frac{1638,29225}{\sqrt{1708,006} \times \sqrt{1812,564}} = 0,956666513$$

Dari hasil perhitungan *Cosine Similarity*, dokumen ke-132 memiliki nilai kemiripan tertinggi dengan dokumen ke-170 yaitu sebesar 96% dan disusul dengan sembilan dokumen lainnya. Sesuai data yang diperoleh, dokumen ke-132 merupakan dokumen dari Good Karma Yogyakarta- Hostel. Dengan demikian 10 urutan rekomendasi hotel yang mirip dengan Good Karma Yogyakarta- Hostel dari sistem rekomendasi yang telah dibangun ditampilkan pada Tabel 5.10.

Tabel 5.10 Hasil Rekomendasi Hotel

No	Dokumen ke-	Nama Hotel	Nilai Cosim
1	170	Happy Buddha Yogyakarta - Hostel	0.956666513
2	303	Nextdoor Homestay	0.946570717
3	210	Hotel Puspita	0.917459394
4	360	OYO 426 Hotel Gading Resto	0.912534877
5	315	Omah Jegog Homestay	0.886439718
6	422	Prawirotaman Homestay	0.88221982
7	501	RedDoorz near Prawirotaman	0.881775275
8	45	Ayodhya Garden Hostel Yogyakarta by HOM	0.875845794
9	63	Bringin House Yogyakarta	0.872030219
10	223	House 24 Yogyakarta	0.871514859

Pada saat pengguna memilih atau mengklik Good Karma Yogyakarta- Hostel maka sistem akan memberikan 10 rekomendasi hotel lain yang memiliki kemiripan

tertinggi dengan hotel yang pengguna pilih. Kemiripan tersebut diukur berdasarkan kemiripan deskripsi dari masing-masing hotel. Contohnya, peneliti akan membandingkan kemiripan dari deskripsi dokumen ke-132 yaitu Good Karma Yogyakarta- Hostel dengan deskripsi dokumen ke-170 yang mana merupakan rekomendasi hotel yang paling mirip yaitu Happy Buddha Yogyakarta – Hostel. Perbandingan dari dua hotel tersebut disajikan dalam Tabel 5.11.

Tabel 5.11 Keseluruhan Deskripsi dari Dokumen ke-132 dan Dokumen ke-170
Setelah *Preprocessing Data*

Good Karma Yogyakarta - Hostel	Happy Buddha Yogyakarta - Hostel
property located near malioboro street smoke free hostel feature 24 hour front desk tourticket assistance garden wifi public area free room offer free wifi shower view terrace garden use amenity complimentary wireless internet access front desk property arrangement check contact property ahead time using information booking confirmation home one 5 guestroom complimentary wireless internet access available keep connected bathroom shower provided <i>nearby attraction</i> batik plentong 0 5 km 0 3 mi lana gallery 1 km 0 6 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 1 4 km 0 9 mi sasono hinggil 1 6 km 1 mi water castle tamansari 2 1 km 1 3 mi taman pintar science park 2 1 km 1 3 mi gumuling well 2 2 km 1 3 mi kampoeng cyber 2 2 km 1 4 mi malioboro street 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4 mi sonobudoyo museum 2 3 km 1 4 mi senopati 2 4 km 1 5 mi palace	property located near palace yogyakarta along 24 hour front desk smoke free hostel tourticket assistance garden wifi public area free 8 room offer free wifi room service shower view terrace garden use amenity complimentary wireless internet access front desk property arrangement check contact property ahead time using information booking confirmation home one 8 guestroom prepare meal sharedcommunal kitchen complimentary wireless internet access available keep connected bathroom shower provided <i>nearby attraction</i> batik plentong 0 5 km 0 3 mi lana gallery 0 9 km 0 5 mi masjid jogokariyan 1 3 km 0 8 mi alun alun kidul 2 3 km 1 4 mi de mata trick eye 3d museum 2 3 km 1 4 mi water castle tamansari 2 3 km 1 5 mi gumuling well 2 4 km 1 5 mi sasono hinggil 2 4 km 1 5 mi kampoeng cyber 2 4 km 1 5 mi taman pintar science park 2 6 km 1 6 mi museum sasmitaloka panglima besar

Good Karma Yogyakarta - Hostel	Happy Buddha Yogyakarta - Hostel
yogyakarta 2 4 km 1 5 mi fort vredeburg museum 2 5 km 1 6 mi nearest major airport yogyakarta jog adisucipto intl 11 9 km 7 4 mi	jenderal sudirman 2 7 km 1 6 mi pakualaman palace 2 7 km 1 7 mi malioboro street 2 7 km 1 7 mi sonobudoyo museum 2 8 km 1 7 mi senopati 2 9 km 1 8 mi nearest major airport yogyakarta jog adisucipto intl 11 2 km 7 mi

Berdasarkan hasil perhitungan kemiripan dokumen deskripsi antara Good Karma Yogyakarta- Hostel dengan Happy Buddha Yogyakarta – Hostel sebesar 96% yang tertera pada Tabel 5.10, berikut adalah rincian dari kemiripan antar dua hotel tersebut yang disajikan pada Tabel 5.12.

Tabel 5.12 Perbandingan deskripsi Good Karma Yogyakarta- Hostel dan Happy Buddha Yogyakarta – Hostel

Deskripsi Hotel	Nama Hotel	
	Good Karma Yogyakarta - Hostel	Happy Buddha Yogyakarta - Hostel
Facility	front desk, tourticket assistance, garden, bathroom shower, view terrace garden, complimentary wireless internet access	front desk, tourticket assistance, garden, bathroom shower, view terrace garden, complimentary wireless internet access
<i>Nearby attraction</i>	batik plentong, lana gallery, masjid jogokriyan, alun-alun kidul, water castle tamansari, taman pintar science park, gumuling well, kampoeng cyber, malioboro street, de mata trick eye 3D museum,	batik plentong, lana gallery, masjid jogokriyan, alun-alun kidul, water castle tamansari, taman pintar science park, gumuling well, kampoeng cyber, malioboro street, de mata trick eye 3D museum,

Deskripsi Hotel	Nama Hotel	
	Good Karma Yogyakarta - Hostel	Happy Buddha Yogyakarta - Hostel
	sonobudoyu museum, yogyakarta adisucipto airport, senopati, sasono hinggil	sonobudoyu museum, yogyakarta adisucipto airport

5.5 Kaji Ulang Rekomendasi Hotel

Hasil rekomendasi pada subbab 5.4 didapati hasil yang terkadang kurang tepat sehingga perlu ditingkatkan. Hal tersebut akibat dari variabel deskripsi yang digunakan sebagai basis dari rekomendasi memuat semua deskripsi dari hotel. Selain itu, hasil dari rekomendasi tidak memberikan informasi hal-hal apa saja yang dianggap mirip dengan hotel yang sebelumnya dicari atau di klik oleh user. Sebagai contoh adalah pada Hotel Eastparc dimana ditemukan hasil yang bias pada hotel tersebut Berikut merupakan hasil rekomendasi Eastparc Hotel dan nilai *cosine similarity* yang disajikan dalam Tabel 5.13.

Tabel 5.13 Hasil Rekomendasi Hotel Eastparc berdasarkan Deskripsi

No	Dokumen ke-	Nama Hotel	Nilai Cosim
1	361	OYO 444 Rafitha Homestay	0.682801
2	462	RedDoorz @ Seturan Raya 2	0.67937
3	463	RedDoorz @ Solo Street	0.676323
4	317	Omah Ndeliq Jogja	0.669392
5	678	Woodpecker Pavilion	0.666206
6	430	Puri Janti Guest House	0.662063
7	655	Victory Kost Eksklusif	0.661634
8	533	RedDoorz Plus near Jogja Expo Center	0.658178

No	Dokumen ke-	Nama Hotel	Nilai Cosim
9	461	RedDoorz @ Seturan Raya	0.656208
10	553	RedDoorz Syariah near UIN Yogyakarta	0.655131

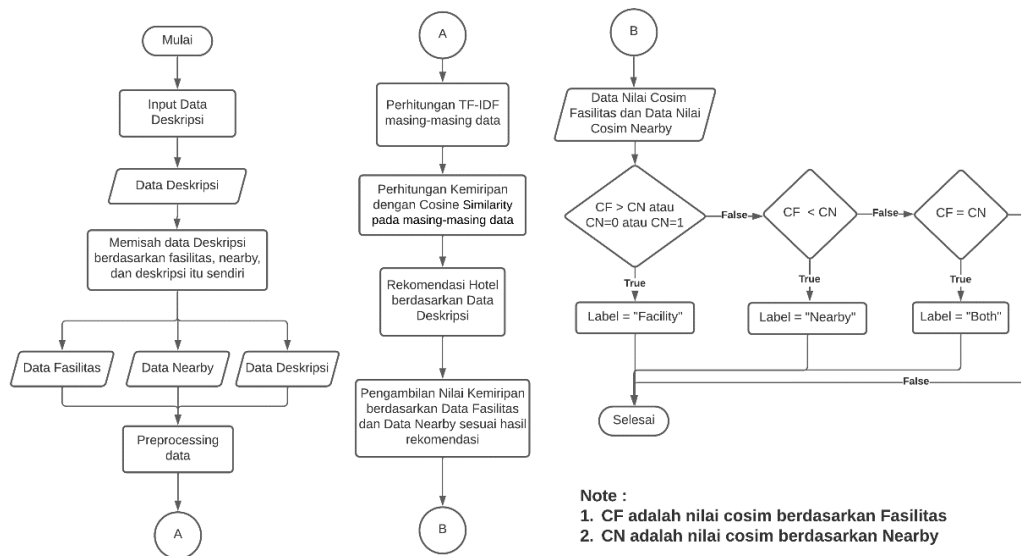
Hotel Eastparc merupakan hotel bintang 5 dan memiliki fasilitas yang setara dengan klasifikasi bintang tersebut, namun sistem rekomendasi memberikan rekomendasi hotel seperti OYO, RedDoorz, Guest House, dan lain-lain yang merupakan hotel kategori lowbudget. Jika dilihat dari hasil rekomendasi ini, hal tersebut tidak sesuai dengan persepsi masyarakat, sehingga rekomendasi tersebut tidak sesuai. Berdasarkan hal tersebut, untuk mendapatkan hasil yang lebih jelas, peneliti melakukan kaji ulang terhadap hasil rekomendasi dengan memisahkan deskripsi yang mana memuat semua informasi dari hotel secara umum menjadi dua bagian yaitu *Facility* dan *Nearby attraction* untuk dicari masing-masing nilai kemiripannya. Setelah data dipisahkan, kemudian dihitung nilai kemiripannya dari masing-masing data yaitu data *Facility* dan data *Nearby attraction* kemudian didapatkan hasil nilai kemiripan yang ditampilkan pada Tabel 5.14.

Tabel 5.14 Hasil Rekomendasi Hotel Eastparc dan Nilai *Cosine Similarity* berdasarkan Fasilitas serta *Nearby attraction*

No	Data ke-	Rekomendasi Hotel berdasarkan Deskripsi	Nilai Cosim berdasarkan <i>Facility</i>	Nilai Cosim berdasarkan <i>Nearby attraction</i>
1	361	OYO 444 Rafitha Homestay	0.33832	0.95521
2	462	RedDoorz @ Seturan Raya 2	0.30507	0.95800
3	463	RedDoorz @ Solo Street	0.33631	0.94090
4	317	Omah Ndeliq Jogja	0.30897	0.93271

No	Data ke-	Rekomendasi Hotel berdasarkan Deskripsi	Nilai Cosim berdasarkan <i>Facility</i>	Nilai Cosim berdasarkan <i>Nearby attraction</i>
5	678	Woodpecker Pavilion	0.40378	0.91650
6	430	Puri Janti Guest House	0.31406	0.93181
7	655	Victory Kost Eksklusif	0.30594	0.97266
8	533	RedDoorz Plus near Jogja Expo Center	0.36059	0.90672
9	461	RedDoorz @ Seturan Raya	0.30118	0.92272
10	553	RedDoorz Syariah near UIN Yogyakarta	0.28848	0.93100

Dari hasil pemisahan data tersebut dan menghitung masing-masing nilai kemiripannya berdasarkan fasilitas dan *nearby attraction* yang kemudian dibandingkan nilai kemiripannya didapati bahwa perbandingan nilai kemiripan berdasarkan *nearby attraction* lebih besar dari nilai kemiripan berdasarkan fasilitas. Untuk mendapatkan rekomendasi yang lebih jelas, peneliti menambahkan variabel info dari kemiripannya berdasarkan hasil perbandingan nilai kemiripan yang dihasilkan. Jika nilai kemiripan berdasarkan fasilitas lebih besar dari nilai kemiripan berdasarkan *nearby attraction* maka hasil rekomendasi hotel akan diberi label “Facility”, sedangkan jika nilai kemiripan berdasarkan fasilitas lebih kecil dari nilai kemiripan berdasarkan *nearby attraction* maka hasil rekomendasi hotel akan diberi label “Nearby”. Jika nilai kemiripan berdasarkan fasilitas sama dengan nilai kemiripan berdasarkan *nearby attraction* maka hasil rekomendasi hotel akan diberi label “Both”. Berikut adalah diagram alir dari algoritma tersebut yang disajikan dalam Gambar 5.3.



Gambar 5.3 Diagram Alir Algoritma Tambahan

Setelah menambahkan algoritma tersebut, dihasilkan rekomendasi hotel Eastparc yang ditampilkan pada Tabel 5.15.

Tabel 5.15 Hasil Rekomendasi Hotel Eastparc dan Info Kemiripannya

No	Data ke-	Rekomendasi Hotel berdasarkan Deskripsi	Info Kemiripan
1	361	OYO 444 Rafitha Homestay	Nearby
2	462	RedDoorz @ Seturan Raya 2	Nearby
3	463	RedDoorz @ Solo Street	Nearby
4	317	Omah Ndeliq Jogja	Nearby
5	678	Woodpecker Pavilion	Nearby
6	430	Puri Janti Guest House	Nearby
7	655	Victory Kost Eksklusif	Nearby
8	533	RedDoorz Plus near Jogja Expo Center	Nearby
9	461	RedDoorz @ Seturan Raya	Nearby
10	553	RedDoorz Syariah near UIN Yogyakarta	Nearby

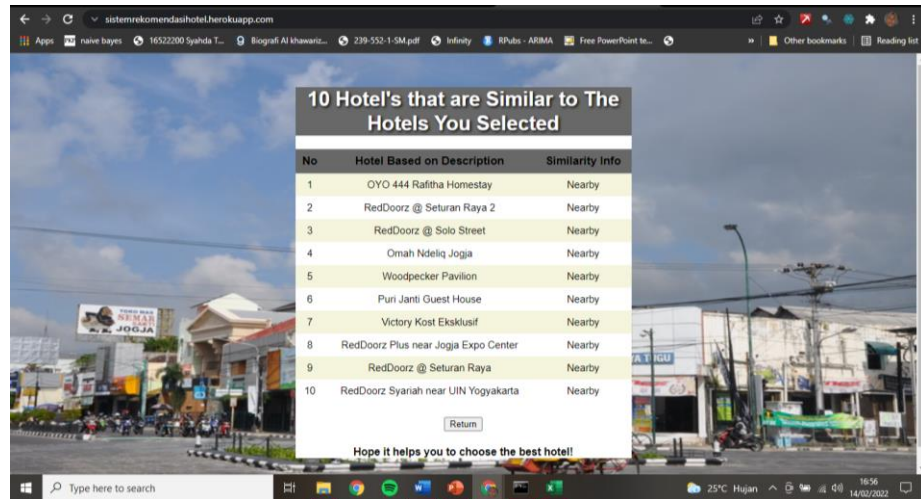
Melihat hasil nilai kemiripan berdasarkan fasilitasnya untuk hotel Eastparc pada Tabel 5.14 belum bagus karena *dataset* fasilitas hotel tersebut tidak lengkap. Adanya hal ini, disarankan kepada *website* Nusatrip untuk melengkapi deskripsi dari masing-masing hotel yang ada pada *website* tersebut agar mendapatkan hasil rekomendasi yang lebih baik dan akurat.

5.6 Pengaplikasian Sistem Rekomendasi pada *Website*

Langkah terakhir pada penelitian ini adalah *men-deploy* sistem rekomendasi yang telah dibuat ke dalam sebuah *website* sederhana menggunakan *Flask Python* kemudian *di-deploy* kedalam Heroku agar dapat diakses secara online untuk memudahkan pengguna dalam mencari hotel yang mirip dengan hotel pilihan sebelumnya. Tampilan *website* pengaplikasian sistem rekomendasi yang dibangun dalam penelitian ini disajikan dalam Gambar 5.4 dan Gambar 5.5.



Gambar 5.4 Halaman Pencarian



Gambar 5.5 Halaman Hasil Rekomendasi

Dapat diakses melalui <https://sistemrekomendasihotel.herokuapp.com/>

BAB 6 PENUTUP

6.1 Kesimpulan

Berdasarkan analisis dan pembahasan dari penelitian ini, didapatkan kesimpulan sebagai berikut:

1. Perhitungan kemiripan dokumen menggunakan *Cosine Similarity Methods* pada penelitian ini sebagai contoh digunakan dokumen ke-132. Didapatkan 10 dokumen dengan nilai kemiripan tertinggi yaitu dokumen ke-170, 303, 210, 360, 315, 422, 510, 45, 63, dan 223. Dokumen tersebut memiliki nilai *Cosine Similarity* secara berturut-turut sebesar 0.956666513, 0.946570717, 0.917459394, 0.912534877, 0.886439718, 0.88221982, 0.881775275, 0.875845794, 0.872030219, dan 0.871514859.
2. Alur memberikan rekomendasi hotel di Yogyakarta dengan *Cosine Similarity Methods* dan *Content-Based Filtering Methods* adalah melakukan *preprocessing* data deskripsi dari, melakukan pembobotan kata dengan TF-IDF, dilanjutkan dengan menghitung nilai kemiripan antar dokumen yang kemudian nilai tersebut diurutkan dari nilai terbesar dan mengambil 10 urutan pertama.
3. Hasil rekomendasi hotel di Yogyakarta dengan *Content-Based Filtering Methods* yang dihasilkan dalam penelitian ini berdasarkan kemiripan deskripsi hotel adalah 10 hotel teratas dengan nilai kemiripan yang terbesar.
4. Langkah membangun aplikasi *website* dari sistem rekomendasi berdasarkan model yang diperoleh adalah dengan menggunakan *web framework* berbahasa *Python* yaitu *Flask* yang kemudian di-*deploy* pada Heroku agar dapat diakses secara online.

6.2 Saran

Beberapa saran yang peneliti berikan berdasarkan penelitian ini adalah sebagai berikut.

1. Untuk penelitian selanjutnya dapat mengembangkan atau membandingkan metode kemiripan dokumen yang lain selain metode *Cosine Similarity*.
2. Penelitian ini hanya menggunakan konten dari item sebagai dasar rekomendasi, sehingga untuk penelitian selanjutnya bisa mengkombinasikan faktor lain sebagai dasar rekomendasi.
3. Pada penelitian selanjutnya diharapkan dapat melakukan penyempurnaan pada proses *preprocessing data* terutama pada bagian penghapusan kata yang tidak memiliki makna atau *remove stopwords* dengan memperhatikan kata-kata yang kiranya perlu dihilangkan.
4. Pada penelitian selanjutnya dapat mempertimbangkan penggunaan multivariabel untuk memberikan rekomendasi serta mengembangkan algoritma multivariabel.

DAFTAR PUSTAKA

- Aamir, M., & Bhusry, M. (2015). Recommendation System: State of the Art Approach. *International Journal of Computer Applications*, 120(12), 25–32.
- Ahsan. (2020.) Staycation Kala Pandemi, Bandung dan Yogyakarta Jadi Incaran. Diambil kembali dari <https://travel.wego.com/berita/staycation-kala-pandemi-bandung-dan-yogyakarta-jadi-incaran/>
- Alkaff, M., Khatimi, H., & Eriadi, A. (2020). Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 193–202.
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19–27.
- Arekar, T., Sonar, M. R. S., & Uke, D. N. J. (2015). A Survey on Recommendation System. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, 2(1), 6–10.
- Asrianti, S., & Azizah, N. (2020). Pandemi Corona Ubah Kecenderungan Wisatawan Pilih Akomodasi. Diambil kembali pada <https://www.republika.co.id/berita/qc6j66463/pandemi-corona-ubah-kecenderungan-wisatawan-pilih-akomodasi>
- BPS. (2020). Tingkat Penghunian Kamar Hotel Daerah Istimewa Yogyakarta. Yogyakarta: Badan Pusat Statistik
- B.Thorat, P., M. Goudar, R., & Barve, S. (2015). Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System. *International Journal of Computer Applications*, 110(4), 31–36.
- Das, D., Sahoo, L., & Datta, S. (2017). A Survey on Recommendation System. *International Journal of Computer Applications*, 160(7), 6–10.
- Devi, A. A. P., & Tonora, D. B. (2015). Rancang Bangun Recommender System

- dengan Menggunakan Metode Collaborative Filtering untuk Studi Kasus Tempat Kuliner di Surabaya. *JUI SI Jurnal Informatika Dan Sistem Informasi*, 01, 102–112.
- Fiarni, C., Maharani, H., & Calista, N. (2019). Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 3(2), 42.
- Fitri, E., Lumbanraja, F. R., & Ardiansyah, A. (2020). Klasifikasi Abstrak Jurnal Komputasi Menggunakan Metode Text Mining Dan Algoritma Support Vector Machine. *Jurnal Pepadun*, 1(1), 83–88.
- Hendriyati, L. (2019). Online travel agent. *Jurnal Media Wisata*, 17(1), 1–10.
- Husni, Negara, Y. D. P., & Syarief, M. (2015). Clusterisasi Dokumen Web (Berita) Bahasa Indonesia Menggunakan Algoritma K-Means (Clustering Web Documents (News) Indonesian Language Using K-Means Algorithm). *Jurnal Simantec*, 4(3), 159–166.
- Irsyad, R. (2018). *Penggunaan Python Web Framework Flask Untuk Pemula*.
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2011). *Recommender systems: an introduction*. Cambridge University Press.
- Kao, A., & Poteet, S. R. (2007). *Natural Language Processing and Text Mining*. Springer Science+Business Media.
- Kusuma, H. S., & Musdholifah, A. (2021). Recommendation System for Thesis Topics Using Content-based Filtering. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(1), 65.
- Larasati, F. B. A., & Februariyanti, H. (2021). Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content - Based Filtering. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 45.
- Layton, Julia. 2009. What is a Staycation. Diambil kembali pada <https://money.howstuffworks.com/personal-finance /budgeting/staycation.htm>
- Mondi, R. H., Wijayanto, A., & Winarno. (2019). Recommendation System With

- Content-Based Filtering Method for Culinary Tourism in Mangan Application. *Itsmart*, 8(2), 65–72.
- Nisrina, F. D. (2020). Implementasi Deteksi Topik Putusan Hakim dengan Latent Dirichlet Allocation (LDA)(Studi terhadap Putusan Tindak Pidana di Pengadilan Negeri Sleman, DI Yogyakarta). *Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Indonesia*.
- Nugraha, K. A., & Sebastian, D. (2018). Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity. *Jurnal Teknik Informatika Dan Sistem Informasi*, 4, 376–386.
- Nurjannah, M., & Fitri Astuti, I. (2013). Penerapan Algoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining. Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. *Jurnal Informatika Mulawarman*, 8(3), 110–113.
- Pasaribu, J. (2019). Penetapan Harga Kamar Batiqa Hotel Pekanbaru. *JOM FISIP*, 6(July), 1–23.
- Prasetya, C. S. D. (2017). Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 194.
- Prasetya, Y. N., Winarso, D., & Syahril. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19. *Junal Fasilkom*, 11(2), 97–103.
- Purnaramadhan, R. (2021). Recommendation System Model Untuk Merekomendasikan Produk Pada Website Menggunakan Metode Content-Based. *Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Indonesia*.
- Raghuwanshi, S. K., & Pateriya, R. K. (2019). Recommendations System: Techniques, Challenges, Application, and Evaluation. *Advances in Intelligent Systems and Computing*, 2(January), 107–119.
- Ricci, F., Rokch, L., & Shapira, B. (2015). *Recommender Systems Handbook* (F. Ricci, L. Rokach, & B. Shapira (eds.); Second Edi). Springer Science+Business

Media.

- Rismanto, R., Syulistyo, A. R., & Agusta, B. P. C. (2020). Research supervisor recommendation system based on topic conformity. *International Journal of Modern Education and Computer Science*, 12(1), 26–34.
- Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31.
- Sandi Wachyuni, S., & Wiweka, K. (2020). Kepuasan Wisatawan Dalam Penggunaan E-Commerce Agoda Dalam Pemesanan Hotel. *Journal of Tourism Destination and Attraction*, 8(1), 275.
- Shah, L., Gaudani, H., & Balani, P. (2016). Survey on Recommendation System. *International Journal of Computer Applications*, 137(7), 43–49.
- Siregar, R. R. A., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Computatio : Journal of Computer Science and Information Systems*, 1(2), 171.
- Supriyati, E., & Iqbal, M. (2018). Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 361–370.
- Wahyuni, R. T., Prastiyanto, D., & Suprptono, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro*, 9(1), 18–23.
- Yasni, L., Subroto, I. M. I., & Haviana, S. F. C. (2018). Implementasi Cosine Similarity Matching Dalam Penentuan Dosen Pembimbing Tugas Akhir. *Transmisi*, 20(1), 22.
- Yusuf, M., & Cherid, A. (2020). Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar. 9, 8–16.

LAMPIRAN

Lampiran 1 *Dataset*

<https://bit.ly/DatasetHotel>

Lampiran 2 Perhitungan Manual TF-IDF dan *Cosine Similarity*

<https://bit.ly/PerhitunganManualTF-IDFdanCosim>

Lampiran 3 *Script Python* Sistem Rekomendasi

<https://bit.ly/SyntaxPython>

Lampiran 4 *Script Flask* dan *HTML*

<https://github.com/cherylam/sistemrekomendasi.git>