

**IMPLEMENTASI TEXT MINING DALAM PERBAIKAN KUALITAS RESTORAN
BERDASARKAN ULASAN PELANGGAN**

(Studi Kasus: Gudeg Yu Djum)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1

Program Studi Teknik Industri - Fakultas Teknologi Industri

Universitas Islam Indonesia



Disusun Oleh:

Nama : Kevin Vieri El Sabri

No. Mahasiswa : 19 522 151

**PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2023

PERNYATAAN KEASLIAN

Demi Allah saya akui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 29 Mei 2023



Kevin Vieri El Sabri
19 522 151

SURAT BUKTI PENELITIAN



FAKULTAS
TEKNOLOGI INDUSTRI

Gedung KH. Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kalitirang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. fti@uii.ac.id
W. fti.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor : 250/Ka.Lab.Datmin/70/Lab.Datmin/VIII/2023

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Kevin Vieri El Sabri
No. Mhs : 19522151
Dosen Pembimbing : Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

Telah selesai melaksanakan penelitian yang berjudul Implementasi Text Mining Dalam Perbaikan Kualitas Restoran Berdasarkan Ulasan Pelanggan (Studi Kasus: Gudeg Yu Djum)" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 5 Mei sampai dengan tanggal 5 Juni 2023.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 07 Agustus 2023

Kepala Laboratorium
Data Mining

Annisa Uswatun Khasanah, ST., M.B.A., M.Sc.

LEMBAR PENGESAHAN PEMBIMBING

**IMPLEMENTASI TEXT MINING DALAM PERBAIKAN KUALITAS RESTORAN
BERDASARKAN ULASAN PELANGGAN
(Studi Kasus: Gudeg Yu Djum)**

TUGAS AKHIR

Disusun Oleh :

Nama : Kevin Vieri El Sabri

No. Mahasiswa : 19 522 151

Yogyakarta, 29 Mei 2023

Dosen Pembimbing



Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

LEMBAR PENGESAHAN DOSEN PENGUJI

**IMPLEMENTASI TEXT MINING DALAM PERBAIKAN KUALITAS RESTORAN
BERDASARKAN ULASAN PELANGGAN
(Studi Kasus: Gudeg Yu Djum)**

TUGAS AKHIR

Disusun Oleh :

Nama : Kevin Vieri El Sabri

No. Mahasiswa : 19 522 151

**Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk
memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri
Universitas Islam Indonesia**

Yogyakarta, 06 September 2023

Tim Penguji

**Annisa Uswatun Khasanah, S.T., M.B.A.,
M.Sc.**

Ketua



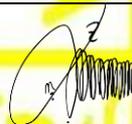
Danang Setiawan, S.T., M.T.

Anggota I



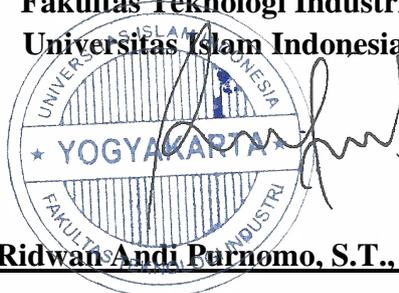
Ir. Abdullah 'Azzam, S.T., M.T., IPM

Anggota II



Mengetahui,

**Ketua Program Studi Teknik Industri Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia**



Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM

HALAMAN PERSEMBAHAN

*Dengan Menyebut Nama Allah yang Maha Pengasih dan Maha Penyayang.
Puji syukur atas kehadiran Allah SWT atas nikmat yang telah diberikan. Skripsi ini saya
persembahkan kepada:*

*Bapak Sabrioda dan Ibu Restu Sundari selaku kedua orang tua saya yang tiada henti
mendo'akan, memberikan dukungan semangat, dan mendidik saya dari kecil hingga saat ini.
Terima kasih atas semua pengorbanan yang telah diberikan yang tidak dapat diungkapkan
melalui kata-kata dan tak akan pernah bisa dibalas oleh apapun*

*Teruntuk keluarga besar dan adik saya Gafri, terimakasih atas do'a, dukungan, dan motivasi
yang diberikan hingga pada tahap ini. Serta sahabat dan rekan-rekan saya yang telah
membantu dan menemani hari-hari saya selama masa perkuliahan.*

MOTTO

“Barangsiapa menempuh jalan untuk mendapatkan ilmu, Allah akan memudahkan baginya jalan menuju surga.” (HR. Muslim)

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahirabbil'alamin, puji dan syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul "IMPLEMENTASI TEXT MINING DALAM PERBAIKAN KUALITAS RESTORAN BERDASARKAN ULASAN PELANGGAN (Studi Kasus: Gudeg Yu Djum)" dengan baik dan tepat pada waktunya. Shalawat serta salam selalu tercurah kepada baginda Rasulullah yakni Nabi Muhammad SAW yang senantiasa menjadi sumber inspirasi dan teladan bagi umat manusia.

Laporan Tugas Akhir ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Teknik di Fakultas Teknologi Industri Universitas Islam Indonesia. Dalam penyusunan Tugas Akhir ini, penulis tentunya mendapatkan banyak bimbingan dan dukungan dari berbagai pihak. Untuk itu, penulis ingin mengucapkan terimakasih kepada:

1. Bapak Prof., Dr., Ir., Hari Purnomo, M.T., IPU., ASEAN.Eng. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. selaku Ketua Program Studi Teknik Industri Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Ibu Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc. selaku dosen pembimbing penulis yang telah memberikan arahan serta bimbingan.
4. Kedua orang tua, dan adik saya untuk semua do'a, dukungan, dan motivasi yang diberikan kepada saya sehingga penyusunan laporan Tugas Akhir ini dapat terlaksana dengan baik dan tepat pada waktunya.
5. Sahabat penulis, *Duntil's family* (Qowi, Rizha, Ilham, Imam, Daffa, Oka, Dio, Tiara, Gadis, dan Fairuz) yang senantiasa dengan senang hati berbagi ilmu, memberikan dukungan dan mendengarkan segala keluh kesah. Terimakasih atas segala dukungan dan motivasi yang diberikan kepada penulis.
6. Teman-teman Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia, khususnya angkatan 2019.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan. Oleh sebab itu, penulis berharap mendapatkan kritik dan saran yang bersifat membangun untuk penulisan yang lebih baik di masa yang akan datang. Akhir kata, semoga laporan Tugas Akhir ini dapat bermanfaat bagi semua pihak.

Yogyakarta, 29 Mei 2023



Kevin Vieri El Sabri
19 522 151

ABSTRAK

Yogyakarta merupakan kota yang dikenal dengan berbagai kuliner khasnya. Salah satu restoran yang terkenal dalam menyajikan masakan khas daerah yaitu restoran Gudeg Yu Djum. Gudeg Yu Djum merupakan salah satu restoran yang berdiri pada tahun 1950 dan bergerak di bidang kuliner khas daerah berupa makanan gudeg. Restoran ini memiliki ulasan di situs *travel* terbesar di dunia yaitu TripAdvisor. Di dalam situs tersebut, terdapat 675 ulasan Bahasa Indonesia yang dibuat oleh pengunjung mengenai restoran Gudeg Yu Djum. Ulasan yang diberikan oleh pelanggan mengenai restoran beraneka macam, mulai dari ulasan positif hingga ulasan negatif. Dilakukan penelitian terkait ulasan restoran Gudeg Yu Djum dengan pendekatan aplikasi *text mining* yaitu analisis sentimen. Metode yang digunakan dalam menganalisis data yaitu *K-Means Clustering* dengan tujuan membuat klasterisasi dari data ulasan restoran Gudeg Yu Djum, mengidentifikasi ulasan yang bersifat negatif, dan menganalisis ulasan tersebut dengan diagram *fishbone* untuk perbaikan kualitas layanan restoran. Hasil dari penelitian ini yaitu terbentuknya *cluster* data ulasan restoran sebanyak 3 *cluster* dengan topik pada masing-masing *cluster* yaitu *cluster* 0 yang merupakan kelompok ulasan mengenai rasa makanan, pelayanan, harga, dan kenyamanan. *Cluster* 1 merupakan kelompok ulasan mengenai menu, rasa, harga, dan eksistensi restoran. Terakhir *cluster* 2 merupakan kelompok ulasan yang membahas tentang rasa dan tekstur dari gudeg dan krecek. Hasil *cluster* yang terbentuk kemudian dilakukan proses visualisasi *wordcloud* dan asosiasi kata menggunakan metode *association rule* agar mendapatkan informasi yang lebih akurat. Dari hasil asosiasi kata, didapatkan aturan asosiasi kata yang bersifat negatif. Aturan asosiasi kata yang bersifat negatif tersebut dianalisis menggunakan diagram *fishbone* dan terdeteksi 7 permasalahan yang terjadi pada restoran Gudeg Yu Djum yang terbagi menjadi 5 aspek yaitu *place, process, product, people, dan price*.

Kata Kunci: Analisis Sentimen, *Text Mining*, *Clustering*, *K-Means Clustering*, *Association Rule*, Diagram *Fishbone*, Gudeg Yu Djum

DAFTAR ISI

PERNYATAAN KEASLIAN	ii
SURAT BUKTI PENELITIAN.....	iii
LEMBAR PENGESAHAN PEMBIMBING	iv
LEMBAR PENGESAHAN DOSEN PENGUJI	v
HALAMAN PERSEMBAHAN.....	vi
MOTTO.....	vii
KATA PENGANTAR	viii
ABSTRAK.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	7
1.3 Tujuan Penelitian.....	7
1.4 Manfaat Penelitian.....	7
1.5 Batasan Penelitian.....	8
BAB II TINJAUAN PUSTAKA.....	9
2.1 Kajian Induktif.....	9
2.2 Kajian Deduktif	19
2.2.1 Text Mining	19
2.2.2 Term Frequency-Inverse Document Frequency (TF-IDF).....	20
2.2.3 Clustering.....	21
2.2.4 Machine Learning.....	22
2.2.5 K-Means Clustering.....	23
2.2.6 Metode Elbow.....	24
2.2.7 Association Rule.....	24
2.2.8 Fishbone Diagram.....	25
BAB III METODE PENELITIAN	27
3.1 Objek dan Subjek Penelitian.....	27
3.2 Populasi dan Sampel Penelitian.....	27
3.3 Metode Pengumpulan Data	28

3.4	Jenis dan Sumber Data	28
3.5	Metode Analisis Data	28
3.6	Alur Penelitian	29
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA		34
4.1	Pengumpulan Data.....	34
4.2	Pengolahan Data	39
4.2.1	Analisis Deskriptif.....	39
4.2.2	Preprocessing Data	40
4.2.3	Clustering dengan Algoritma K-Means Clustering	43
4.2.4	Visualisasi Cluster	45
4.2.5	Asosiasi Kata	47
4.2.6	Analisis Diagram Fishbone	52
BAB V PEMBAHASAN		54
5.1	Hasil Analisa Deskriptif Ulasan Restoran Gudeg Yu Djum	54
5.2	Hasil Klasterisasi Ulasan Menggunakan <i>K-Means Clustering</i>	55
5.3	Hasil Analisa Profil <i>Cluster</i>	56
5.4	Hasil Usulan Perbaikan Masalah	58
BAB VI PENUTUP		60
6.1	Kesimpulan.....	60
6.2	Saran	61
DAFTAR PUSTAKA		63
LAMPIRAN.....		A-1

DAFTAR TABEL

Tabel 2.1 Rangkuman Kajian Induktif	13
Tabel 4.1 Hasil <i>Scraping Data</i>	36
Tabel 4.2 Proses <i>Case Folding</i>	41
Tabel 4.3 Proses <i>Tokenizing</i>	42
Tabel 4.4 Proses <i>Filtering</i>	42
Tabel 4.5 Proses <i>Stemming</i>	43
Tabel 4.6 Hasil Perbandingan Nilai <i>Support</i> dan <i>Confidence</i>	48
Tabel 4.7 Aturan Asosiasi <i>Cluster 0</i>	48
Tabel 4.8 Aturan Asosiasi <i>Cluster 1</i>	49
Tabel 4.9 Aturan Asosiasi <i>Cluster 2</i>	51
Tabel 4.10 Permasalahan pada Restoran Gudeg Yu Djum	53
Tabel 5.1 Kata Kunci <i>Cluster</i>	56
Tabel 5.2 Rancangan Usulan Perbaikan Masalah	58

DAFTAR GAMBAR

Gambar 1.1 Profil Pengguna Internet di Indonesia	2
Gambar 1.2 Situs Populer Kategori <i>Travel and Tourism</i>	3
Gambar 1.3 Rekomendasi Restoran Terbaik di Yogyakarta.....	4
Gambar 1.4 Ulasan Pelanggan Restoran Gudeg Yu Djum	5
Gambar 2.1 Proses <i>Text Mining</i>	19
Gambar 2.2 <i>Fishbone Diagram</i>	26
Gambar 3.1 Alur Penelitian.....	30
Gambar 4.1 <i>Interface</i> Aplikasi WebHarvy	35
Gambar 4.2 Proses <i>Record</i> Variabel Data Ulasan.....	36
Gambar 4.3 Grafik Persentase <i>Rating</i> Ulasan	39
Gambar 4.4 Grafik <i>Elbow</i>	44
Gambar 4.5 Grafik <i>Clustering</i> Data Ulasan	45
Gambar 4.6 Visualisasi <i>Wordcloud Cluster 0</i>	46
Gambar 4.7 Visualisasi <i>Wordcloud Cluster 1</i>	46
Gambar 4.8 Visualisasi <i>WordCloud Cluster 2</i>	47
Gambar 4.9 Analisis Diagram <i>Fishbone</i>	53

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi merupakan suatu teknologi yang digunakan dalam memproses dan mengolah data dengan berbagai cara untuk mendapatkan informasi yang berkualitas yaitu yang bersifat relevan, akurat dan *real time*, yang digunakan untuk berbagai macam keperluan, baik keperluan pribadi, bisnis, maupun instansi pemerintahan dan informasi yang dihasilkan bersifat strategis yang digunakan untuk pengambilan keputusan (Cholik, 2021). Perkembangan teknologi saat ini mendatangkan zaman serba digital (*Digital Technology*) yang dimana proses kehidupan yang serba instan, segala sesuatu serba cepat seperti persebaran informasi dapat tersebar dalam hitungan detik (Megawati, 2021). Selain mendatangkan zaman serba digital, perkembangan teknologi ini juga melahirkan beberapa inovasi di bidang teknologi sehingga teknologi informasi memiliki pengaruh yang besar dalam berbagai aspek kehidupan saat ini karena dapat memenuhi kebutuhan dan juga memudahkan pekerjaan manusia (Siregar & Nasution, 2020).

Salah satu contoh dari perkembangan teknologi informasi yaitu adanya Internet. Internet merupakan sebuah jaringan komputer yang terhubung menggunakan sistem standar TCP/IP yang berfungsi sebagai protokol dalam pertukaran paket data yang dapat melayani miliaran pengguna di seluruh dunia (Nasution et al., 2020). Saat ini, perkembangan internet di Indonesia dapat dikatakan cukup pesat. Menurut survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) jumlah penduduk yang terkoneksi dengan Internet pada tahun 2022-2023 yakni sebanyak 215 juta jiwa dengan tingkat penetrasi internet di Indonesia pada tahun 2023 yaitu sebesar 78,19% (Asosiasi Penyelenggara Jasa Internet Indonesia, 2023).



Gambar 1.1 Profil Pengguna Internet di Indonesia

Sumber: APJII (2023)

Jumlah pengguna internet di Indonesia meningkat setiap tahun nya. Hal ini ditandai dengan bertambahnya jumlah pengguna internet pada tahun 2022-2023 yakni sebanyak 5 juta jiwa. Terjadi peningkatan persentase penetrasi internet Indonesia pada tahun 2022-2023 sebesar 1,17%. Meningkatnya pengguna internet di Indonesia diikuti dengan berkembangnya media *online* di Indonesia (Adriansah & Santoso, 2019). Media *online* atau sosial media merupakan wadah yang dijadikan oleh pengguna internet untuk mengutarakan opini mereka terhadap suatu hal (Amardita et al., 2022). Promosi dan penyebaran informasi tentang destinasi wisata kini semakin bergantung pada media *online*. Pada umumnya, sebelum memilih tujuan wisata, masyarakat akan selalu mencari informasi terkait tujuan wisata yang akan dikunjungi untuk dijadikan sebagai bahan pertimbangan. Saat ini, masyarakat dapat saling bertukar informasi mengenai objek wisata melalui internet, seperti melalui media sosial, *travel blog*, dan situs ulasan objek wisata (Af'idah et al., 2022). Untuk menemukan informasi mengenai objek wisata yang akan dikunjungi, sebagian besar wisatawan mencari informasi tersebut melalui situs *website* seperti TripAdvisor (Azzahra & Wibowo, 2020). TripAdvisor merupakan situs digital berbasis *website* yang dirancang khusus untuk memberikan panduan terkait perjalanan wisata melalui informasi tentang akomodasi hotel, restoran dan jasa transportasi ke lokasi destinasi yang diinginkan oleh wisatawan (Singgalen, 2022). Menurut survei yang dilakukan oleh similarweb.com pada tahun 2023, terdapat beberapa situs yang sering dikunjungi oleh wisatawan yang dapat diilustrasikan pada gambar 1.2 berikut.

Rank	Website	Category	Rank Change	Avg. Visit Duration	Pages / Visit	Bounce Rate
1	 tripadvisor.com	Travel and Tourism > Travel and Tourism - Other	=	00:03:02	4.16	57.89%
2	 navitime.co.jp	Travel and Tourism > Travel and Tourism - Other	=	00:02:40	3.17	53.93%
3	 irctc.co.in	Travel and Tourism > Travel and Tourism - Other	- 3	00:07:33	5.70	28.09%
4	 makemytrip.com	Travel and Tourism > Travel and Tourism - Other	- 1	00:04:57	5.03	42.20%
5	 kayak.com	Travel and Tourism > Travel and Tourism - Other	- 2	00:04:27	5.63	35.10%
6	 rome2rio.com	Travel and Tourism > Travel and Tourism - Other	- 1	00:02:40	3.43	27.65%
7	 tutu.ru	Travel and Tourism > Travel and Tourism - Other	- 2	00:04:17	4.10	38.58%

Gambar 1.2 Situs Populer Kategori *Travel and Tourism*
Sumber: Similar Web (2023)

Situs TripAdvisor menduduki peringkat 1 di dunia dalam kategori situs travel dan turis, dengan rata-rata kunjungan per halaman sebesar 4,16. Kemudian merujuk pada survei yang dilakukan oleh perusahaan TripAdvisor bahwa perusahaan unggul dalam kualitas ulasan yang dimiliki. Sebanyak 88% responden survei menilai ulasan yang dimiliki oleh situs tersebut memiliki kualitas tinggi dan sebagian wisatawan merujuk ke TripAdvisor untuk membuat keputusan mengenai destinasi wisata yang akan dikunjungi (TripAdvisor, 2023).

Di Indonesia, Yogyakarta dikenal dengan berbagai kuliner khasnya. Salah satu jenis makanan daerah yang menjadi simbol Yogyakarta adalah gudeg. Nangka muda dimasak dengan santan, rempah-rempah, dan gula merah untuk membuat gudeg. Rasa gudeg yang manis, gurih dan lezat menjadikannya makanan khas favorit di lidah masyarakat Yogyakarta. Gudeg lebih dari sekedar makanan yang dimakan penduduk setempat, gudeg ini juga menjadi tujuan wisata populer bagi wisatawan yang ingin mencoba kuliner khas Yogyakarta.

Salah satu tujuan destinasi wisata paling populer di Indonesia, Yogyakarta juga dikenal memiliki beragam restoran dengan pengalaman kuliner yang mewah dan lezat. Daerah Istimewa Yogyakarta (DIY) memiliki banyak jenis restoran, mulai dari restoran tradisional hingga restoran mewah yang memadukan citarasa tradisional dan internasional. Berdasarkan survei yang dilakukan oleh *mybest*, terdapat beberapa rekomendasi restoran terbaik di Daerah Istimewa Yogyakarta.

Popularity				Filter				
Produk	Gambar	Harga terendah	Poin	Perincian				
				Wilayah	Rating di Google	Jenis makanan	Cocok dikunjungi...	
1 MRBYKAMIL Mediterranean Restaurant		Mediterranean	Trip Advisor	Hadirkan cita rasa Prancis yang autentik dari chef ternama	Jl. Tirtodipuran No. 24A,...	4,7 dari 5	Makanan khas luar negeri...	Keluarga, teman, pasangan,...
2 Gudeg Yu Djum Pusat		Gudeg Yu Djum Pusat	Trip Advisor	Rumah makan gudeg terkenal yang sudah ada sejak 1951	Jl. Kaliurang Km. 4,5 Karangasem...	4,6 dari 5	Makanan khas Jogja	Keluarga, teman
3 Mang Engking... Gubug Makanan Mang Engking		Gubug Makanan	Trip Advisor	Arena bermain luas, anak pun bisa bermain dengan puas	Jl. Gedongan-Tempel, Jamur,...	4,6 dari 5	Makanan khas Indonesia...	Keluarga, rombongan besar
4 Jejamuran		Jejamuran	Trip Advisor	Nikmati berbagai olahan jamur yang unik dan lezat	Jl. Pandowoharjo, Niron,...	4,6 dari 5	Makanan khas Indonesia	Keluarga, teman, kolega rombongan...
5 The House Of Raminten		The House of...	Trip Advisor	Tawarkan konsep restoran unik yang dipenuhi koleksi antik	Jl. Faridan M Noto No. 7, Kotabaru,...	4,5 dari 5	Makanan khas Indonesia	Keluarga, teman

Gambar 1.3 Rekomendasi Restoran Terbaik di Yogyakarta
Sumber: MyBest (2023)

Gudeg Yu Djum menempati posisi kedua dalam rekomendasi restoran terbaik yang terdapat di Daerah Istimewa Yogyakarta. Penilaian yang dilakukan berdasarkan *rating* dan *review* dari pelanggan. Gudeg Yu Djum merupakan salah satu restoran yang berdiri pada tahun 1950 dan bergerak di bidang kuliner khas daerah yang berlokasi di Daerah Istimewa Yogyakarta. Produk yang ditawarkan berupa makanan khas daerah seperti Nasi Gudeg, Gudeg Kendil, Gudeg Besek, dan lain-lain.

Restoran merupakan salah satu objek yang banyak di *review* di media *online* berdasarkan pengalaman pribadi seseorang (Indrayuni, 2017). Di dalam situs TripAdvisor, terdapat 586 data historis ulasan Bahasa Indonesia pada bulan Juli 2012 - November 2022 yang dibuat oleh pelanggan mengenai restoran Gudeg Yu Djum. Dengan banyaknya ulasan yang ada, akan membuat suatu restoran kesulitan untuk mengetahui apakah restoran tersebut memiliki lebih banyak ulasan positif atau negatif (Julianto et al., 2022). Ulasan yang diberikan oleh pelanggan mengenai restoran Gudeg Yu Djum beraneka ragam, mulai dari ulasan positif hingga ulasan negatif. Berdasarkan pengamatan melalui situs TripAdvisor, terdapat beberapa ulasan negatif yang diberikan oleh pelanggan. Mayoritas pelanggan mengeluhkan dari segi rasa gudeg terlalu manis, kemudian dari segi pelayanan kurang ramah, dan harga yang relatif cukup mahal dari gudeg lainnya.



Gambar 1.4 Ulasan Pelanggan Restoran Gudeg Yu Djum

Sumber: TripAdvisor (2023)

Berdasarkan permasalahan diatas, maka perlu dilakukan penelitian tentang ekstraksi data ulasan pelanggan restoran Gudeg Yu Djum untuk membantu pengelola restoran dalam menemukan informasi yang berguna untuk meningkatkan kualitas layanan restoran. Penilaian wisatawan terhadap objek wisata menjadi suatu hal yang sangat penting karena dengan adanya penilaian tersebut dapat dijadikan sebagai pendukung keputusan baik untuk wisatawan maupun pengelola objek wisata tersebut (Somantri, 2019).

Analisis permasalahan pada penelitian ini menggunakan pendekatan *Text Mining*. *Text Mining* merupakan metode algoritmik berbasis komputer untuk mengungkap informasi yang sebelumnya tidak diketahui dalam kumpulan teks (Priyanto & Ma'arif, 2018). Untuk mengetahui kepuasan pelanggan terhadap kualitas layanan yang diberikan, dapat dilihat melalui opini yang diberikan oleh pelanggan. Salah satu aplikasi *text mining* dalam menganalisis opini yaitu analisis sentimen. Analisis sentimen, juga disebut sebagai penambangan opini merupakan metode pemrosesan bahasa alami (NLP) untuk menentukan nada emosional teks. Metode ini banyak digunakan untuk mendapatkan umpan balik pengguna tentang produk dan layanan dari ulasan dan membuat informasi yang berguna bagi suatu organisasi (Shaik et al., 2023). Analisis sentimen pada penelitian ini menggunakan pendekatan *Clustering*. Digunakan pendekatan *clustering* karena proses analisis sentimen yang dilakukan tidak berdasarkan spesifik label (Wijaya et al., 2021).

Terdapat beberapa algoritma dalam *clustering* seperti *K-Means Clustering*, *DBSCAN*, *Affinity Propagation Algorithm*, dan lain-lain. Pada penelitian ini algoritma *K-Means Clustering* dipilih karena algoritma ini mudah untuk diimplementasikan, diadaptasi,

dan juga waktu pembelajaran yang dibutuhkan relatif cepat (Utami & Eka Putra, 2022). Algoritma *K-Means Clustering* cukup banyak digunakan untuk mengelompokkan data tanpa label ke dalam *cluster* dengan teknik partisi untuk menganalisis data dimana data yang diamati ditempatkan dalam sebuah *cluster* berdasarkan seberapa dekat dengan *centroid*, atau juga lebih dikenal dengan pusat *cluster* (Sanko et al., 2023). Pemilihan algoritma ini juga didasarkan karena pada algoritma DBSCAN, memiliki kompleksitas komputasi yang tinggi (Salman, 2023). Kemudian Algoritma *Affinity Propagation* tidak selalu menghasilkan *cluster* ideal karena memiliki keterbatasan yaitu sulit untuk menentukan nilai *preference* terbaik untuk menghasilkan *cluster* yang optimal (Izzatillah & Mutiara, 2020).

Beberapa penelitian terdahulu telah dilakukan mengenai analisis sentimen data ulasan dengan pendekatan *clustering*. Penelitian yang dilakukan oleh Saputra & Arianty (2019) mengenai implementasi algoritma *K-Means Clustering* pada analisis sentimen keluhan pengguna Indosat. Hasil dari penelitian ini menunjukkan bahwa algoritma tersebut mampu mengelompokkan data ulasan menjadi 3 *cluster*, *cluster* 1 membahas tentang jaringan indosat yang parah, *cluster* 2 membahas tentang permintaan perbaikan jaringan sinyal indosat, dan terakhir *cluster* 3 membahas tentang jaringan sinyal Indosat parah di daerah Bogor. Kemudian penelitian yang dilakukan oleh Safitri, et al. (2021) mengenai implementasi algoritma *K-Means Clustering* untuk *clustering* sentimen pada opini kualitas pelayanan jasa penerbangan. Hasil dari penelitian ini menunjukkan presentase keberhasilan algoritma *K-Means Clustering* dalam mengelompokkan data ulasan maskapai Garuda Indonesia sebesar 71,8%, Air Asia sebesar 52,8% dan Lion Air sebesar 62,1%. Kemudian dari penelitian ini juga menghasilkan 2 profil *cluster*, yaitu *cluster* ulasan positif dan *cluster* ulasan negatif. Maskapai Garuda Indonesia terdiri dari 116 ulasan positif dan 416 ulasan negatif, Lion Air terdiri dari 17 ulasan positif dan 102 ulasan negatif, dan terakhir Air Asia terdiri dari 62 ulasan positif dan 347 ulasan negatif. Berdasarkan hasil penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma *K-Means Clustering* cukup baik dalam melakukan pengelompokan data ulasan.

Setelah dilakukan proses klasterisasi data ulasan, kemudian dilakukan visualisasi *cluster* ulasan menggunakan visualisasi *Word Cloud*. Pada *cluster* yang memiliki kata ulasan yang bersifat negatif akan dianalisis menggunakan metode diagram *fishbone*. Diagram *fishbone* digunakan untuk mengidentifikasi akar permasalahan dari hasil klasterisasi yang kemudian diberikan usulan perbaikan atas permasalahan tersebut.

Penelitian ini diharapkan dapat memaparkan klasterisasi data ulasan restoran dengan baik dan dapat memberikan informasi yang berguna untuk evaluasi terkait pelayanan restoran Gudeg Yu Djum.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah pada penelitian ini yaitu sebagai berikut.

1. Bagaimana hasil klasterisasi yang didapat pada ulasan restoran Gudeg Yu Djum dengan pendekatan *K-Means Clustering*?
2. Faktor apa saja yang perlu diperbaiki dalam meningkatkan kualitas layanan restoran Gudeg Yu Djum berdasarkan hasil klasterisasi yang didapat?
3. Apa usulan perbaikan yang dapat diberikan guna meningkatkan kualitas layanan restoran Gudeg Yu Djum berdasarkan hasil klasterisasi yang didapat?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu sebagai berikut.

1. Untuk mengetahui hasil klasterisasi data ulasan restoran Gudeg Yu Djum dengan pendekatan *K-Means Clustering*.
2. Untuk mengetahui faktor apa saja yang perlu diperbaiki dalam meningkatkan kualitas layanan restoran Gudeg Yu Djum berdasarkan hasil klasterisasi data ulasan.
3. Untuk memberikan usulan perbaikan yang sesuai guna meningkatkan kualitas layanan restoran Gudeg Yu Djum berdasarkan hasil klasterisasi data ulasan.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini yaitu sebagai berikut.

1. Bagi Akademisi
 - a. Memberikan informasi mengenai gambaran umum pengalaman pengunjung terhadap restoran Gudeg Yu Djum.
 - b. Dapat menjadi referensi untuk penelitian selanjutnya, terkait ekstraksi data ulasan untuk perbaikan kualitas dengan pendekatan *Text Mining* dan *Clustering*.
2. Bagi Restoran

- a. Hasil ekstraksi data ulasan dapat dijadikan sebagai pendukung keputusan bagi restoran untuk mengatasi permasalahan terkait ulasan negatif yang diberikan oleh pelanggan terhadap kualitas layanan.
- b. Dapat memberikan informasi mengenai keluhan-keluhan yang dialami oleh pelanggan.

1.5 Batasan Penelitian

Penelitian ini memiliki batasan penelitian dengan tujuan agar penelitian yang dilakukan memiliki arah dan tujuan yang jelas. Adapun batasan pada penelitian ini yaitu sebagai berikut.

1. Penelitian ini menggunakan metode *web scraping* untuk pengambilan data dengan bantuan aplikasi yaitu WebHarvy versi 5.5.0.168.
2. Data yang digunakan untuk penelitian ini yaitu data historis ulasan restoran Gudeg Yu Djum sebanyak 586 ulasan pada bulan Juli 2012 – November 2022 yang diambil dari situs TripAdvisor.
3. Data ulasan yang diambil yaitu data ulasan yang menggunakan Bahasa Indonesia.
4. Pengolahan dan analisis data menggunakan beberapa *tools* seperti Microsoft Excel dan Jupyter Notebook dengan bahasa pemrograman yang digunakan yaitu Python versi 3.10.5.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Induktif

Penelitian yang dilakukan oleh Afida, et al. (2021) terkait penerapan *Text Mining* untuk klusterisasi aduan masyarakat Kota Semarang dengan algoritma yang digunakan yaitu *K-Means Clustering*. Data yang digunakan yaitu data hasil proses dari *crawling data* aduan masyarakat dari twitter dengan kata kunci yang digunakan yaitu “laporhendi”. Hasil dari penelitian ini menunjukkan bahwa hasil klusterisasi data aduan masyarakat menggunakan metode *K-Means Clustering* terbagi menjadi 3 *cluster* dengan evaluasi hasil *cluster* yang digunakan yaitu menggunakan nilai *purity*. Model dari hasil *cluster* kemudian diimplementasikan kedalam *prototype* aplikasi analisis *text* aduan masyarakat Kota Semarang.

Irsyad & Pribadi (2020) melakukan penelitian tentang implementasi *Text Mining* dalam pengelompokan data tweet pertanian Indonesia dengan algoritma yang digunakan yaitu *K-Means Clustering*. Penelitian ini menggunakan 2 *tools* untuk pengolahan data yaitu Orange Data Mining untuk *text preprocessing* dan RapidMiner untuk melakukan pengolahan data menggunakan algoritma *K-Means Clustering*. Hasil dari penelitian ini yaitu dengan menerapkan algoritma *K-Means Clustering* terbentuk 5 *cluster*, yaitu Pangan, Produksi, Lahan, Ekspor dan Teknologi. Dari 5 *cluster* tersebut, dilakukan evaluasi menggunakan operator *%performance* pada RapidMiner dengan hasil rata-rata akurasi yang didapat yaitu sebesar 0,344%. Dari hasil penelitian juga terdapat 2 *cluster* memiliki nilai rata-rata akurasi tinggi yaitu *cluster* 0 (Pangan) dengan nilai sebesar 0,528% dan *cluster* 2 (Produksi) dengan nilai sebesar 0,523% dan untuk *cluster* dengan nilai terendah yaitu *cluster* 3 (Ekspor) dengan nilai sebesar 0,123%.

Sabna, et al. (2020) dalam penelitiannya meneliti tentang *Text Mining* untuk memprediksi keinginan pasar terkait perjalanan wisata menggunakan algoritma *K-Means Clustering*. Data yang digunakan pada penelitian ini yaitu sebanyak 250 data tweet terkait wisata. *Tools* yang digunakan pada penelitian ini yaitu RapidMiner Studio versi 9.7. Hasil

penelitian menunjukkan bahwa dengan kata kunci pencarian “wisata pantai” dan dilakukan proses klusterisasi menggunakan RapidMiner Studio menghasilkan kata “pantai” dan “kidul” menempati posisi paling tinggi. Sedangkan jika dilihat berdasarkan nama pantai, diperoleh hasil kata yaitu kata “lombok”. Dari hasil penelitian ini dapat disimpulkan bahwa wisata pantai yang populer yaitu Gunung Kidul di Yogyakarta dan pantai Lombok di Nusa Tenggara Barat.

Sentiya & Suroyo (2019) melakukan penelitian tentang *Text Clustering* komentar *followers* akun *Fanpage* Shopee Indonesia menggunakan *tools* Orange Data Mining. Pengambilan data menggunakan teknik *scraping* menggunakan ekstensi *Web Scraper*. Kemudian hasil pengambilan data dilakukan proses *preprocessing text* yang meliputi *transformation*, *tokenization*, *normalization*, dan *filtering*. Hasil dari penelitian ini menunjukkan bahwa data komentar *followers* akun twitter Shopee Indonesia terbentuk sebanyak 3 *cluster* dan *followers* akun twitter Shopee Indonesia paling banyak merespon status yang berhubungan dengan kata kuis berhadiah, *retweet* yang memberikan *handphone*, dan tebakan dalam bentuk *vote*.

Utami & Eka Putra (2022) melakukan penelitian tentang pengelompokan topik dokumen penelitian menggunakan algoritma *K-Means Clustering* dan *Cosine Similarity*. Data yang digunakan pada penelitian yaitu judul penelitian dosen tahun 2019 – 2021 berjumlah 52 data. Proses ekstraksi dokumen menggunakan pendekatan *Text Mining*, sedangkan untuk pengelompokan dokumen menggunakan algoritma *K-Means Clustering*. Hasil dari penelitian menunjukkan bahwa algoritma *K-Means Clustering* mampu menghasilkan *cluster* optimal sebanyak 6 *cluster*. Tren topik penelitian yang dilakukan dosen STMIK Primakara berdasarkan hasil klusterisasi yaitu Pengembangan dan Evaluasi Sistem Informasi, *E-Government*, *Data Mining*, Teknologi Pendidikan, *Machine Learning/Artificial Intelligence*, serta Manajemen dan Bisnis.

Chen, et al. (2022) melakukan penelitian tentang konten yang dibuat oleh pengguna dalam sistem Q&A dan komentar *online* berdasarkan pendekatan *Text Mining*. Data yang digunakan yaitu sebanyak 90.886 komentar dari sejumlah 1.822 produk laptop di situs Amazon. Algoritma yang digunakan pada penelitian ini yaitu *Improved K-Means Algorithm*. Hasil dari penelitian ini menunjukkan bahwa algoritma ini mampu menghasilkan *cluster* optimal sebanyak 9 *cluster*, dengan penentuan jumlah *cluster* terbaik menggunakan metode *silhouette coefficient*.

Huang, et al. (2021) melakukan studi empiris terkait rekomendasi iklan yang dipersonalisasi berdasarkan pengelompokan DBSCAN *User Generated Content* (UGC) Sina Weibo. Proses yang dilakukan pada penelitian ini yaitu mengelompokkan teks UGC, menambang data minat pengguna, mengekstraksi tema teks iklan dengan model LSI, dan menghitung kesamaan antara teks iklan dan minat pengguna. Teks iklan dilakukan proses filterisasi dan digunakan sebagai kumpulan rekomendasi, lalu diukur tingkat keakuratan rekomendasinya. Hasil penelitian ini menunjukkan bahwa DBSCAN efektif dalam menggambarkan topik yang diminati pengguna.

Jiachen, et al. (2019) melakukan penelitian tentang penerapan metode *Clustering* dan *Co-occurrence* untuk mengidentifikasi peristiwa penting dan hubungannya dengan Pasar Saham Cina. Penelitian ini melakukan pengelompokan teks untuk menemukan peristiwa dan merancang model jaringan korelasi peristiwa berita berdasarkan vektor kata. Data yang digunakan yaitu data berita tentang keuangan sepanjang tahun 2017 dan ditemukan kata kunci berita utama secara otomatis melalui pengelompokan teks. Algoritma yang digunakan untuk pengelompokan teks yaitu *K-Means Clustering* dengan jumlah *cluster* optimal yang terbentuk sebanyak 13 *cluster*. Hasil dari penelitian menunjukkan bahwa metode yang digunakan efektif dalam mengekstrak peristiwa *hotspot* berita.

Nasim & Haider (2022) melakukan penelitian dengan judul “*Cluster analysis of urdu tweets*”. Algoritma yang digunakan pada penelitian ini yaitu Algoritma *K-Means*, *Bisecting K-Means*, dan *Affinity Propagation Algorithm*. Kemudian dilakukan perbandingan dengan pendekatan pemodelan topik. Hasil dari penelitian menunjukkan bahwa algoritma *K-Means* memiliki performa lebih baik dalam pengelompokan data tweet Urdu dibandingkan dengan metode lainnya.

Zuliani, et al. (2021) melakukan penelitian dengan judul “*Topics and trends in Mountain Livestock Farming research: a text mining approach*”. Data yang digunakan untuk penelitian ini yaitu sebanyak 2679 dokumen, dimana 92% diantaranya merupakan dokumen publikasi ilmiah. Dengan pendekatan *text mining*, diperoleh informasi bahwa susu, kambing, dan domba merupakan kata dengan frekuensi yang sering muncul dalam korpus data. Dan diperoleh 10 topik bermakna berdasarkan analisis topik seperti pengelolaan ternak, dan dinamika vegetasi, kesehatan hewan dan epidermiologi, studi metodologi pada sapi, sistem produksi dan keberlanjutan, studi metodologis, penelitian satwa liar dan konservasi, reproduksi dan kinerja, produksi dan kualitas susu atau daging,

penggunaan lahan dan perubahannya dan topik terakhir yaitu genetika atau genomik. Analisis topik menggunakan pendekatan *Agglomerative Hierarchical Clustering* untuk pengelompokan keterkaitan antar topik. Hasil penelitian menunjukkan bahwa topik dikelompokkan ke dalam 3 *cluster*, yang pertama berfokus pada keberlanjutan, konservasi dan aspek sosial ekonomi, yang kedua berfokus pada produksi dan kualitas pangan, dan yang terakhir berfokus pada studi metodologi pada flora dan fauna pegunungan.

Probierz, et al. (2022) melakukan penelitian tentang klusterisasi artikel ilmiah menggunakan pendekatan *Natural Language Processing* (NLP). Data yang digunakan pada penelitian ini yaitu sebanyak 1500 artikel ilmiah yang diterbitkan pada tahun 2017 – 2022. Tujuan penelitian ini yaitu mengelompokkan artikel ilmiah sesuai dengan topik yang diidentifikasi berdasarkan isi abstrak dan pendahuluan. Algoritma yang digunakan yaitu *K-Means Clustering*. Hasil penelitian ini menunjukkan bahwa artikel dikelompokkan menjadi 3, 5, dan 7 kelompok dan membandingkan hasilnya berdasarkan akurasi NLP yang diterapkan. Kesimpulan pada penelitian ini yaitu pengelompokan artikel terbaik dapat diperoleh setelah menerapkan fitur bobot kata TF-IDF dengan *cluster* yang terbentuk sebanyak 7 *cluster*.

Pawar, et al. (2022) melakukan penelitian dengan judul “*Text Summarization Using Document and Sentence Clustering Method*”. Tujuan dari dilakukannya penelitian ini yaitu untuk meringkas dokumen untuk pengambilan informasi lebih cepat. Data yang digunakan yaitu *dataset* yang didapat dari *website* Kaggle. Algoritma yang digunakan pada penelitian ini yaitu menggunakan metode *Jaccard Similarity* untuk klusterisasi dokumen. Hasil dari penelitian ini menunjukkan bahwa metode ini menghasilkan kinerja yang baik dalam pengelompokan dokumen dan mengurangi redundansi terkait pengelompokan dokumen.

Škrjanc, et al. (2022) melakukan penelitian dengan judul “*Evolving Gaussian on-line clustering in social network analysis*”. Tujuan dari penelitian ini yaitu mengelompokkan data pengguna Twitter secara otomatis sesuai dengan perilaku pengguna. Data yang digunakan yaitu data dari Twitter API dengan jumlah sebanyak 2866 data. Analisis data menggunakan Algoritma eGauss+. Hasil penelitian menunjukkan bahwa metode eGauss+ cocok untuk melakukan klasifikasi dan klusterisasi. Untuk pengelompokan data pengguna Twitter ditemukan 4 *cluster* pengguna yaitu *ordinary*, *very active*, *influential*, dan *unusual*.

Supianto, et al. (2023) melakukan penelitian dengan judul “*Cluster-based text mining for extracting drug candidates for the prevention of COVID-19 from the biomedical*

literature". Data yang digunakan pada penelitian ini diambil dari situs PubMed, PubChem, dan DrugBank. Analisis data menggunakan teknik *Agglomerative Hierarchical Clustering*. Hasil dari penelitian ini menunjukkan bahwa penggunaan metode klasterisasi berbasis *Text Mining* terbukti memberikan dampak positif terhadap penemuan kandidat obat COVID-19. 24 dari 26 delapan senyawa dikonfirmasi sebagai kandidat obat COVID-19 menurut DrugBank.

Reddy, et al. (2021) melakukan penelitian dengan judul "*Use and validation of text mining and cluster algorithms to derive insights from Corona Virus Disease-2019 (COVID-19) medical literature*". Data yang digunakan pada penelitian ini yaitu diambil dari *The COVID-19 Open Research Dataset (CORD-19)*. Analisis yang digunakan pada penelitian ini menggunakan metode *Spectral Clustering* dan *Agglomerative Clustering*. Hasil dari penelitian ini menunjukkan bahwa hasil pengelompokan dataset memiliki dampak positif karena dapat dengan cepat mengakses informasi yang akurat dan tepat dari infodemic COVID-19 dalam bentuk visual.

Kajian induktif memuat tentang kajian penelitian terdahulu yang berkaitan dengan topik penelitian yang diangkat. Berdasarkan kajian induktif yang telah dipaparkan diatas, berikut merupakan tabel rangkuman mengenai kajian induktif penelitian terdahulu.

Tabel 2.1 Rangkuman Kajian Induktif

Peneliti	Metode	Objek	Hasil
Afida, et al. (2021)	<i>K-Means Clustering</i> dan nilai <i>purity</i>	Data aduan masyarakat Kota Semarang.	Penelitian ini menghasilkan klasterisasi data aduan masyarakat sebanyak 3 <i>cluster</i> dengan evaluasi hasil <i>cluster</i> menggunakan nilai <i>purity</i> . Model dari hasil <i>cluster</i> kemudian diimplementasikan kedalam <i>prototype</i> aplikasi analisis <i>text</i> aduan masyarakat Kota Semarang.
Irsyad & Pribadi (2020)	<i>K-Means Clustering</i>	Data tweet pertanian Indonesia.	Penelitian ini menghasilkan 5 <i>cluster</i> dari data data tweet pertanian Indonesia, yaitu Pangan, Produksi, Lahan, Ekspor dan Teknologi. Dari 5 <i>cluster</i> tersebut, dilakukan evaluasi <i>cluster</i> menggunakan

			operator <i>%performance</i> dengan hasil rata-rata akurasi yang didapat sebesar 0,344%. Terdapat 2 <i>cluster</i> memiliki nilai rata-rata akurasi tinggi yaitu <i>cluster</i> 0 (Pangan) sebesar 0,528% dan <i>cluster</i> 2 (Produksi) sebesar 0,523% dan <i>cluster</i> dengan nilai terendah yaitu <i>cluster</i> 3 (Ekspor) sebesar 0,123%.
Sabna, et al. (2020)	<i>K-Means Clustering</i>	Data tweet terkait wisata.	Proses penelitian ini menggunakan kata pencarian “wisata pantai” dan dilakukan proses klasterisasi menggunakan RapidMiner Studio. kemudian menghasilkan kata “pantai” dan “kidul” yang menempati posisi paling tinggi. Sedangkan jika dilihat dari nama pantai, diperoleh hasil kata yaitu “Lombok”. Dapat disimpulkan bahwa wisata pantai yang populer yaitu di Gunung Kidul Yogyakarta dan pantai Lombok di Nusa Tenggara Barat.
Sentiya & Suroyo (2019)	<i>K-Means Clustering</i>	Data komentar <i>followers</i> akun <i>Fanpage</i> Shopee Indonesia.	Penelitian ini melakukan analisis <i>text clustering</i> komentar <i>followers</i> akun <i>Fanpage</i> Shopee Indonesia. Hasil dari penelitian ini menunjukkan bahwa data komentar <i>followers</i> akun twitter Shopee Indonesia terbentuk sebanyak 3 <i>cluster</i> dan <i>followers</i> akun twitter Shopee Indonesia paling banyak merespon status yang berhubungan dengan kata kuis berhadiah, <i>retweet</i> yang memberikan <i>handphone</i> , dan tebakkan dalam bentuk <i>vote</i> .

Utami & Eka Putra (2022)	<i>K-Means Clustering dan Cosine Similarity</i>	Data penelitian dosen tahun 2019 – 2021.	judul Penelitian ini melakukan pengelompokan topik dokumen penelitian menggunakan <i>K-Means Clustering</i> dan <i>Cosine Similarity</i> . Hasil dari penelitian ini menunjukkan bahwa dengan algoritma <i>K-Means</i> menghasilkan pengelompok dokumen sebanyak 6 <i>cluster</i> . Tren topik penelitian yang dilakukan dosen STMIK Primakara berdasarkan hasil klasterisasi yaitu Pengembangan dan Evaluasi Sistem Informasi, <i>E-Government</i> , <i>Data Mining</i> , Teknologi Pendidikan, <i>Machine Learning/Artificial Intelligence</i> , serta Manajemen dan Bisnis.
Chen, et al. (2022)	<i>Improved K-Means Algorithm dan Silhoutte Coefficient</i>	Data komentar produk laptop di situs Amazon.	Penelitian ini melakukan pengelompokan konten yang dibuat oleh pengguna dalam sistem Q&A. Hasil dari penelitian ini menunjukkan bahwa algoritma ini mampu menghasilkan <i>cluster</i> optimal sebanyak 9 <i>cluster</i> , dengan penentuan jumlah <i>cluster</i> terbaik menggunakan metode <i>silhouette coefficient</i> .
Huang, et al. (2021)	DBSCAN	Data <i>User Generated Content</i> (UGC) di situs Sina Weibo.	Penelitian ini melakukan studi empiris terkait rekomendasi iklan yang dipersonalisasi menggunakan metode pengelompokan DBSCAN <i>User Generated Content</i> (UGC) Sina Weibo. Hasil dari penelitian ini menunjukkan bahwa DBSCAN efektif dalam menggambarkan topik yang diminati pengguna.

Jiachen, et al. (2019)	<i>K-Means Clustering dan Co-occurence</i>	Data berita tentang keuangan sepanjang tahun 2017.	Penelitian ini melakukan identifikasi peristiwa penting dan analisis hubungannya dengan Pasar Saham Cina menggunakan pendekatan <i>Clustering</i> dan <i>Co-occurence</i> . Hasil dari penelitian ini menunjukkan bahwa metode DBSCAN efektif dalam mengekstrak peristiwa <i>hotspot</i> berita dengan <i>cluster</i> yang terbentuk sebanyak 13 <i>cluster</i> .
Nasim & Haider (2022)	<i>K-Means Clustering, Bisecting K-Means dan Affinity Propagation Algorithm</i>	Data tweet bahasa Urdu.	Penelitian ini melakukan komparasi algoritma <i>clustering</i> dalam melakukan pembentukan <i>cluster</i> terhadap tweet Urdu. Hasil penelitian ini menunjukkan bahwa algoritma <i>K-Means Clustering</i> memiliki performa yang lebih baik dalam pengelompokan tweet Urdu dibandingkan dengan metode lainnya.
Zuliani, et al. (2021)	<i>Agglomerative Hierarchical Clustering</i>	Data dokumen publikasi ilmiah.	Dengan pendekatan <i>text mining</i> , diperoleh informasi bahwa susu, kambing, dan domba merupakan kata dengan frekuensi yang sering muncul dalam korpus data. Dan diperoleh 10 topik bermakna berdasarkan analisis topik seperti pengelolaan ternak, dan dinamika vegetasi, kesehatan hewan dan epidermiologi, studi metodologi pada sapi, sistem produksi dan keberlanjutan, studi metodologis, penelitian satwa liar dan konservasi, reproduksi dan kinerja, produksi dan kualitas susu atau daging, penggunaan lahan dan perubahannya dan topik terakhir yaitu genetik atau genomik. Kemudian

			<p>topik dikelompokkan menggunakan algoritma <i>Agglomerative Hierarchical Clustering</i>. Hasil penelitian menunjukkan bahwa topik dikelompokkan kedalam tiga <i>cluster</i>, yang pertama berfokus pada keberlanjutan, konservasi dan aspek sosial ekonomi, yang kedua berfokus pada produksi dan kualitas pangan, dan yang terakhir berfokus pada studi metodologi pada flora dan fauna pegunungan</p>
<p>Probierz, et al. (2022)</p>	<p><i>K-Means Clustering</i></p>	<p>Data ilmiah diterbitkan pada tahun 2017 – 2022.</p>	<p>Penelitian ini melakukan klasterisasi artikel ilmiah menggunakan pendekatan NLP. Hasil penelitian ini menunjukkan bahwa artikel dikelompokkan menjadi 3, 5, dan 7 kelompok dan membandingkan hasilnya berdasarkan akurasi NLP yang diterapkan. Kesimpulan pada penelitian ini yaitu pengelompokan artikel terbaik dapat diperoleh setelah menerapkan fitur bobot kata TF-IDF dengan <i>cluster</i> yang terbentuk sebanyak 7 <i>cluster</i>.</p>
<p>Pawar, et al. (2022)</p>	<p><i>Jaccard Similarity</i></p>	<p><i>Dataset</i> pada <i>website</i> Kaggle.</p>	<p>Tujuan dari penelitian ini yaitu meringkas dokumen untuk pengambilan informasi lebih cepat. Hasil dari penelitian ini menunjukkan bahwa metode ini menghasilkan kinerja yang baik dalam pengelompokan dokumen dan mengurangi redundansi terkait pengelompokan dokumen.</p>

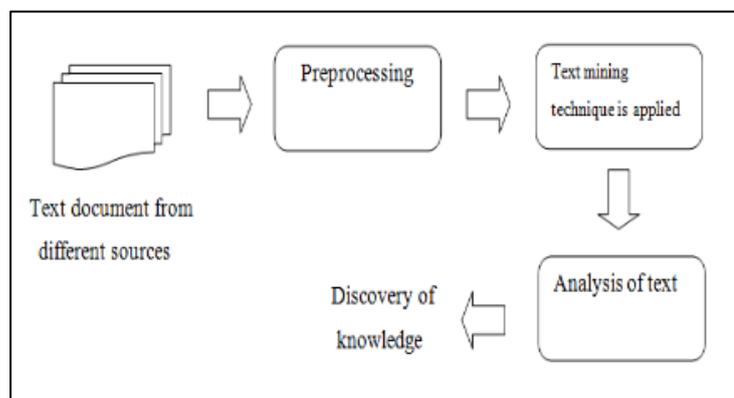
Škrjanc, et al. (2022)	<i>eGauss+</i> <i>Algorithm</i>	Data pengguna twitter yang di dapat pada Twitter API sebanyak 2866 data.	Tujuan dari penelitian ini yaitu mengelompokkan data pengguna Twitter secara otomatis sesuai dengan perilaku pengguna. Hasil penelitian menunjukkan bahwa metode eGauss+ cocok untuk melakukan klasifikasi dan klasterisasi.
Supianto, et al. (2023)	<i>Agglomerative Hierarchical Clustering</i>	Data pada situs PubMed, PubChem, dan DrugBank.	Penelitian melakukan pengelompokan obat untuk mendapatkan kandidat obat untuk pencegahan COVID-19. Hasil dari penelitian ini menunjukkan bahwa penggunaan metode klasterisasi berbasis <i>Text Mining</i> terbukti memberikan dampak positif terhadap penemuan kandidat obat COVID-19. 24 dari 26 delapan senyawa dikonfirmasi sebagai kandidat obat COVID-19 menurut DrugBank.
Reddy, et al. (2021)	<i>Agglomerative Hierarchical Clustering dan Spectral Clustering</i>	Data pada situs <i>COVID-19 Open Research Dataset</i> (CORD-19).	Penelitian ini melakukan penggunaan <i>text mining</i> dan validasi menggunakan algoritma <i>clustering</i> untuk memperoleh informasi dari literatur medis COVID-19. Hasil dari penelitian ini menunjukkan bahwa hasil pengelompokan dataset memiliki dampak positif karena dapat dengan cepat mengakses informasi yang akurat dan tepat dari infodemik COVID-19 dalam bentuk visual.

Berdasarkan rangkuman kajian induktif diatas, dapat disimpulkan bahwa beragam metode yang digunakan dalam melakukan pengelompokan data teks yang dilakukan oleh peneliti terdahulu. Hasil penelitian yang dilakukan oleh peneliti terdahulu menunjukkan pendekatan *clustering* cukup populer digunakan untuk pengelompokan data yang bersifat tekstual. Maka dari itu, dilakukan penelitian terkait klasterisasi data ulasan pelanggan Gudeg Yu Djum untuk perbaikan kualitas layanan restoran. Harapannya dengan dilakukannya penelitian ini dapat memberikan informasi mengenai keluhan pelanggan melalui pengelompokan data ulasan dan diberikan usulan perbaikan menggunakan metode diagram *fishbone*. Usulan perbaikan tersebut dapat dijadikan sebagai bahan pertimbangan bagi restoran dalam meningkatkan kualitas layanan restoran.

2.2 Kajian Deduktif

2.2.1 Text Mining

Text mining merupakan variasi dari bidang ilmu yang disebut dengan *data mining* yang mencoba untuk mencari pola yang menarik dari suatu *database* besar (Sagayam et al., 2012). *Text mining* melakukan proses pencarian atau penggalian informasi yang berguna melalui data tekstual (Gurusamy & Kannan S, 2014). Dalam *text mining*, proses penambangan teks dimulai dengan pengumpulan dokumen dari berbagai sumber. *Tools* dari *text mining* akan melakukan pengambilan dokumen tertentu dan melakukan pra proses data dengan memeriksa format dan rangkaian karakter pada dokumen (Gaikwad et al., 2014). Berikut merupakan proses dalam melakukan *text mining*.



Gambar 2.1 Proses *Text Mining*

Sumber: Gaikwad, et al. (2014)

Terdapat komponen utama sebelum melakukan analisis dokumen, yaitu *text preprocessing*. *Text preprocessing* merupakan proses mengubah data tekstual yang tidak terstruktur menjadi bentuk data yang lebih terstruktur dan kemudian disimpan ke dalam basis data (Langgeni et al., 2010). Dalam melakukan *text preprocessing*, terdapat beberapa tahapan yang harus dilakukan, yaitu sebagai berikut.

1. *Case folding*

Case folding merupakan langkah pertama dari seluruh rangkaian *preprocessing* dokumen. Proses yang dilakukan pada tahap *case folding* yaitu mengubah semua huruf yang terdapat pada dokumen menjadi huruf kecil (Naf'an et al., 2019).

2. *Tokenizing*

Tokenizing merupakan proses pemotongan kalimat pada dokumen menjadi kumpulan kata (Amrizal, 2018).

3. *Filtering*

Filtering merupakan proses melakukan pemilihan kata-kata yang memiliki makna yang didapat dari hasil tokenisasi, yaitu kata-kata yang bisa digunakan untuk mewakili isi dari sebuah dokumen. Proses ini disebut juga dengan *stopword removal* (Sholehhudin et al., 2018).

2.2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Pembobotan kata merupakan langkah awal pra-pemrosesan teks dalam penerapan pengelompokan teks dimana setiap istilah diberi bobot yang sesuai dengan tujuan meningkatkan performa dari pengelompokan teks tersebut (Omar, 2020). Salah satu teknik yang sering digunakan dalam melakukan pembobotan kata yaitu teknik pembobotan TF-IDF.

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik pembobotan kata dengan cara menentukan frekuensi relatif kata dalam suatu dokumen tertentu melalui proporsi terbalik dari kata-kata di seluruh korpus dokumen. Dalam menentukan nilai bobotnya, metode ini menggunakan dua elemen, yaitu : TF – frekuensi

term dari *term* i dalam dokumen j dan IDF – *inverse document frequency* dari *term* i (Trstenjak et al., 2014). Berikut ini merupakan rumus dari pembobotan kata TF-IDF.

$$a_{ij} = tf_{ij}idf_i = tf_{ij} \times \log_2 \left(\frac{N}{df_i} \right) \quad (2.1)$$

Dimana:

a_{ij} = bobot term i pada dokumen j .

N = jumlah dokumen dalam korpus.

tf_{ij} = frekuensi istilah dari istilah i dalam dokumen j .

df_i = frekuensi dokumen istilah i dalam korpus.

2.2.3 Clustering

Menurut Sinaga & Yang (2020), *Clustering* merupakan metode dalam menentukan struktur *cluster* dalam suatu kumpulan data yang ditandai dengan kesamaan terbesar dalam *cluster* yang sama dan perbedaan terbesar dalam *cluster* yang berbeda. *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa pengawasan (*unsupervised*) (Agusta, 2007). Algoritma *Clustering* digunakan untuk mengatur data, mengkategorikan data, untuk kompresi dan konstruksi model, untuk mendeteksi *outlier*, dan lain sebagainya (Guruvayur & Suchithra, 2018).

Analisis *cluster* banyak digunakan dalam aplikasi seperti intelijen bisnis, pengenalan pola gambar, pencarian *web*, biologi, dan keamanan. *Clustering* juga dikenal dengan segmentasi data di beberapa aplikasi karena metode ini mempartisi kumpulan data besar ke dalam kelompok sesuai dengan kesamaannya. Contoh penerapan *clustering* ini yaitu untuk deteksi *outlier*, pendeteksi *outlier* ini termasuk untuk kasus-kasus umum seperti pendeteksi penipuan kartu kredit dan pemantauan aktivitas kriminal dalam *e-commerce* (Han et al., 2012).

Terdapat 2 jenis *Clustering* dalam melakukan pengelompokan data, yaitu secara hierarki dan partisi. Algoritma hirearki dalam menemukan *cluster* yang berurutan menggunakan *cluster* yang telah didirikan sebelumnya, sedangkan algoritma partisi menentukan *cluster* pada waktu (Madhulatha, 2007). Contoh dari algoritma hirearki yaitu

Agglomerative Clustering dan *Divisive Clustering*, sedangkan untuk partisi seperti algoritma *K-Means Clustering*.

2.2.4 *Machine Learning*

Machine learning didefinisikan sebagai cabang dari *Artificial Intelligence* yang berfokus pada pembelajaran dari data dan pengembangan sistem yang mampu belajar secara mandiri tanpa harus berulang kali di *program* ulang oleh manusia (Cholissodin & Soebroto, 2021). Cara kerja dari *machine learning* bergantung pada algoritma yang digunakan dalam memecahkan masalah data (Horvitz & Mulligan, 2015).

Strategi pembelajaran yang digunakan pada *machine learning* ini yaitu menemukan pola dalam kumpulan *training data*. Pola ini kemudian digunakan untuk mengklasifikasikan atau memprediksi perilaku baru (Bose & Mahapatra, 2001). Menurut Somvanshi (2016), terdapat 3 jenis pembelajaran dalam *machine learning*, yaitu *supervised learning*, *unsupervised learning* dan *reinforcement learning*. Ketiga jenis pembelajaran ini dapat dijabarkan sebagai berikut.

1. *Supervised learning*

Metode *supervised learning* dilengkapi dengan *input* dan *output* yang sudah diketahui sebelumnya. Skenario pembelajaran dari *supervised learning* ini akan memungkinkan dapat menentukan label kelas dengan benar. Contoh algoritma dari *supervised learning* yaitu *Support Vector Machine*, *Decision Tree*, dan lain-lain.

2. *Unsupervised learning*

Metode *unsupervised learning* dilengkapi dengan beberapa *input* tetapi sistem tidak mengetahui *output* yang benar. Tujuan dari algoritma ini yaitu mendefinisikan fungsi untuk menggambarkan struktur tersembunyi dari data yang tidak berlabel.

3. *Reinforcement learning*

Metode *reinforcement learning* merupakan subdomain dari *machine learning* yang berkaitan dengan bagaimana agen perangkat lunak dalam mengambil tindakan di lingkungan interaktif.

2.2.5 *K-Means Clustering*

Menurut Husni et al. (2015) *K-Means Clustering* didefinisikan sebagai metode pengelompokan data non hierarki yang berusaha mempartisi data ke dalam bentuk dua atau lebih kelompok. Tujuan dari *K-Means Clustering* yaitu meminimalkan jarak rata-rata dokumen dari pusat *cluster* nya, dimana pusat *cluster* didefinisikan sebagai rata-rata pusat massa dokumen dalam sebuah *cluster* (Singh et al., 2011).

Algoritma *K-Means Clustering* menggunakan proses yang bersifat iteratif dalam pengelompokan data. Dibutuhkan inialisasi jumlah *cluster* yang diinginkan untuk melakukan proses *clustering* (Bangoria et al., 2013). Menurut Wakhidah (2014), berikut merupakan proses dalam melakukan *clustering* menggunakan algoritma *K-Means Clustering*.

1. Menentukan jumlah *cluster*.
2. Menentukan nilai centroid.

Dalam menentukan nilai centroid untuk iterasi pertama, penentuan nilai centroid dilakukan secara acak. Sedangkan jika nilai centroid merupakan tahap dari iterasi, dapat digunakan rumus sebagai berikut.

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2.2)$$

Dimana:

- \bar{v}_{ij} = centroid atau rata-rata *cluster* ke-i untuk variabel ke-j.
- N_i = jumlah data yang menjadi anggota *cluster* ke-i.
- i, k = indeks *cluster*.
- j = koordinat centroid.
- x_{kj} = nilai data ke-k yang ada didalam *cluster* tersebut untuk variabel ke-j.

3. Menghitung jarak antara centroid dengan titik tiap objek

Untuk menghitung jarak centroid, dapat menggunakan rumus *Euclidean distance*, yaitu:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.3)$$

Dimana:

D_e = *Euclidean distance*

i = banyak nya objek

x, y = koordinat objek

s, t = koordinat centroid

4. Pengelompokan data

Dalam menentukan anggota *cluster*, dilakukan perhitungan jarak minimum objek. Nilai keanggotaan data pada *distance* matriks yaitu 0 atau 1, yang dimana nilai 1 untuk data yang ditempatkan ke *cluster* dengan karakteristik yang sama dan nilai 0 untuk data yang ditempatkan ke *cluster* yang lain.

5. Kembali ke langkah 2, lakukan perhitungan ulang hingga mendapatkan nilai centroid tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

2.2.6 Metode *Elbow*

Metode *Elbow* merupakan metode dalam penentuan jumlah *cluster* terbaik dengan cara menggunakan kuadrat jarak antara titik sampel di setiap *cluster* dan pusat *cluster* (Yuan & Yang, 2019). Metode *Elbow* dapat menjelaskan konsistensi dan memverifikasi analisis *clustering*, dan bertujuan untuk membantu dalam menemukan jumlah *cluster* yang sesuai dalam kumpulan data (Liu & Deng, 2021). Menurut Cui (2020), berikut merupakan rumus dari metode *Elbow* dengan metrik *Within Cluster Sum of Squares* (WCSS).

$$WCSS = \sum_{P_i \text{ in Cluster 1}} distance(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} distance(P_i, C_2)^2 + \dots \quad (2.4)$$

2.2.7 *Association Rule*

Menurut Kaur & Madan (2015) *Association Rule* merupakan teknik dalam *data mining* yang bertujuan untuk mengekstraksi korelasi yang menarik, pola kemunculan, dan asosiasi atau struktur kasual diantara kumpulan *item* dalam *database* transaksi atau penyimpanan data

lainnya. Aturan asosiasi digunakan untuk menemukan hubungan antara objek yang sering digunakan bersama. Contoh aplikasi dari aturan asosiasi yaitu *market basket analysis*, *clustering*, klasifikasi, dan lain-lain (Kumbhare & V Chobe, 2014). Rajak & Gupta (2008) mengemukakan bahwa terdapat tiga ukuran dasar dalam melakukan aturan asosiasi, yaitu *support*, *confidence*, dan *lift*.

1. *Support*

Merupakan jumlah transaksi yang mencakup semua *item* di bagian anteseden dan konsekuen aturan (nilai *support* dijabarkan dalam bentuk persentase dari jumlah *total record* dalam *database*).

2. *Confidence*

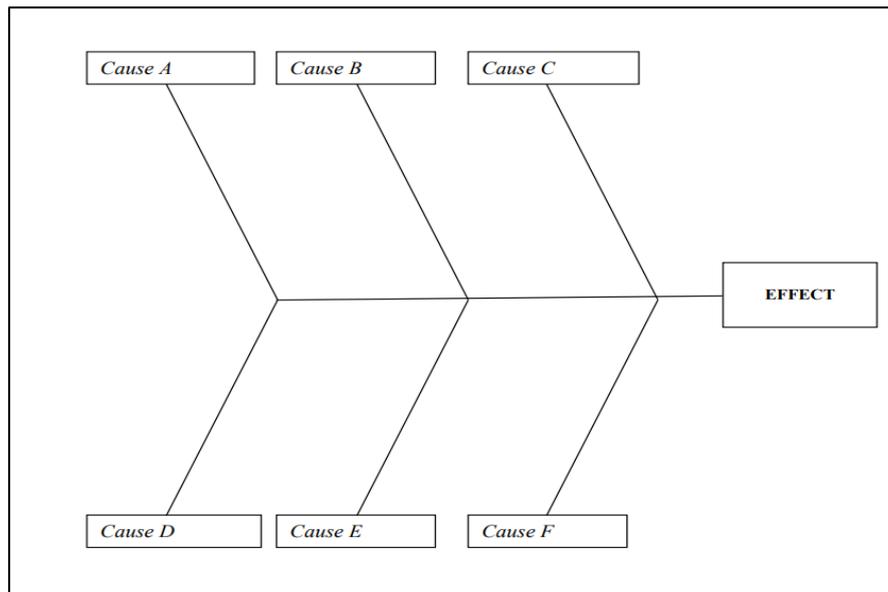
Merupakan rasio jumlah transaksi yang mencakup semua *item* dalam konsekuen serta anteseden (yaitu *support*) ke jumlah transaksi yang mencakup semua *item* dalam anteseden.

3. *Lift*

Merupakan nilai yang memberikan informasi tentang probabilitas dari bagian ‘maka’ (konsekuen) yang diberikan bagian ‘jika’ (anteseden).

2.2.8 *Fishbone Diagram*

Fishbone Diagram (atau lebih dikenal dengan *Ishikawa Diagram*) merupakan *tools* yang digunakan untuk menganalisis akar permasalahan dan dicetus pertama kali oleh seorang warga negara Jepang yang bernama Kaoru Ishikawa (1915-1989). Menurut Gaspers (2002) dalam Hamidy (2016), *Fishbone Diagram* didefinisikan sebagai pendekatan terstruktur dengan dilakukannya analisis yang lebih terperinci dalam menemukan penyebab-penyebab suatu masalah, ketidaksesuaian, dan kesenjangan yang ada. *Fishbone diagram* merupakan suatu alat visual untuk mengidentifikasi dan menggambarkan secara detail semua penyebab yang berhubungan dengan suatu permasalahan (Kuswardana, 2017). Berikut merupakan bentuk dari *fishbone diagram*.



Gambar 2.2 *Fishbone Diagram*

Sumber: Coccia (2018)

BAB III

METODE PENELITIAN

3.1 Objek dan Subjek Penelitian

Objek pada penelitian ini yaitu ulasan yang diberikan oleh pelanggan restoran Gudeg Yu Djum di *website* TripAdvisor. Sedangkan subjek pada penelitian ini yaitu pelanggan restoran Gudeg Yu Djum. Restoran Gudeg Yu Djum dipilih menjadi objek pada penelitian ini karena restoran Gudeg Yu Djum merupakan restoran yang sering dikunjungi wisatawan dan salah satu restoran terbaik di Yogyakarta yang menyajikan makanan khas daerah berupa makanan gudeg. Namun, masih banyak pelanggan yang memberikan keluhan mengenai restoran tersebut terkait pelayanan yang diberikan. Oleh karena itu, restoran perlu melakukan evaluasi agar kualitas layanan restoran meningkat dan mampu memenuhi kebutuhan pelanggan.

3.2 Populasi dan Sampel Penelitian

Populasi pada penelitian ini yaitu data historis ulasan pelanggan restoran Gudeg Yu Djum yang diambil dari situs TripAdvisor dengan jumlah ulasan sebanyak 586 ulasan. Sampel data yang digunakan merupakan data ulasan pada bulan Juli 2012 – November 2022. Pemilihan sampel dengan rentang waktu tersebut karena awal mula pelanggan memberikan ulasan mengenai restoran tersebut. Faktor lainnya yaitu jika mengambil hanya pada tahun tertentu, data yang didapat sangat sedikit dan hasil analisis data yang didapat kurang baik. Sehingga pada penelitian ini berfokus pada analisis seluruh data ulasan restoran yang terdapat pada situs TripAdvisor dengan ulasan yang diteliti yaitu ulasan Bahasa Indonesia.

3.3 Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini menggunakan teknik *web scraping*. Alat yang digunakan untuk *scraping* data ulasan restoran Gudeg Yu Djum yaitu menggunakan aplikasi WebHarvy.

3.4 Jenis dan Sumber Data

Data yang digunakan pada penelitian ini yaitu data sekunder. Data sekunder ini diambil dari situs TripAdvisor dengan bantuan aplikasi *scraping* data ulasan yaitu WebHarvy dan teknik pengambilan data yang digunakan dengan pendekatan *web scraping*. Sumber data yang digunakan pada penelitian ini berasal dari situs TripAdvisor dengan alamat *domain* yaitu https://www.tripadvisor.com/Restaurant_Review-g294230-d3156129-Reviews-Gudeg_Yu_Djum-Yogyakarta_Region_Java.html. Data yang diperoleh berupa ulasan pelanggan restoran Gudeg Yu Djum yang diambil pada bulan Juli 2012 – November 2022 sebanyak 586 ulasan.

3.5 Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini yaitu menggunakan Jupyter Notebook dengan bahasa pemrograman Python dan Microsoft Excel 2019. Terdapat beberapa metode analisis data yang digunakan pada penelitian ini, yang dapat dijabarkan sebagai berikut.

1. Analisis Deskriptif

Metode analisis deskriptif digunakan untuk memberikan gambaran ulasan pelanggan mengenai restoran Gudeg Yu Djum yang ada pada situs TripAdvisor.

2. Metode *Machine Learning*

Algoritma *machine learning* yang digunakan pada penelitian ini yaitu algoritma *K-Means Clustering*. Fungsi dari algoritma ini adalah mengelompokkan data ulasan ke beberapa *cluster* dengan tingkat kemiripan maksimum dan data antar *cluster* memiliki tingkat kemiripan yang minimum.

3. *WordCloud*

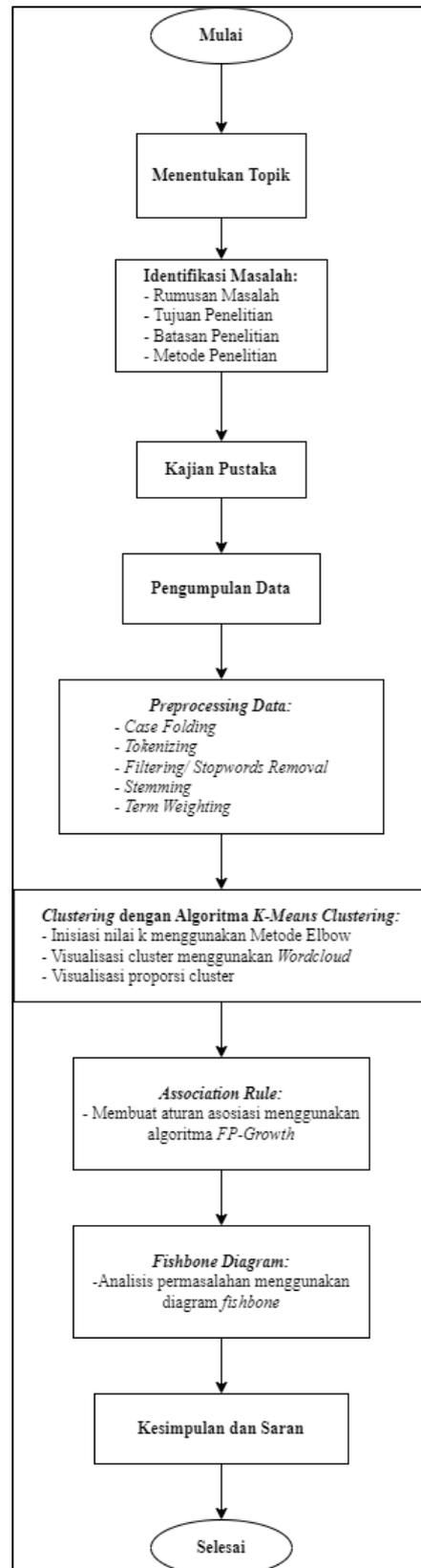
Metode *wordcloud* digunakan untuk proses visualisasi yang memuat informasi berupa kata-kata yang sering muncul pada masing-masing *cluster* ulasan restoran Gudeg Yu Djum.

4. Diagram *Fishbone*

Metode diagram *fishbone* digunakan untuk melakukan analisa permasalahan berdasarkan *term* yang bersifat negatif.

3.6 Alur Penelitian

Langkah-langkah yang dilakukan pada penelitian ini dijelaskan dalam bentuk gambar *flowchart*. Alur yang terdapat pada penelitian ini yaitu sebagai berikut.



Gambar 3.1 Alur Penelitian

Berikut merupakan penjelasan dari alur penelitian diatas.

1. Mulai

Tahap pertama yaitu peneliti mulai melakukan penelitian dengan melakukan pengamatan terhadap ulasan pelanggan restoran Gudeg Yu Djum melalui situs TripAdvisor.

2. Menentukan Topik

Tahap kedua yang dilakukan pada penelitian ini yaitu menentukan topik penelitian. Berdasarkan pengamatan, terdapat ulasan yang bersifat negatif yang diberikan oleh pelanggan mengenai kualitas layanan restoran Gudeg Yu Djum. Maka dari itu, topik yang diangkat pada penelitian ini yaitu implementasi *text mining* untuk perbaikan kualitas layanan restoran Gudeg Yu Djum berdasarkan ulasan yang diberikan oleh pelanggan.

3. Identifikasi dan Perumusan Masalah

Tahap ketiga yaitu melakukan identifikasi dan perumusan masalah. Berdasarkan pengamatan melalui situs TripAdvisor, terdapat beragam ulasan yang diberikan oleh pelanggan, mulai dari ulasan yang bersifat positif maupun negatif. Terdapat beberapa ulasan negatif yang diberikan oleh pelanggan, yaitu mayoritas pelanggan mengeluhkan dari segi rasa gudeg terlalu manis, pelayanan kurang ramah, dan harga gudeg cukup mahal dari gudeg lainnya. Berdasarkan permasalahan tersebut, dilakukan perumusan masalah pada penelitian ini, yaitu bagaimana hasil pengelompokan data ulasan restoran Gudeg Yu Djum dengan pendekatan *K-Means Clustering*, faktor apa saja yang perlu diperbaiki dalam meningkatkan kualitas layanan restoran Gudeg Yu Djum berdasarkan hasil klasterisasi data, dan bagaimana usulan perbaikan yang dapat diberikan guna meningkatkan kualitas layanan restoran tersebut. Kemudian pada tahap ini juga menentukan tujuan dari penelitian, batasan, dan metode yang digunakan pada penelitian.

4. Kajian Pustaka

Tahap keempat yaitu melakukan kajian pustaka dengan tujuan untuk memperkuat penelitian yang akan dilakukan. Proses yang dilakukan pada kajian pustaka ini yaitu melakukan kajian induktif seperti merangkum penelitian-penelitian terdahulu yang selaras dengan topik yang diangkat dan kajian deduktif berupa teori yang digunakan sebagai pendukung pada penelitian yang dilakukan.

5. Pengumpulan Data

Tahap kelima yaitu melakukan pengumpulan data ulasan pelanggan restoran Gudeg Yu Djum dengan pendekatan *web scraping*. Alat yang digunakan untuk membantu dalam pengambilan data ulasan pelanggan yaitu aplikasi WebHarvy. Data yang digunakan pada penelitian ini yaitu data ulasan pelanggan restoran Gudeg Yu Djum di situs TripAdvisor.

6. *Preprocessing Data*

Tahap keenam yaitu melakukan *preprocessing data*. Terdapat beberapa aktivitas dalam melakukan *preprocessing data*. Aktivitas *preprocessing data* yang dilakukan dapat dijabarkan sebagai berikut.

- a. *Case Folding*: merupakan tahap menyeragamkan ulasan menjadi huruf kecil, menghapus angka, dan menghapus tanda baca.
- b. *Tokenizing*: merupakan tahap pemotongan kalimat ulasan menjadi kata per kata.
- c. *Filtering*: Setelah melakukan *tokenizing*, tahap selanjutnya yaitu melakukan *filtering*. Proses yang dilakukan yaitu pemilihan kata-kata yang mengandung makna yang didapat dari hasil tokenisasi.
- d. *Stemming*: Proses ini merupakan tahap selanjutnya dari proses *filtering*. *Stemming* merupakan tahap dimana kata yang didapat dari hasil *filtering* diubah ke dalam bentuk baku.
- e. *Term Weighting*: Setelah menyelesaikan seluruh rangkaian diatas, maka tahap selanjutnya yaitu *term weighting*. Fungsi dari *term weighting* yaitu memberi bobot terhadap dokumen yang sudah di proses sebelumnya. *Term Weighting* ini perlu dilakukan agar mesin pembelajaran (*machine learning*) dapat membaca dokumen, dan kemudian dapat dilakukan ke tahap analisis selanjutnya.

7. *Clustering* dengan Algoritma *K-Means Clustering*

Data ulasan yang sudah melalui tahap *pre-processing data*, kemudian dilakukan proses pengelompokan data ulasan. Pada proses pengelompokan data ulasan ini, metode *machine learning* yang digunakan yaitu algoritma *K-Means Clustering*.

8. *Association Rule*

Setelah melakukan pengelompokan dan visualisasi data ulasan, selanjutnya membuat aturan asosiasi terhadap frekuensi kata terbesar yang terdapat pada masing-masing *cluster*. Aturan asosiasi dilakukan dengan tujuan agar informasi

yang didapat lebih akurat. Luaran dari proses ini yaitu asosiasi kata serta nilai *support*, *confidence*, dan *lift ratio*.

9. *Fishbone Diagram*

Setelah membuat aturan asosiasi pada data ulasan, tahap selanjutnya yaitu menganalisis permasalahan menggunakan *fishbone diagram*. Analisis *fishbone diagram* digunakan untuk mencari akar dari permasalahan yang ditemukan. Setelah mendapatkan akar dari permasalahan yang ditemukan, kemudian diberikan usulan perbaikan atas permasalahan yang terjadi di restoran Gudeg Yu Djum.

10. Kesimpulan dan Saran

Tahap terakhir yaitu merumuskan kesimpulan dan saran berdasarkan hasil dari penelitian yang dilakukan.

11. Selesai

BAB IV

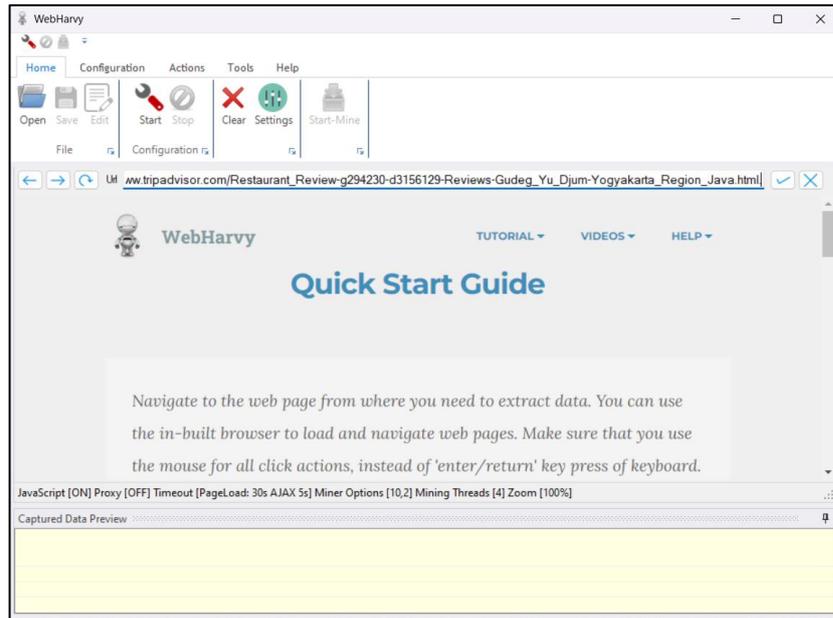
PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu ulasan pelanggan restoran Gudeg Yu Djum yang terdapat pada situs TripAdvisor. Teknik pengambilan data menggunakan teknik *web scraping* dengan bantuan aplikasi WebHarvy. Penelitian ini menggunakan aplikasi WebHarvy versi 5.5.0.168 untuk mengambil ulasan restoran Gudeg Yu Djum dan hasil pengambilan data disimpan ke dalam format *comma separated value* (.csv). Berikut merupakan langkah-langkah dalam melakukan pengumpulan data ulasan restoran Gudeg Yu Djum.

1. Membuka aplikasi WebHarvy

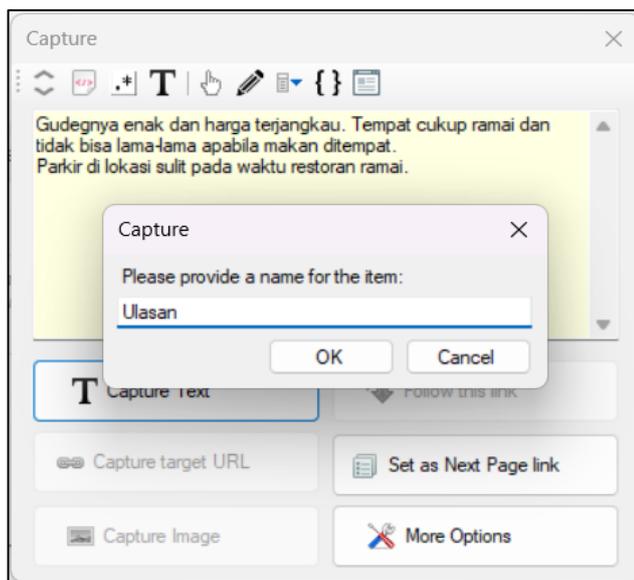
Langkah pertama dalam melakukan pengambilan data ulasan yaitu membuka aplikasi WebHarvy. Kemudian, masukkan alamat *domain* yang akan di *scraping* di kolom Url. Pada penelitian ini, alamat *domain* yang akan di *scraping* yaitu https://www.tripadvisor.com/Restaurant_Review-g294230-d3156129-Reviews-Gudeg_Yu_Djum-Yogyakarta_Region_Java.html.



Gambar 4.1 *Interface* Aplikasi WebHarvy

2. Melakukan proses *record* variabel data ulasan

Tahap selanjutnya yaitu melakukan proses *record* variabel pada data ulasan yang akan di *scraping*. Pada penelitian ini, variabel yang diambil pada ulasan restoran Gudeg Yu Djum yaitu nama, *rating*, tanggal ulasan, dan ulasan pelanggan. Contoh penerapan pada proses ini yaitu apabila ingin *record* variabel ulasan, lakukan klik pada ikon *Start*, kemudian klik ulasan yang ingin di *record*, dan akan muncul tampilan menu *capture*, dan klik *capture text*. Beri nama untuk variabel yang telah di *record*. Hasil dari proses *record* akan tersimpan dalam tampilan *capture data preview*.

Gambar 4.2 Proses *Record* Variabel Data Ulasan

3. Melakukan proses *scraping data*

Tahap terakhir yaitu melakukan proses *scraping data*. Untuk melakukan proses *scraping data*, klik ikon *start mine*, kemudian centang pada bagian *mine all pages* untuk melakukan penambangan seluruh data ulasan. Setelah proses penambangan data selesai, simpan data yang telah di *scraping* ke dalam format yang diinginkan. Pada penelitian ini, dilakukan penggalian data ulasan restoran Gudeg Yu Djum sebanyak 586 ulasan yang disimpan ke dalam format csv. Berikut merupakan hasil dari proses *scraping data*.

Tabel 4.1 Hasil *Scraping Data*

No.	Nama	Rating	Date	Ulasan
1,	BriandanaO	1	21 2017	Selama saya sekolah dulu, gudeg yu djum salah satu gudeg terkenal karena enak. Tapi mengapa saat saya kembali kesana, rasanya bikin eneg lalu kuku ayam tidak dipotong, bulu ayam masih byk yg nempel. Bukannya nambah selera makan tapi malah jadi ilfil untuk

makan. Bedalah sama yang dulu. (Bukan maksud untuk menjelekkkan tapi itu lah kenyataannya, ngg percaya? Buktikan sendiri)

-
- | | | | | |
|----|---------|---|------------|---|
| 2. | Jimmy S | 1 | 6 May 2016 | <p>Memang harga gudeg yu djum relatif murah, tapi pelayannya sungguh buruk. Kami adalah turis lokal dari surabaya yang sudah memesan terlebih dahulu unyuk makanan di depot yang TERKENAL ini. 1. Kami sudah reserved dan penyajian sungguh lambat. 2. Meja tidak dibersihkan saat kami duduk di sana. 3. Dengan SANGAT tidak sopan seorang pelayan datang ke meja kami dan menyuruh kami untuk STOP makan karena dia menyangka kami bukan rombongan yang sudah reserved 4. Tidak konsisten dengan perkataan STOP makan kalau bukan rombongan yg sudah reserved, restoran ini tetap melayani pengunjung lain yang datang. Yang saya sayangkan pelayan tersebut tidak merasa bersalah dan tidak mau minta maaf. Saya sangat kecewa, walaupun dengan rasa yang relatif masih oke, tentu saja makanan tidak akan terasa enak dengan pelayan yang tidak sopan bukan ?</p> |
|----|---------|---|------------|---|
-

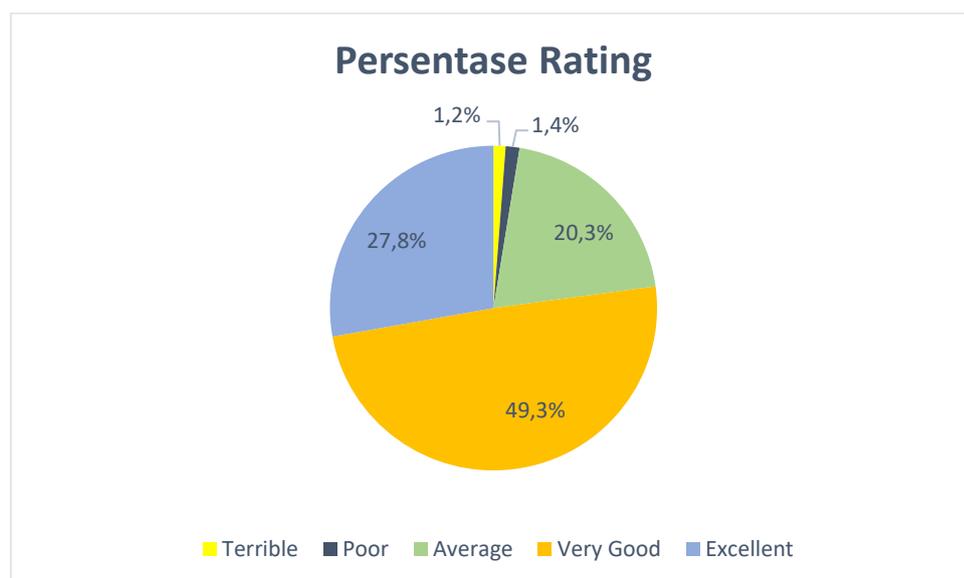
3.	lincelimowa	3	19 Juni 2014	Nama gudeg "yu djum " yang terkenal dijogja, tidaklah sebanding dengan pelayanan yang diberikan. terus terang kami sekeluarga sangat kecewa dengan pelayanan di warung ini. saran saya, kalau menerima pegawai diseleksi dengan baik, krn pelayanan yang buruk dapat merusak citra warung anda.
4.	Putri_Permatasari	5	23 April 2014	Kemarin saya makan di gudeg yu djum. Dekat bandara yogya...pelayan anda sangat tidak sopan dan kurang ajar!!!pesan blm selsai sudah ngloyor pergi, pas dipanggil lagi malah ngredumel "opo meneh", pesanannya salah terus, saya kasih tau malah dia njawab saya "NGEH" DENGAN NADA TINGGI!!!! Sebaiknya anda memberi pengertian kpd pekerja anda untu bersikap dan ungah ungguh...saya sangat marah, apalagi saya membawa orang tua dan anak saya!!!!saya ada fotonya!!!
5.	Michaela_Mukti	5	26 Februari 2016	Tidak cocok untuk yg punya lidah Sumatera. Mgkn klo tidak trlalu manis sy akan suka, krna pada dasarnya bumbunya enak.

4.2 Pengolahan Data

Data yang telah didapat pada proses *scraping* situs TripAdvisor selanjutnya dilakukan pengolahan data agar mendapatkan *knowledge* yang berguna dari data tersebut. Proses pengolahan data diawali dengan melakukan analisis deskriptif. Setelah melalui tahap analisis deskriptif, dilakukan proses pengolahan data dengan pendekatan *text mining* yaitu diawali dengan *preprocessing data* yang terdiri dari beberapa tahap, seperti *case folding*, *tokenizing*, *filtering*, *stemming* dan *term weighting*. Setelah melalui tahap *preprocessing data*, dilakukan analisis data menggunakan metode *K-Means Clustering*. Proses pengolahan data pada penelitian ini dapat dijabarkan sebagai berikut.

4.2.1 Analisis Deskriptif

Analisis deskriptif merupakan analisis yang dilakukan untuk mengetahui gambaran umum ulasan pelanggan mengenai restoran Gudeg Yu Djum yang ada pada situs TripAdvisor. Aspek yang ditampilkan pada analisis ini yaitu persentase jumlah ulasan berdasarkan *rating* yang diberikan oleh pelanggan restoran Gudeg Yu Djum. Berikut merupakan grafik persentase jumlah ulasan berdasarkan *rating* yang diberikan oleh pelanggan restoran Gudeg Yu Djum.



Gambar 4.3 Grafik Persentase *Rating* Ulasan

Berdasarkan gambar 4.4 dapat dilihat bahwa terdapat 5 kategori penilaian yang berbeda yang diberikan oleh pelanggan restoran Gudeg Yu Djum. Pelanggan memberikan *rating* dengan nilai 1 sampai dengan 5. *Rating* 1 termasuk dalam kategori “*terrible*”, *rating* 2 kategori “*poor*”, *rating* 3 kategori “*average*”, *rating* 4 kategori “*very good*”, dan *rating* 5 kategori “*excellent*”. Pada grafik diatas, dapat disimpulkan bahwa rata-rata pelanggan restoran Gudeg Yu Djum memberikan penilaian cukup baik terhadap layanan restoran tersebut. Hal ini dibuktikan dengan jumlah ulasan sebanyak 586 ulasan, 27,8% atau 163 pelanggan memberikan *rating* 5 atau *excellent*. Selanjutnya sebanyak 49,3% atau 289 pelanggan memberikan *rating* 4 atau *very good*. Kemudian diikuti dengan *rating* 3 atau *average* dengan persentase sebesar 20,3% atau sebanyak 119 pelanggan. Terakhir, masih terdapat pelanggan yang memberikan *rating* 2 atau *poor* terhadap layanan restoran dengan persentase sebesar 1,4% atau sebanyak 8 pelanggan, dan terakhir *rating* 1 atau *terrible* dengan persentase sebesar 1,2% atau sebanyak 7 pelanggan.

4.2.2 Preprocessing Data

Data yang digunakan pada penelitian ini merupakan data ulasan yang didapat pada proses *scraping* pada situs TripAdvisor. Data ulasan yang didapat masih memiliki berbagai jenis gaya penulisan dan format yang terdapat pada data masih dalam bentuk tidak terstruktur. Oleh karena itu, perlu dilakukan tahap *preprocessing data* untuk mengubah data menjadi lebih terstruktur dan memudahkan dalam melakukan tahap analisis selanjutnya. Tahap *preprocessing data* menggunakan bantuan aplikasi Jupyter Notebook dengan bahasa pemrograman Python dan *library* yang digunakan yaitu Sastrawi, *Natural Language Toolkit* (NLTK), *Regular Expression* (RegEx), *String*, *Numpy*, dan *Pandas*. *Natural Language Toolkit* (NLTK), *Regular Expression* (RegEx), dan *String* digunakan untuk pemrosesan awal pada data teks, terutama pada proses *case folding* dan *tokenizing*. Kemudian untuk proses *filtering* dan *stemming* menggunakan *library* Sastrawi karena cocok digunakan untuk teks Bahasa Indonesia (Alzami et al., 2020). Adapun tahap *pre-processing data* dapat dijabarkan sebagai berikut.

4.2.2.1 Case Folding

Tahap pertama pada *pre-processing data* yaitu *case folding*. Terdapat beberapa proses yang dilakukan pada *case folding* yaitu menyeragamkan huruf pada dokumen menjadi huruf kecil (*lowercase*), menghapus angka, dan menghapus tanda baca. Proses *case folding* dibantu dengan *library Natural Language Toolkit* (NLTK), *String*, dan *Regular Expression* (RegEx). Berikut merupakan contoh penerapan proses *case folding*.

Tabel 4.2 Proses *Case Folding*

No.	Input	Output
1.	Gudegnya enak dan harga terjangkau. Tempat cukup ramai dan tidak bisa lama-lama apabila makan ditempat. Parkir di lokasi sulit pada waktu restoran ramai.	gudegnya enak dan harga terjangkau tempat cukup ramai dan tidak bisa lama-lama apabila makan ditempat parkir di lokasi sulit pada waktu restoran ramai
2.	Mau ngerasain makanan Khas Jogja? Gudeg yg manis itu? Disini nih tempatnya... Gudeg Yu Jum! Enak, porsi besar.. Lauk banyak. Bisa dibawa buat leh oleh jg lhoo	mau ngerasain makanan khas jogja gudeg yg manis itu disini nih tempatnya gudeg yu jum enak porsi besar lauk banyak bisa dibawa buat leh oleh jg lhoo
3.	Setiap ke Kaliurang, saya selalu mampir sini. Untuk makan ditempat dan dibawa pulang. Pernah juga saya pesan untuk acara gathering, 300 pax. Aman, bisa delivery. Mantap!	setiap ke kaliurang saya selalu mampir sini untuk makan ditempat dan dibawa pulang pernah juga saya pesan untuk acara gathering pax aman bisa delivery mantap

4.2.2.2 Tokenizing

Tahap kedua pada *pre-processing data* yaitu proses *tokenizing*. *Tokenizing* merupakan proses pemotongan kalimat menjadi kata per kata. Proses *tokenizing* menggunakan *library Natural Language Toolkit* (NLTK). Berikut merupakan contoh penerapan proses *tokenizing*.

Tabel 4.3 Proses *Tokenizing*

<i>Input</i>	<i>Output</i>
gudegnya enak dan harga terjangkau tempat cukup ramai dan tidak bisa lama-lama apabila makan ditempat parkir di lokasi sulit pada waktu restoran ramai	['gudegnya', 'enak', 'dan', 'harga', 'terjangkau', 'tempat', 'cukup', 'ramai', 'dan', 'tidak', 'bisa', 'lamalama', 'apabila', 'makan', 'ditempat', 'parkir', 'di', 'lokasi', 'sulit', 'pada', 'waktu', 'restoran', 'ramai']

4.2.2.3 *Filtering*

Tahap ketiga pada *pre-processing data* yaitu *filtering*. *Filtering* merupakan tahap selanjutnya dari proses *tokenizing*. Proses yang dilakukan yaitu pemilihan kata-kata yang mengandung makna yang didapat dari hasil tokenisasi. Proses *filtering* disebut juga dengan *stopword removal*. *Library* yang digunakan untuk melakukan proses *filtering* yaitu *library* Sastrawi. Berikut merupakan contoh penerapan dari proses *filtering*.

Tabel 4.4 Proses *Filtering*

<i>Input</i>	<i>Output</i>
['gudegnya', 'enak', 'dan', 'harga', 'terjangkau', 'tempat', 'cukup', 'ramai', 'dan', 'tidak', 'bisa', 'lamalama', 'apabila', 'makan', 'ditempat', 'parkir', 'di', 'lokasi', 'sulit', 'pada', 'waktu', 'restoran', 'ramai']	['gudegnya', 'enak', 'harga', 'terjangkau', 'tempat', 'cukup', 'ramai', 'lamalama', 'tidak', 'bisa', 'apabila', 'makan', 'ditempat', 'parkir', 'lokasi', 'sulit', 'waktu', 'restoran', 'ramai']

4.2.2.4 *Stemming*

Tahap keempat pada *pre-processing data* yaitu *Stemming*. *Stemming* merupakan tahap dimana kata yang didapat pada proses *filtering* diubah ke dalam bentuk baku. *Library* yang digunakan pada proses *stemming* yaitu Sastrawi. Pada penelitian ini proses *stemming* digunakan karena kelebihan nya yaitu efektif dan efisien. Efektif dan efisien dalam arti menghasilkan dokumen yang lebih relevan dengan waktu yang sesingkat-singkatnya (Supitasari, 2016). Berikut merupakan contoh penerapan dari proses *stemming*.

Tabel 4.5 Proses *Stemming*

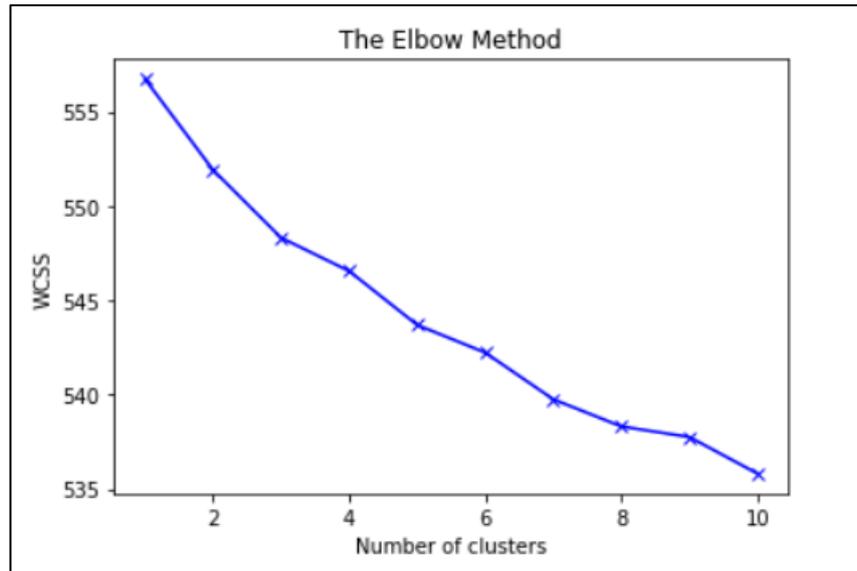
<i>Input</i>	<i>Output</i>
['gudegnya', 'enak', 'harga', 'terjangkau', 'tempat', 'tempat', 'cukup', 'ramai', 'lamalama', 'cukup', 'ramai', 'lamalama', 'apabila', 'apabila', 'makan', 'ditempat', 'parkir', 'makan', 'tempat', 'parkir', 'lokasi', 'sulit', 'lokasi', 'sulit', 'waktu', 'restoran', 'ramai']	['gudeg', 'enak', 'harga', 'jangkau', 'tempat', 'tempat', 'cukup', 'ramai', 'lamalama', 'apabila', 'apabila', 'makan', 'ditempat', 'parkir', 'makan', 'tempat', 'parkir', 'lokasi', 'sulit', 'waktu', 'restoran', 'ramai']

4.2.3 Clustering dengan Algoritma *K-Means Clustering*

Setelah melakukan tahap *pre-processing data*, langkah selanjutnya yaitu melakukan *clustering*. Proses *clustering* dibantu dengan aplikasi Jupyter Notebook dengan bahasa pemrograman yang digunakan yaitu Python. *Library* yang digunakan untuk melakukan proses *clustering* yaitu *scikit-learn* dan algoritma yang digunakan dalam mengelompokkan data ulasan yaitu algoritma *K-Means Clustering*. Adapun proses dalam melakukan *clustering* data ulasan dapat dijabarkan sebagai berikut.

4.2.3.1 Inisiasi Jumlah Cluster

Tahap awal dalam melakukan proses *clustering* menggunakan algoritma *K-Means Clustering* yaitu menentukan inisiasi nilai *k* yang akan digunakan. Pada penelitian ini, untuk menentukan nilai *k* menggunakan metode *Elbow*. Metode *elbow* merupakan metode dalam menentukan jumlah *cluster* terbaik dengan menghitung kuadrat jarak antara titik sampel di setiap *cluster* dengan pusat *cluster*. Metrik yang digunakan dalam perhitungan metode *elbow* yaitu metrik *Within Cluster Sum of Squares* (WCSS). Perhitungan metode *elbow* menggunakan aplikasi Jupyter Notebook dengan bahasa pemrograman yang digunakan yaitu Python. Berikut merupakan grafik perhitungan metode *elbow*.

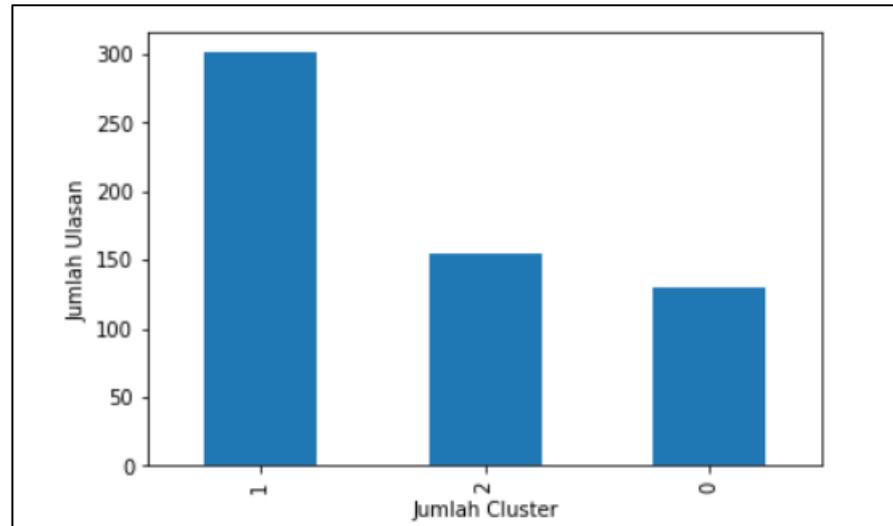


Gambar 4.4 Grafik *Elbow*

Berdasarkan gambar diatas, dapat dilihat hasil perhitungan menggunakan metode *elbow* dengan metrik *Within Cluster Sum of Squares* (WCSS). Diketahui nilai $k = 3$ merupakan titik siku yang terdapat pada grafik *elbow* dengan nilai WCSS sebesar 548,31. Maka, dapat disimpulkan bahwa inisiasi nilai k yang digunakan yaitu $k = 3$ untuk pengelompokan data ulasan. Setelah mendapatkan nilai k yang akan digunakan, maka dapat dilanjutkan ke tahap analisis *clustering*.

4.2.3.2 Analisis *Clustering*

Tahap selanjutnya yaitu melakukan proses *clustering*. Dalam melakukan proses *clustering* data ulasan, menggunakan algoritma *K-Means Clustering* dengan inisiasi jumlah *cluster* yang digunakan yaitu 3 *cluster*. Proses *clustering* dilakukan pada aplikasi Jupyter Notebook dengan bahasa pemrograman Python. Berikut merupakan grafik hasil *clustering* data ulasan restoran Gudeg Yu Djum.



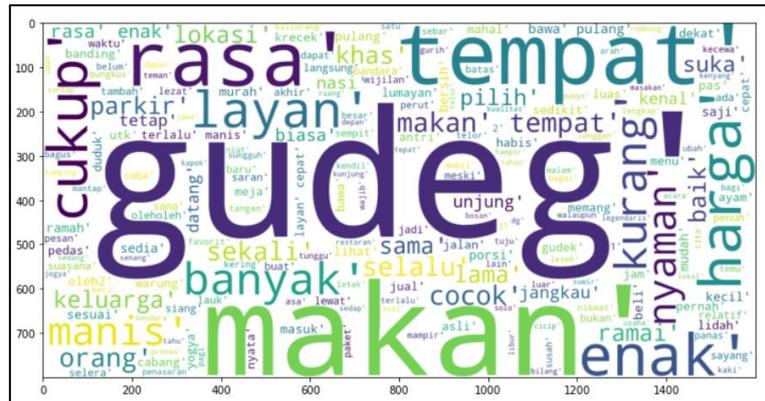
Gambar 4 5 Grafik *Clustering* Data Ulasan

Berdasarkan gambar diatas, dapat dilihat grafik pengelompokan data ulasan restoran Gudeg Yu Djum menggunakan algoritma *K-Means Clustering*. Dapat diketahui *cluster* data ulasan yang terbentuk sebanyak 3 *cluster* dengan jumlah ulasan pada masing-masing *cluster* yaitu *cluster* 0 sebanyak 130 ulasan, *cluster* 1 sebanyak 301 ulasan, dan *cluster* 2 sebanyak 155 ulasan. Setelah melakukan pengelompokan data ulasan, maka tahap selanjutnya yaitu melakukan visualisasi *cluster* untuk mengetahui informasi yang tersembunyi pada ulasan pelanggan restoran Gudeg Yu Djum.

4.2.4 Visualisasi *Cluster*

Setelah melakukan pengelompokan data ulasan, selanjutnya dapat dilakukan proses visualisasi *cluster* data ulasan. Visualisasi *cluster* digunakan untuk menampilkan informasi berupa kata-kata yang banyak dibicarakan oleh pelanggan mengenai restoran Gudeg Yu Djum. Proses dalam melakukan visualisasi data ulasan menggunakan visualisasi *wordcloud*. Berikut merupakan hasil visualisasi *cluster* ulasan restoran Gudeg Yu Djum.

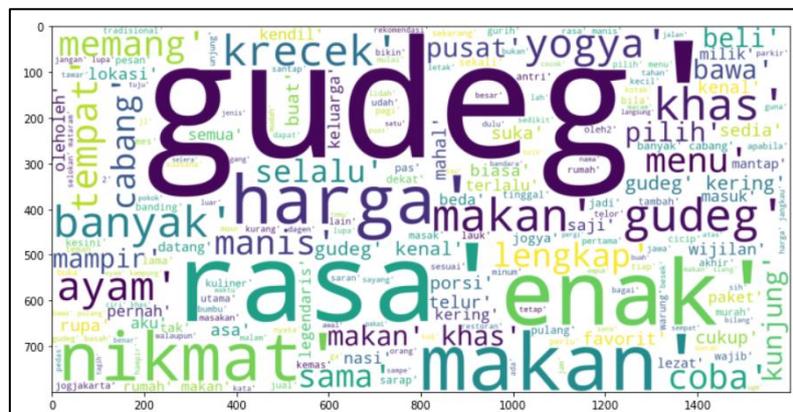
a. *Cluster* 0



Gambar 4.6 Visualisasi Wordcloud Cluster 0

Berdasarkan gambar 4.7 di atas, dapat dilihat bahwa topik-topik yang sering dibicarakan oleh pelanggan Gudeg Yu Djum. *Cluster 0* memiliki jumlah data ulasan sebanyak 130 ulasan dari 586 ulasan secara keseluruhan. Pada *Cluster 0* ini, kata-kata yang sering muncul yaitu kata ‘gudeg’, ‘makan’, ‘kurang’, ‘nyaman’, ‘harga’, ‘cukup’, ‘rasa’, ‘layan’, ‘enak’, dan ‘tempat’.

b. *Cluster 1*



Gambar 4.7 Visualisasi Wordcloud Cluster 1

Berdasarkan gambar 4.8 di atas, dapat dilihat kata-kata yang sering muncul yang dibicarakan oleh pelanggan restoran Gudeg Yu Djum. *Cluster 1* memiliki jumlah data ulasan sebanyak 301 ulasan dari 586 ulasan secara keseluruhan. Pada *Cluster 1* ini, mayoritas kata-kata yang sering muncul yaitu kata ‘makan’, ‘gudeg’, ‘rasa’, ‘enak’, ‘nikmat’, ‘harga’, ‘khas’, ‘banyak’, ‘yogya’, dan ‘kenal’.

confidence yang akan digunakan. Berikut merupakan tabel perbandingan nilai *support* dan *confidence*.

Tabel 4.6 Hasil Perbandingan Nilai *Support* dan *Confidence*

No.	Min. Support	Min. Confidence	Jumlah Rules		
			Cluster 0	Cluster 1	Cluster 2
1.	0,05	0,1	3.821	923	6.833
2.	0,03	0,1	25.420	3.539	31.620
3.	0,02	0,1	100.120	10.572	75.371

Berdasarkan tabel diatas, dapat dilihat nilai perbandingan nilai *support* dan *confidence* pada masing-masing *cluster*. Pada penelitian ini, nilai *support* dan *confidence* yang digunakan yaitu 0,02 dan 0,1 karena *rule* yang terbentuk banyak dan bervariasi. Berikut merupakan hasil asosiasi terhadap kata-kata yang terdapat pada masing-masing *cluster*.

A. Cluster 0

Berikut merupakan hasil asosiasi terhadap kata-kata yang terdapat pada *cluster* 0.

Tabel 4.7 Aturan Asosiasi *Cluster* 0

<i>Antecedents</i>	<i>Consequents</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
(terlalu,rasa,manis)	(gudeg)	0,061	1,000	1,340
(kurang,nyaman,tempat)	(makan)	0,030	0,666	1,125
(gudeg,rasa)	(kurang)	0,012	0,275	1,434
(tempat,kurang,panas)	(nyaman)	0,030	1,000	5,416
(gudeg,mahal,cukup)	(harga)	0,030	1,000	3,421
(gudeg,mahal)	(cukup)	0,030	0,500	2,407
(gudeg,enak,manis)	(rasa)	0,038	0,833	1,444
(ramah,kurang)	(layan)	0,023	0,600	3,000
(lama)	(layan)	0,030	0,250	1,250
(gudeg,rasa)	(enak)	0,018	0,413	1,014
(parkir,batas)	(tempat)	0,030	1,000	1,444

Berdasarkan tabel 4.6 diatas, dapat dilihat aturan asosiasi kata-kata yang sering muncul pada *cluster* 0. Informasi yang didapat pada *cluster* 0 dapat dijabarkan sebagai berikut.

1. Kata 'gudeg' berasosiasi dengan kata 'terlalu', 'rasa', dan 'manis'. Informasi yang didapat pada aturan ini yaitu pelanggan mengeluhkan rasa dari gudeg terlalu manis.
2. Kata 'makan' berasosiasi dengan kata 'kurang', 'nyaman', dan 'tempat'. Informasi yang didapat pada aturan ini yaitu pelanggan merasa kurang nyaman apabila makan di tempat restoran.
3. Kata 'kurang' berasosiasi dengan kata 'gudeg' dan 'rasa'. Informasi yang didapat pada aturan ini yaitu rasa dari gudeg telah berkurang dari sebelumnya.
4. Kata 'nyaman' berasosiasi dengan kata 'tempat', 'kurang', dan 'panas'. Informasi yang didapat pada aturan ini yaitu ketika pelanggan makan di tempat restoran, pelanggan merasa kurang nyaman karena panas.
5. Kata 'harga' berasosiasi dengan kata 'gudeg', 'mahal', dan 'cukup'. Informasi yang didapat pada aturan ini yaitu pelanggan mengeluhkan harga dari gudeg cukup mahal dari gudeg lainnya.
6. Kata 'cukup' berasosiasi dengan kata 'gudeg', dan 'mahal'. Informasi yang didapat pada aturan ini yaitu pelanggan merasa bahwa harga gudeg cukup mahal.
7. Kata 'rasa' berasosiasi dengan kata 'gudeg', 'enak', dan 'manis'. Informasi yang didapat pada aturan ini yaitu rasa makanan gudeg enak dan manis.
8. Kata 'layan' berasosiasi dengan kata 'ramah', dan 'kurang'. Informasi yang didapat pada aturan ini yaitu pelanggan kecewa karena pelayanan dari restoran kurang ramah. Kemudian juga berasosiasi dengan kata 'lama' yang berarti pelayanan dari restoran lama.
9. Kata 'enak' berasosiasi dengan kata 'gudeg', dan 'rasa'. Informasi yang didapat pada aturan ini yaitu pelanggan berpendapat rasa dari gudeg enak.
10. Kata 'tempat' berasosiasi dengan kata 'parkir', dan 'batas'. Informasi yang didapat pada aturan ini yaitu keluhan pelanggan mengenai tempat parkir terbatas.

B. *Cluster 1*

Berikut merupakan aturan asosiasi terhadap kata-kata yang terdapat pada *cluster 1*.

Tabel 4.8 Aturan Asosiasi *Cluster 1*

<i>Antecedents</i>	<i>Consequents</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
(gudeg, lengkap)	(makan)	0,069	0,656	1,400

(mahal,harga)	(gudeg)	0,059	1,000	1,030
(nikmat,khas,gudeg)	(rasa)	0,053	0,761	1,341
(manis,gudeg)	(enak)	0,073	0,448	1,239
(rasa,khas)	(nikmat)	0,053	0,301	1,933
(mahal)	(harga)	0,059	0,782	3,464
(bumbu,gudeg)	(khas)	0,029	0,562	2,039
(cabang)	(banyak)	0,086	0,590	2,615
(gudeg)	(yogya)	0,023	0,155	1,170
(gudeg)	(kenal)	0,014	0,154	1,030

Berdasarkan tabel 4.7 diatas, dapat dilihat aturan asosiasi kata-kata yang sering muncul pada *cluster* 1. Informasi yang didapat pada *cluster* 1 dapat dijabarkan sebagai berikut.

1. Kata 'makan' berasosiasi dengan kata 'gudeg' dan 'lengkap'. Informasi yang didapat pada aturan ini yaitu restoran ini memiliki varian menu gudeg yang lengkap.
2. Kata 'gudeg' berasosiasi dengan kata 'mahal' dan 'harga'. Informasi yang didapat pada aturan ini yaitu mengenai harga gudeg dari restoran Gudeg Yu Djum cukup mahal.
3. Kata 'rasa' berasosiasi dengan kata 'nikmat', 'khas', dan 'gudeg'. Informasi yang didapat pada aturan ini yaitu pendapat pelanggan mengenai gudeg yang memiliki rasa yang khas dan nikmat.
4. Kata 'enak' berasosiasi dengan kata 'manis' dan 'gudeg'. Informasi yang didapat pada aturan ini yaitu pendapat pelanggan mengenai rasa dari gudeg enak dan manis.
5. Kata 'nikmat' berasosiasi dengan kata 'rasa' dan 'khas'. Informasi yang didapat pada aturan ini yaitu mengenai cita rasa dari gudeg yang khas dan nikmat.
6. Kata 'harga' berasosiasi dengan kata 'mahal'. Informasi yang didapat pada aturan ini yaitu berkaitan dengan harga gudeg cukup mahal dari gudeg lainnya.
7. Kata 'khas' berasosiasi dengan kata 'bumbu' dan 'gudeg'. Informasi yang didapat pada aturan ini yaitu gudeg memiliki bumbu yang khas.
8. Kata 'cabang' berasosiasi dengan kata 'banyak'. Informasi yang didapat pada aturan ini yaitu Gudeg Yu Djum merupakan restoran yang memiliki banyak cabang yang tersebar di beberapa daerah di Yogyakarta.

9. Kata 'yogya' berasosiasi dengan kata 'gudeg'. Informasi yang didapat pada aturan ini yaitu Gudeg Yu Djum merupakan restoran gudeg ternama di Daerah Istimewa Yogyakarta,
10. Kata 'kenal' berasosiasi dengan kata 'gudeg'. Informasi yang didapat pada aturan ini yaitu gudeg yang terkenal di Daerah Istimewa Yogyakarta (DIY) yaitu Gudeg Yu Djum.

C. Cluster 2

Berikut merupakan aturan asosiasi terhadap kata-kata yang terdapat pada *cluster 2*.

Tabel 4.9 Aturan Asosiasi *Cluster 2*

<i>Antecedents</i>	<i>Consequents</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
(enak,krecek,pas)	(makan)	0,051	0,727	1,943
(kering,manis)	(gudeg)	0,012	1,000	1,115
(gurih,pedas)	(rasa)	0,051	1,000	1,741
(rasa,gudeg)	(manis)	0,458	0,865	1,082
(kering,krecek,manis,rasa)	(enak)	0,025	0,666	1,722
(masakan,gudeg,manis)	(suka)	0,077	0,800	2,339
(telur,pedas)	(krecek)	0,058	0,450	1,162
(masakan,gudeg,krecek,rasa)	(pas)	0,025	1,000	2,627
(rasa,lidah,gudeg,manis)	(terlalu)	0,064	0,588	1,599
(kering,manis,rasa)	(ayam)	0,051	0,666	2,296

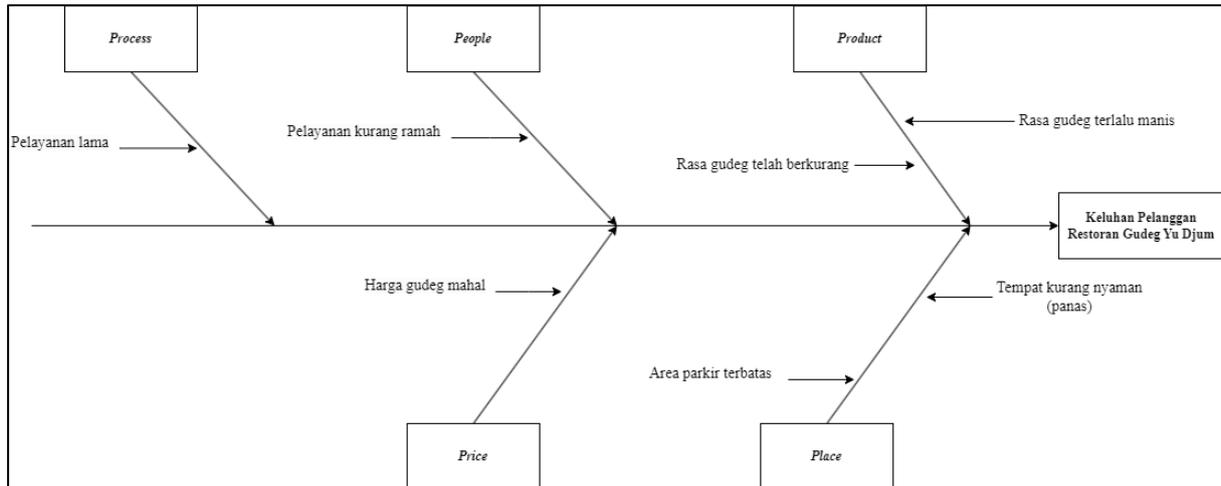
Berdasarkan tabel 4.8 diatas, dapat dilihat aturan asosiasi kata-kata yang sering muncul pada *cluster 2*. Informasi yang didapat pada *cluster 2* dapat dijabarkan sebagai berikut.

1. Kata 'makan' berasosiasi dengan kata 'enak', 'krecek', dan 'pas'. Informasi yang didapat pada aturan ini yaitu rasa dari makanan krecek enak dan pas.
2. Kata 'gudeg' berasosiasi dengan kata 'kering' dan 'manis'. Informasi yang didapat pada aturan ini yaitu tentang rasa makanan gudeg manis dan bertekstur kering.
3. Kata 'rasa' berasosiasi dengan kata 'gurih' dan 'pedas'. Informasi yang didapat pada aturan ini yaitu rasa dari makanan krecek gurih dan pedas.
4. Kata 'manis' berasosiasi dengan kata 'rasa' dan 'gudeg'. Informasi yang didapat pada aturan ini yaitu gudeg memiliki rasa yang manis.

5. Kata 'enak' berasosiasi dengan kata 'kering', 'krecek', 'manis', dan 'rasa'. Informasi yang didapat pada aturan ini yaitu pelanggan menyukai rasa dari krecek yang manis, enak dan bertekstur kering.
6. Kata 'suka' berasosiasi dengan kata 'masakan', 'gudeg' dan 'manis'. Informasi yang didapat pada aturan ini yaitu pelanggan menyukai rasa manis dari masakan gudeg.
7. Kata 'krecek' berasosiasi dengan kata 'telur' dan 'pedas'. Informasi yang didapat pada aturan ini yaitu telur dan krecek memiliki rasa yang pedas.
8. Kata 'pas' berasosiasi dengan kata 'masakan', 'gudeg', 'krecek', dan 'rasa'. Informasi yang didapat pada aturan ini yaitu rasa dari masakan gudeg dan krecek sangat pas.
9. Kata 'terlalu' berasosiasi dengan kata 'rasa', 'lidah', 'gudeg', dan 'manis'. Informasi yang didapat pada aturan ini yaitu rasa dari gudeg terlalu manis di lidah pelanggan.
10. Kata ayam berasosiasi dengan kata 'kering', 'manis', dan 'rasa'. Informasi yang didapat pada aturan ini yaitu rasa dari ayam nya manis dan bertekstur kering.

4.2.6 Analisis Diagram *Fishbone*

Berdasarkan hasil pengelompokan ulasan diatas, maka diperoleh informasi mengenai masalah yang terjadi terkait dengan ulasan negatif pada restoran Gudeg Yu Djum. Informasi yang didapat tersebut kemudian dianalisis menggunakan diagram *fishbone*. Hasil dari diagram *fishbone* dapat dilihat pada gambar berikut.



Gambar 4.9 Analisis Diagram *Fishbone*

Berdasarkan gambar diagram *fishbone* diatas, dapat diketahui penyebab dari munculnya ulasan negatif yang diberikan oleh pelanggan. Penyebab adanya ulasan negatif tersebut disebabkan oleh 5 faktor, yaitu faktor *product*, *people*, *process*, *place*, dan *price*. Permasalahan yang terdapat pada diagram didasarkan pada informasi yang didapat pada tahap visualisasi dan asosiasi kata pada masing – masing *cluster* yang telah dilakukan sebelumnya. Berikut merupakan informasi yang didapat pada aturan asosiasi kata pada masing – masing *cluster*.

Tabel 4.10 Permasalahan pada Restoran Gudeg Yu Djum

No.	Faktor	Permasalahan
1.	<i>Product</i>	- Rasa gudeg terlalu manis. - Rasa gudeg telah berkurang atau tidak konsisten.
2.	<i>People</i>	- Pelayanan kurang ramah.
3.	<i>Process</i>	- Pelayanan lama
4.	<i>Price</i>	- Harga gudeg mahal.
5.	<i>Place</i>	- Tempat restoran kurang nyaman (panas) - Area parkir terbatas.

BAB V

PEMBAHASAN

5.1 Hasil Analisa Deskriptif Ulasan Restoran Gudeg Yu Djum

Gudeg Yu Djum merupakan salah satu restoran yang berdiri pada tahun 1950 yang bergerak di bidang kuliner khas daerah yang berlokasi di Daerah Istimewa Yogyakarta. Produk yang ditawarkan berupa makanan khas daerah seperti Nasi Gudeg, Gudeg Kendil, Gudeg Besek, dan lain-lain. Restoran ini memiliki ulasan di situs TripAdvisor dengan jumlah ulasan sebanyak 586 ulasan Bahasa Indonesia. Restoran ini memiliki *rating* sebesar 4 dari 5. Secara *rating*, restoran ini memiliki performa yang cukup baik, namun ketika dilakukan analisis mendalam, masih terdapat ulasan negatif yang diberikan oleh pelanggan restoran Gudeg Yu Djum.

Pada penelitian ini dilakukan analisis deskriptif terhadap ulasan pelanggan restoran Gudeg Yu Djum untuk mengetahui gambaran umum mengenai kepuasan pelanggan terhadap pelayanan yang diberikan oleh restoran. Hasil dari analisis deskriptif menunjukkan bahwa rata-rata pelanggan restoran Gudeg Yu Djum memberikan penilaian cukup baik terhadap restoran Gudeg Yu Djum. Hasil ini dapat dilihat dengan jumlah ulasan sebanyak 586 ulasan, sebanyak 27,8% atau 163 pelanggan memberikan *rating* 5 atau *excellent*. Selanjutnya sebanyak 49,3% atau 289 pelanggan memberikan *rating* 4 atau *very good*. Kemudian diikuti dengan *rating* 3 atau *average* dengan persentase sebesar 20,3% atau sebanyak 119 pelanggan. Terakhir, masih terdapat pelanggan yang memberikan *rating* 2 atau *poor* terhadap restoran dengan persentasae sebesar 1,4% atau sebanyak 8 pelanggan, dan *rating* 1 atau *terrible* dengan persentase sebesar 1,2% atau sebanyak 7 pelanggan.

Berdasarkan hasil dari analisis deskriptif diatas, dapat dilihat masih terdapat pelanggan yang memberikan penilaian yang kurang baik terhadap restoran Gudeg Yu Djum. Maka dari itu, perlu adanya evaluasi terhadap layanan restoran untuk meningkatkan kualitas layanan dari restoran tersebut.

5.2 Hasil Klasterisasi Ulasan Menggunakan *K-Means Clustering*

Pada penelitian ini, dilakukan klasterisasi data ulasan restoran Gudeg Yu Djum yang terdapat pada situs TripAdvisor. Dalam melakukan proses klasterisasi data ulasan restoran, menggunakan metode *K-Means Clustering*. Proses klasterisasi data ulasan diawali dengan melakukan *pre-processing data*, yang terdiri dari beberapa proses yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Setelah melalui tahap *pre-processing data*, dilakukan tahap *term weighting* menggunakan metode TF-IDF agar memudahkan dalam proses analisis selanjutnya.

Setelah melalui tahap pembobotan kata, kemudian dilakukan analisis data ulasan menggunakan metode *K-Means Clustering*. Sebelum melakukan proses *clustering*, dilakukan penentuan banyak *cluster* yang akan dibuat terlebih dahulu menggunakan metode *elbow*. Dari hasil metode *elbow*, didapat bahwa dalam penelitian ini digunakan nilai $k = 3$ untuk klasterisasi data ulasan restoran Gudeg Yu Djum dengan nilai metrik *Within Cluster Sum of Squares* (WCSS) sebesar 548,31. Setelah mendapatkan nilai k akan digunakan, maka dapat dilanjutkan ke proses *clustering*.

Proses *clustering* data ulasan menggunakan metode *K-Means Clustering*. Dari proses *clustering* yang dilakukan, didapat hasil pengelompokan data ulasan sebanyak 3 *cluster* dengan jumlah ulasan pada masing-masing *cluster* yaitu *cluster 0* sebanyak 130 ulasan, *cluster 1* sebanyak 301 ulasan, dan *cluster 3* sebanyak 155 ulasan. Jika dibandingkan dengan penelitian terdahulu dengan topik serupa, terdapat perbedaan hasil yang didapatkan dalam melakukan proses *clustering* data ulasan. Penelitian yang dilakukan oleh Safitri et al. (2021) mengenai klasterisasi pada data sentimen kualitas jasa penerbangan mendapatkan hasil pengelompokan data sentimen sebanyak 2 *cluster*. Kemudian penelitian yang dilakukan oleh Saputra & Arianty (2019) mengenai klasterisasi sentimen keluhan pengguna Indosat juga mendapatkan hasil pengelompokan data ulasan sebanyak 2 *cluster*. Perbedaan hasil ini dapat terjadi karena pada penelitian terdahulu tersebut dalam menentukan jumlah *cluster* berdasarkan asumsi peneliti. Sedangkan pada penelitian ini, dalam menentukan jumlah *cluster* yang akan dibuat menggunakan metode *elbow* sehingga hasil *cluster* yang didapat berbeda.

5.3 Hasil Analisa Profil Cluster

Pada poin 5.2 sebelumnya telah dijelaskan bahwa *cluster* data ulasan restoran Gudeg Yu Djum yang terbentuk sebanyak 3 *cluster* menggunakan metode *K-Means Clustering* dengan jumlah ulasan pada masing-masing *cluster* yaitu *cluster* 0 sebanyak 130 ulasan, *cluster* 1 sebanyak 301 ulasan, dan *cluster* 3 sebanyak 155 ulasan. Setelah melakukan proses *clustering*, dilanjutkan dengan proses visualisasi *cluster* menggunakan *wordcloud*. Dalam menentukan profil dari *cluster* yang terbentuk, diawali dengan menemukan kata kunci yang terdapat pada *cluster* menggunakan hasil *output* dari *model cluster center*. *Model cluster center* ini mengurutkan kata-kata yang paling banyak muncul pada masing-masing *cluster*. Berikut merupakan kata kunci yang terdapat pada masing-masing *cluster*.

Tabel 5.1 Kata Kunci Cluster

<i>Cluster</i>	Kata Kunci
0	['tempat', 'makan', 'gudeg', 'rasa', 'enak', 'cukup', 'layan', 'harga', 'nyaman', 'kurang']
1	['gudeg', 'makan', 'rasa', 'enak', 'khas', 'banyak', 'nikmat', 'harga', 'yogya', 'kenal']
2	['manis', 'gudeg', 'pas', 'terlalu', 'krecek', 'suka', 'rasa', 'enak', 'ayam', 'makan']

Berdasarkan tabel diatas, dapat dilihat kata kunci untuk *cluster* 0 yaitu kata 'tempat', 'makan', 'gudeg', 'rasa', 'enak', 'cukup', 'layan', 'harga', 'nyaman', dan 'kurang'. Kemudian kata kunci untuk *cluster* 1 yaitu kata 'gudeg', 'makan', 'rasa', 'enak', 'khas', 'banyak', 'nikmat', 'harga', 'yogya', dan 'kenal'. Terakhir kata kunci untuk *cluster* 2 yaitu kata 'manis', 'gudeg', 'pas', 'terlalu', 'krecek', 'suka', 'rasa', 'enak', 'ayam', dan 'makan'. Dapat disimpulkan profil yang terbentuk pada masing-masing *cluster* yaitu *cluster* 0 merupakan kelompok ulasan yang membahas tentang rasa makanan, pelayanan, harga, dan kenyamanan. *Cluster* 1 merupakan kelompok ulasan yang membahas tentang menu, rasa, harga dan eksistensi restoran. Terakhir *cluster* 2 merupakan kelompok ulasan yang membahas tentang rasa dan tekstur dari gudeg dan krecek. Dari tabel diatas, dapat diketahui untuk kata kunci yang terbentuk pada masing – masing *cluster* hampir mirip satu sama lain. Contohnya yaitu pada *cluster* 0 terdapat kata yang sering muncul yaitu kata 'harga', kemudian kata ini juga muncul pada *cluster* 1. Hal ini dapat terjadi karena proses pengelompokan ulasan menggunakan metode *clustering* yang dimana data awal tidak

menggunakan label. Kemudian pada kata – kata yang sering muncul ini dilakukan proses asosiasi kata agar mendapatkan informasi yang lebih akurat.

Berdasarkan hasil dari asosiasi kata, diketahui informasi yang didapat pada masing-masing *cluster*. Karakteristik pelanggan terdapat pada *cluster* 0 cenderung membahas terkait suasana dan pelayanan yang diberikan. Informasi yang didapat pada *cluster* ini yaitu berbagai macam pelanggan membahas mengenai permasalahan yang terdapat pada rasa dari makanan gudeg. Beberapa pelanggan merasa kurang cocok dengan rasa dari gudeg yang dibuat oleh restoran. Faktor kurang cocok disebabkan karena rasa nya yang terlalu manis. Opini pelanggan lainnya yaitu pelanggan merasa bahwa rasa dari gudeg telah berkurang dari yang sebelumnya atau kualitas rasa dari gudeg telah menurun. Penilaian dari segi suasana, pelanggan merasa kurang nyaman yang disebabkan oleh beberapa faktor yaitu tempat restoran gelap dan suhu ruangan cukup panas ketika restoran sedang ramai. Kemudian pelanggan juga mengeluhkan terkait area parkir yang dimana area parkir di restoran sangat terbatas. Dari segi harga, pelanggan *cluster* ini merasa bahwa harga gudeg mahal dibandingkan dengan gudeg lainnya. Terakhir dari segi pelayanan yang diberikan, pelanggan kurang menyukai pelayanan dari restoran ini. Pelanggan menilai karyawan restoran kurang ramah dan juga pelayanan nya lama. Meskipun *cluster* ini informasi yang didapat cenderung negatif, masih terdapat beberapa penilaian positif yang diberikan oleh pelanggan. Penilaian positif yang diberikan oleh pelanggan yaitu beberapa pelanggan pada *cluster* ini merasa bahwa rasa gudeg masih tetap enak walaupun cenderung manis.

Karakteristik pelanggan *cluster* 1 yaitu cenderung membahas tentang menu dan eksistensi dari restoran itu sendiri. Informasi yang didapat pada *cluster* ini cenderung mengenai menu makanan yang terdapat pada restoran dapat dikatakan cukup lengkap. Pelanggan juga berpendapat bahwa rasa dari gudeg enak, manis dan bumbu nya sangat mencirikan gudeg khas Daerah Istimewa Yogyakarta. Kemudian restoran ini juga terkenal dan memiliki banyak cabang yang tersebar di Daerah Istimewa Yogyakarta. Penilaian negatif yang diberikan pelanggan pada *cluster* ini hanya pada aspek harga yang dimana pelanggan menilai hampir sama dengan pelanggan yang terdapat pada *cluster* 0 bahwa harga gudeg cukup mahal.

Karakteristik pelanggan yang terdapat pada *cluster* 2 cenderung membahas rasa dan tekstur dari gudeg dan krecek. Informasi yang didapat pada *cluster* ini hampir mirip dengan *cluster* 1. Pelanggan berpendapat rasa gudeg nya enak dan manis. Kemudian untuk makanan

krecek memiliki rasa pedas, untuk ayam nya memiliki tekstur kering dan manis. Dari hasil analisis asosiasi kata yang terbentuk pada *cluster* ini, tidak terdapat informasi mengenai keluhan-keluhan yang dialami oleh pelanggan.

Berdasarkan hasil profilisasi *cluster* diatas, dapat diketahui informasi penting pada masing-masing *cluster*. Pada penelitian ini, pihak restoran dapat menggunakan informasi yang terdapat pada *cluster* 0 untuk evaluasi perbaikan karena pada *cluster* tersebut informasi yang didapat mengenai keluhan-keluhan yang dialami oleh pelanggan.

5.4 Hasil Usulan Perbaikan Masalah

Dalam melakukan perancangan usulan perbaikan, diawali dengan mendapatkan informasi yang lebih akurat pada kata-kata yang terdapat pada masing-masing *cluster* yang terbentuk. Untuk mendapatkan informasi yang lebih akurat tersebut, dilakukan proses asosiasi kata. Proses asosiasi kata mengacu pada kata kunci yang terdapat pada masing-masing *cluster*. Setelah mendapatkan informasi dari kata kunci tersebut, kemudian melakukan analisis terhadap permasalahan yang terjadi di restoran Gudeg Yu Djum menggunakan metode diagram *fishbone*.

Hasil dari diagram *fishbone* menunjukkan beberapa permasalahan yang terjadi pada restoran. Beberapa permasalahan tersebut dibagi menjadi 5 faktor, faktor tersebut yaitu *people, place, price, process, dan product*. Setelah mendapatkan faktor-faktor yang menjadi permasalahan pada restoran, kemudian diberikan rancangan usulan perbaikan untuk meningkatkan kualitas layanan dari restoran tersebut. Rancangan usulan perbaikan yang dibuat berdasarkan usulan dari beberapa penelitian terdahulu yang membahas tentang peningkatan kualitas layanan restoran dan juga beberapa usulan dari peneliti. Berikut merupakan rancangan usulan perbaikan untuk restoran Gudeg Yu Djum.

Tabel 5.2 Rancangan Usulan Perbaikan Masalah

No.	Faktor	Permasalahan	Usulan Perbaikan
1.	<i>Product</i>	- Rasa gudeg terlalu manis. - Rasa gudeg telah berkurang atau tidak konsisten.	Permasalahan pada segi rasa, usulan perbaikan yang dapat diberikan yaitu membuat SOP terkait catatan mengenai takaran dari masing-masing menu yang terdapat pada

		restoran agar tidak terjadi rasa makanan yang inkonsisten.
2.	<i>People</i>	- Pelayanan kurang ramah.
		Permasalahan pada pelayanan dapat diberikan usulan perbaikan dengan memberikan pelatihan tentang <i>Service Excellent</i> . Melayani pelanggan dengan hormat, menyapa dengan wajah yang ramah dan mempersilahkan pelanggan untuk duduk dan menanyakan apa yang dibutuhkan oleh pelanggan (Renaldi & Mulyati, 2022).
3.	<i>Process</i>	- Pelayanan lama
		Untuk pelayanan lama, usulan perbaikan yang dapat diberikan yaitu memberikan pelatihan <i>Service Excellent</i> kepada semua karyawan.
4.	<i>Price</i>	- Harga gudeg mahal.
		Permasalahan pada aspek harga dapat dilakukan dengan cara memastikan bahan baku yang digunakan memiliki kualitas yang baik sehingga harga menyesuaikan bahan baku (Nofandi et al., 2023). Usulan lainnya seperti memberikan promo pada hari atau bulan tertentu.
5.	<i>Place</i>	- Tempat restoran kurang nyaman (panas) - Area parkir terbatas.
		Permasalahan pada tempat diatasi dengan cara menambahkan fasilitas pendingin ruangan seperti AC, kipas, dan lain-lain. Untuk parkir, dapat dilakukan dengan cara ekspansi lahan agar area parkir lebih luas.

BAB VI

PENUTUP

6.1 Kesimpulan

Adapun kesimpulan dari penelitian ini yaitu sebagai berikut.

1. Dilakukan proses *clustering* data ulasan restoran Gudeg Yu Djum menggunakan pendekatan algoritma *K-Means Clustering*. Proses *clustering* data ulasan diawali dengan menentukan jumlah *k* yang akan digunakan menggunakan metode *elbow*. Berdasarkan hasil metode *elbow*, didapat nilai $k = 3$ dengan nilai metrik *Within Cluster Sum of Squares* (WCSS) sebesar 548,31. Berdasarkan hasil algoritma *K-Means Clustering*, didapat hasil klasterisasi data ulasan restoran Gudeg Yu Djum sebanyak 3 *cluster* dengan jumlah ulasan pada masing-masing *cluster* yaitu *cluster 0* sebanyak 130 ulasan, *cluster 1* sebanyak 301 ulasan, dan *cluster 2* sebanyak 155 ulasan. Pada *cluster 0*, kata-kata yang sering muncul yaitu kata 'gudeg', 'makan', 'kurang', 'nyaman', 'harga', 'cukup', 'rasa', 'layan', 'enak', dan 'tempat'. Kemudian pada *cluster 1*, kata-kata yang sering muncul yaitu 'makan', 'gudeg', 'rasa', 'enak', 'nikmat', 'harga', 'khas', 'banyak', 'yogya', dan 'kenal'. Dan terakhir *cluster 2* dengan kata-kata yang sering muncul yaitu 'makan', 'gudeg', 'rasa', 'manis', 'enak', 'suka', 'krecek', 'pas', 'terlalu', dan 'ayam'. Profil yang terdapat pada masing-masing *cluster* yaitu *cluster 0* merupakan kelompok ulasan yang membahas tentang rasa makanan, pelayanan, harga, dan kenyamanan. Kemudian *cluster 1* merupakan kelompok ulasan yang membahas tentang menu, rasa, harga, dan eksistensi restoran. Dan *cluster 2* merupakan kelompok ulasan yang membahas tentang rasa dan tekstur dari gudeg dan krecek.
2. Berdasarkan hasil klasterisasi data ulasan, dilakukan analisis menggunakan diagram *fishbone* untuk perbaikan kualitas layanan restoran. Berdasarkan hasil diagram *fishbone*, didapat 7 permasalahan yang terdapat pada restoran Gudeg Yu Djum yang dibagi menjadi 4 faktor yaitu faktor *people*, *place*, *price*, dan *product*. Tujuh permasalahan tersebut antara lain rasa gudeg telah berkurang, rasa gudeg terlalu manis, pelayanan

kurang ramah, pelayanan lama, harga gudeg mahal, tempat restoran kurang nyaman (panas), dan area parkir terbatas.

3. Usulan yang dapat diberikan untuk perbaikan kualitas layanan restoran Gudeg Yu Djum yaitu untuk faktor *product* yaitu membuat SOP terkait catatan mengenai takaran dari masing-masing menu yang terdapat pada restoran agar tidak terjadi rasa makanan yang inkonsisten. Kemudian untuk faktor *people* memberikan pelatihan *service excellent*, melayani pelanggan dengan hormat, menyapa dengan wajah yang ramah dan mempersilahkan pelanggan untuk duduk dan menanyakan apa yang dibutuhkan oleh pelanggan. Kemudian untuk faktor *process* yaitu juga memberikan pelatihan *service excellent* kepada semua karyawan. Kemudian untuk faktor *price* memastikan bahan baku yang digunakan memiliki kualitas baik dan harga menyesuaikan harga bahan baku, kemudian memberikan promo pada hari atau bulan tertentu. Terakhir untuk faktor *place* menambahkan fasilitas pendingin ruangan seperti AC, kipas, dan lain-lain. Untuk permasalahan parkir dapat dilakukan dengan cara ekspansi lahan agar area parkir lebih luas.

6.2 Saran

Saran yang dapat diberikan berdasarkan hasil penelitian yang telah dilakukan yaitu sebagai berikut.

1. Bagi Restoran

Penelitian yang dilakukan menghasilkan klasterisasi data ulasan mengenai restoran. Dari hasil klasterisasi yang didapat, masih terdapat ulasan negatif yang diberikan oleh pelanggan. Berdasarkan ulasan negatif yang didapat, diberikan usulan perbaikan untuk mengatasi permasalahan tersebut. Hasil usulan perbaikan dapat dijadikan sebagai bahan pertimbangan bagi restoran untuk meningkatkan kualitas layanan restoran.

2. Bagi Peneliti Selanjutnya

- a. Peneliti menyadari bahwa masih terdapat kekurangan dari penelitian yang dilakukan. Saran yang dapat diberikan untuk peneliti selanjutnya yaitu dapat menggunakan parameter lain dalam menentukan jumlah *cluster* optimal seperti metode *Silhouette coefficient*, *Davies Bouldin Index*, dan *Adjusted Rand Index Score*. Kemudian juga dapat menggunakan metode *clustering* lainnya seperti DBSCAN

dan *Spectral Clustering*. Pada penelitian ini juga terbatas pada ulasan Bahasa Indonesia, maka dari itu untuk peneliti selanjutnya dapat menganalisis seluruh data ulasan berdasarkan objek yang diteliti. Kemudian dalam pembobotan kata, dapat menggunakan metode lain seperti *Bag of Words* (BoW) dan *N-Gram*.

- b. Peneliti selanjutnya diharapkan dapat menganalisis data ulasan berdasarkan periode waktu tertentu sehingga dapat memunculkan analisis baru mengenai topik yang sering dibicarakan berdasarkan periode waktu tertentu.

DAFTAR PUSTAKA

- Adriansah, R., & Santoso, I. (2019). *ANALISIS SENTIMEN BADAN PUSATSTATISTIK Sentiment Analysis of Badan Pusat Statistik Based on Online Media*. 217–225.
- Afidah, D. I., Dairoh, D., Handayani, S. F., Pratiwi, R. W., & Sari, S. I. (2022). Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 607–618. <https://doi.org/10.30812/matrik.v21i3.1402>
- Afida, D., Udayanti, E. D., & Kartikadarma, E. (2021). Aplikasi Text Mining untuk Klasterisasi Aduan Masyarakat Kota Semarang Menggunakan Algoritma K-means. *Jurnal Transformatika*, 18(2), 215. <https://doi.org/10.26623/transformatika.v18i2.2362>
- Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem Dan Informatika*, 3(Pebruari), 47–60.
- Alzami, F., Paundrianagari, N. P., Pramunendar, R. A., Megantara, R. A., & Prabowo, D. P. (2020). Sentiment Analysis Untuk Deteksi Ujaran Kebencian Pada Domain Politik. *Science and Engineering National Seminar*, 5(Sens 5), 213–218. <http://conference.upgris.ac.id/index.php/sens/article/view/1606%0Ahttp://conference.upgris.ac.id/index.php/sens/article/download/1606/711>
- Amardita, R. S., Adiwijaya, A., & Purbolaksono, M. D. (2022). Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN. *JURIKOM (Jurnal Riset Komputer)*, 9(1), 62. <https://doi.org/10.30865/jurikom.v9i1.3793>
- Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, 11(2), 149–164. <https://doi.org/10.15408/jti.v11i2.8623>
- Azzahra, S. A., & Wibowo, A. (2020). Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi dengan Algoritme Naïve Bayes untuk Ulasan Wisata Kuliner Pada Web Tripadvisor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 737. <https://doi.org/10.25126/jtiik.2020731907>
- Bangoria, B., Mankad, N., & Pambhar, V. (2013). A survey on Efficient Enhanced K-Means Clustering Algorithm. *IJSRD -International Journal for Scientific Research & Development*, 1(9online), 2321–2613.
- Bose, I., & Mahapatra, R. K. (2001). Business data mining - A machine learning perspective. *Information and Management*, 39(3), 211–225. [https://doi.org/10.1016/S0378-7206\(01\)00091-X](https://doi.org/10.1016/S0378-7206(01)00091-X)
- Chen, Y., Liu, D., Liu, Y., Zheng, Y., Wang, B., & Zhou, Y. (2022). Research on user generated content in Q&A system and online comments based on text mining. *Alexandria Engineering Journal*, 61(10), 7659–7668. <https://doi.org/10.1016/j.aej.2022.01.020>

- Cholik, C. A. (2021). *PERKEMBANGAN TEKNOLOGI INFORMASI KOMUNIKASI / ICT DALAM BERBAGAI BIDANG*. 2(February), 6.
- Cholissodin, I., & Soebroto, A. A. (2021). *AI, MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi)*. December.
- Coccia, M. (2018). The Fishbone Diagram to Identify, Systematize and Analyze the Sources of General Purpose Technologies. *Journal of Social and Administrative Sciences*, 4(4), 291–303.
<https://ssrn.com/abstract=3100011> Electronic copy available at: <https://ssrn.com/abstract=3100011> Electronic copy available at: <https://ssrn.com/abstract=3100011>
- Cui, M. (2020). *Introduction to the K-Means Clustering Algorithm Based on the Elbow Method*. 5–8. <https://doi.org/10.23977/accaf.2020.010102>
- Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text Mining Methods and Techniques. *International Journal of Computer Applications*, 85(17), 42–45.
<https://doi.org/10.5120/14937-3507>
- Gurusamy, V., & Kannan S. (2014). Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Guruvayur, S. R., & Suchithra, R. (2018). A detailed study on machine learning techniques for data mining. *Proceedings - International Conference on Trends in Electronics and Informatics, ICEI 2017, 2018-Janua*, 1187–1192.
<https://doi.org/10.1109/ICOEI.2017.8300900>
- Hamidy, F. (2016). Pendekatan Analisis Fishbone Untuk Mengukur Kinerja Proses Bisnis Informasi E-Koperasi. *Jurnal Teknoinfo*, 10(1), 11. <https://doi.org/10.33365/jti.v10i1.12>
- Han, J., Kamber, M., & Pei, J. (2012). Third Edition : Data Mining Concepts and Techniques. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
<http://library.books24x7.com/toc.aspx?bkid=44712>
- Horvitz, E., & Mulligan, D. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 253–255.
- Huang, Y., Huang, W. J., Xiang, X. L., & Yan, J. J. (2021). An empirical study of personalized advertising recommendation based on DBSCAN clustering of sina weibo user-generated content. *Procedia Computer Science*, 183, 303–310.
<https://doi.org/10.1016/j.procs.2021.02.063>
- 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.
- Indonesia, A. P. J. I. (2023). *Survei Penetrasi & Perilaku Internet 2023*. 2–63.
- Indrayuni, E. (2017). Text Mining dalam Analisis Sentimen Review Restoran Menggunakan Algoritma K-Nearest-Neighbor (KNN). *Agustus, III(2)*, 109. www.tripadvisor.com.
- Irsyad, H., & Pribadi, M. R. (2020). Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means. *Kurawal - Jurnal Teknologi, Informasi Dan Industri*, 3(2), 164–172. <https://doi.org/10.33479/kurawal.v3i2.347>

- Izzatillah, M., & Mutiara, A. B. (2020). Pengujian Algoritma Clustering Affinity Propagation dan Adaptive Affinity Propagation terhadap IPK dan Jarak Rumah. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 4(3), 330. <https://doi.org/10.30998/string.v4i3.6197>
- Jiachen, W., Yaling, W., & Yue, W. (2019). Applying Clustering and Co-occurrence Methods to Identifying Key Events and Their Relations in Chinese Stock Market. *Procedia Computer Science*, 162(61309030), 102–110. <https://doi.org/10.1016/j.procs.2019.11.264>
- Julianto, Y., Setiabudi, D. H., & Rostianingsih, S. (2022). Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine. *Jurnal Infra*, 10(1), 1–7.
- Kaur, J., & Madan, N. (2015). Association Rule Mining: A Survey. *International Journal of Hybrid Information Technology*, 8(7), 239–242. <https://doi.org/10.14257/ijhit.2015.8.7.22>
- Kumbhare, T. A., & V Chobe, S. (2014). An Overview of Association Rule Mining Algorithms. *International Journal of Computer Science and Information Technologies(IJCSIT) International Journal of Computer Science and Information Technologies*, 5(1), 927–930. www.ijcsit.com
- Kuswardana. (2017). Analisis Penyebab Kecelakaan Kerja Menggunakan Metode RCA (Fishbone Diagram Method And 5 – Why Analysis) di PT . PAL Indonesia. *Conference on Safety Engineering and Its Application*, 141–146.
- Langgeni, D. P., Baizal, Z. K. A., & W, Y. F. A. (2010). Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. *Seminar Nasional Informatika 2010, 2010(semnasIF)*, 1–10.
- Liu, F., & Deng, Y. (2021). Determine the Number of Unknown Targets in Open World Based on Elbow Method. *IEEE Transactions on Fuzzy Systems*, 29(5), 986–995. <https://doi.org/10.1109/TFUZZ.2020.2966182>
- Madhulatha, T. S. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6), 583–605. <https://doi.org/10.3233/ida-2007-11602>
- Megawati, S. (2021). Pengembangan Sistem Teknologi Internet of Things Yang Perlu Dikembangkan Negara Indonesia. *Journal of Information Engineering and Educational Technology*, 5(1), 19–26. <https://doi.org/10.26740/jieet.v5n1.p19-26>
- Naf'an, M. Z., Burhanuddin, A., & Riyani, A. (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen. *Jurnal Linguistik Komputasional*, 2(1), 23–27.
- Nasim, Z., & Haider, S. (2022). Cluster analysis of urdu tweets. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 2170–2179. <https://doi.org/10.1016/j.jksuci.2020.08.008>
- Nasution, E. Y., Hariani, P., Hasibuan, L. S., & Pradita, W. (2020). Perkembangan Transaksi Bisnis E-Commerce terhadap Pertumbuhan Ekonomi di Indonesia. *Jesya*, 3(2), 506–519. <https://doi.org/10.36778/jesya.v3i2.227>
- Nofandi, A., Setiawan, N. Y., & Brata, D. W. (2023). Analisis sentimen ulasan pelanggan dengan Metode Support Vector Machine (SVM) untuk peningkatan kualitas layanan pada Restoran Warung Wareg. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(1), 458–466. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12218>

- Omar, A. (2020). Feature selection in text clustering applications of literary texts: A hybrid of term weighting methods. *International Journal of Advanced Computer Science and Applications*, 11(2), 99–107. <https://doi.org/10.14569/ijacsa.2020.0110214>
- Pawar, S., Manjula Gururaj, H., & Chiplunar, N. N. (2022). Text Summarization Using Document and Sentence Clustering. *Procedia Computer Science*, 215, 361–369. <https://doi.org/10.1016/j.procs.2022.12.038>
- Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scraping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi Laman Web Tentang Hidroponik. *Indonesian Journal of Information Systems (IJIS)*, 1(1), 25–33.
- Probierz, B., Kozak, J., & Hrabia, A. (2022). Clustering of scientific articles using natural language processing. *Procedia Computer Science*, 207(Kes), 3443–3452. <https://doi.org/10.1016/j.procs.2022.09.403>
- Rajak, A., & Gupta, M. K. (2008). Association Rule Mining: Applications in Various Areas. *International Conference on Data Management, January 2008*, 3–7. <https://pdfs.semanticscholar.org/e5db/a964a791763d7cd9c9d8979f2c00604e7b9a.pdf>
- Reddy, S., Bhaskar, R., Padmanabhan, S., Verspoor, K., Mamillapalli, C., Lahoti, R., Makinen, V.-P., Pradhan, S., Kushwah, P., & Sinha, S. (2021). Use and validation of text mining and cluster algorithms to derive insights from Corona Virus Disease-2019 (COVID-19) medical literature. *Computer Methods and Programs in Biomedicine Update*, 1(February), 100010. <https://doi.org/10.1016/j.cmpbup.2021.100010>
- Renaldi, R., & Mulyati, D. S. (2022). Usulan Perbaikan Kualitas Pelayanan Restoran Menggunakan Metode Servqual dan Kano. *Jurnal Riset Teknik Industri*, 109–116. <https://doi.org/10.29313/jrti.v2i2.1245>
- Sabna, E., Mustika, B., Fonda, H., Irfan, D., Hang, S., & Pekanbaru, T. (2020). Text Mining Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Keinginan Pasar Terkait Perjalanan Wisata Text Mining Uses K-Means Clustering Algorithm To Predict Market Desires for Tourism Travel. *Journal of Information Technology and Computer Science (INTECOMS)*, 3(2), 380–386.
- Safitri, S. I., Suhery, C., & Bahri, S. (2021). Implementasi Algoritma K-Means Untuk Clustering Sentimen Pada Opini Kualitas Pelayanan Jasa Penerbangan. *Coding Jurnal Komputer Dan Aplikasi*, 09(02), 186–197. <https://jurnal.untan.ac.id/index.php/jcskommipa/article/view/47377>
- Sagayam, R., Srinivasan, S., & Roshni, S. (2012). A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques. *International Journal Of Computational Engineering Research (Ijceronline.Com)*, 2(5), 2250–3005. <http://pakademicsearch.com/pdf-files/com/319/1443-1446> Volume 2, Issue 5, September, 2012.pdf
- Salman, N. (2023). *Density-Based Clustering Analysis*. 8, 1–8.
- Sanko, G., Wasino, S., & Sutrisno, T. (2023). KLASIFIKASI SENTIMEN URAIAN PENGGUNA APLIKASI BCA MOBILE PADA PLATFORM GOOGLE PLAY STORE DENGAN ALGORITMA K-MEANS CLUSTERING. *Jurnal Ilmu Komputer Dan Sistem Informasi*.

- Saputra, T. I., & Arianty, R. (2019). Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat. *Jurnal Ilmiah Informatika Komputer*, 24(3), 191–198. <https://doi.org/10.35760/ik.2019.v24i3.2361>
- Sentiya, A., & Suroyo, H. (2019). Analisis Text Clustering Akun Fanpage Shopee Indonesia Dengan Komentar Followers Menggunakan Tools Orange Data Mining. *Bina Darma Conference on ...*, 1055–1067. <http://conference.binadarma.ac.id/index.php/BDCCS/article/view/660>
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2(October 2022), 100003. <https://doi.org/10.1016/j.nlp.2022.100003>
- Sholehhudin, M., Fauzi Ali, M., & Adinugroho, S. (2018). *Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus : Universitas Brawijaya)*. 2(11), 5518–5524.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Singgalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1343–1352. <https://doi.org/10.47065/bits.v4i3.2486>
- Singh, V. K., Tiwari, N., & Garg, S. (2011). Document clustering using K-means, heuristic K-means and fuzzy C-means. *Proceedings - 2011 International Conference on Computational Intelligence and Communication Systems, CICN 2011*, 297–301. <https://doi.org/10.1109/CICN.2011.62>
- Siregar, L. Y., & Nasution, M. I. P. (2020). Perkembangan Teknologi Informasi Terhadap Peningkatan Bisnis Online. *HIRARKI Jurnal Ilmiah Manajemen Dan Bisnis (HJIMB)*, 02(01), 71–75. <http://journal.upp.ac.id/index.php/Hirarki%0APERKEMBANGAN>
- Škrjanc, I., Andonovski, G., Iglesias, J. A., Sesmero, M. P., & Sanchis, A. (2022). Evolving Gaussian on-line clustering in social network analysis. *Expert Systems with Applications*, 207(June), 117881. <https://doi.org/10.1016/j.eswa.2022.117881>
- Somantri, O. (2019). Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining. *Jepin*, 5(2), 191–196. www.google.com/maps
- Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). *A Review of Machine Learning Techniques using Decision Tree and Support Vector Machine*.
- Supianto, A. A., Nurdiansyah, R., Weng, C. W., Zilvan, V., Yuwana, R. S., Arisal, A., Pardede, H. F., Lee, M. M., Huang, C. H., & Ng, K. L. (2023). Cluster-based text mining for extracting drug candidates for the prevention of COVID-19 from the biomedical literature. *Journal of Taibah University Medical Sciences*, 18(4), 787–801. <https://doi.org/10.1016/j.jtumed.2022.12.015>
- Supitasari, E. (2016). *Peningkatan Algoritma Stemming Porter Bahasa Indonesia Berdasarkan Metode Morfologi dengan Mengaplikasikan Aturan Kombinasi Awalan Akhiran*. 1210651022. <http://repository.unmuhjember.ac.id/id/eprint/468>
- Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based framework for text

- categorization. *Procedia Engineering*, 69, 1356–1364. <https://doi.org/10.1016/j.proeng.2014.03.129>
- Utami, N. W., & Eka Putra, I. G. J. (2022). Text Mining Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity. *Jurnal Informatika Teknologi Dan Sains*, 4(3), 255–259. <https://doi.org/10.51401/jinteks.v4i3.1907>
- Wakhidah, N. (2014). Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering). *Fakultas Teknologi Informasi*, 21(1), 70–80.
- Wijaya, D. V, Alfiansyah, Y. F., & ... (2021). Analisis Sentimen Pada Buletin Menggunakan Algoritme DbSCAN. ... *on Innovation and ...*, *Ciastech*, 417–426. <http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/3337%0Ahttp://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/viewFile/3337/1794>
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, 2(2), 226–235. <https://doi.org/10.3390/j2020016>
- Zuliani, A., Contiero, B., Schneider, M. K., Arsenos, G., Bernués, A., Dovc, P., Gauly, M., Holand, Martin, B., Morgan-Davies, C., Zollitsch, W., & Cozzi, G. (2021). Topics and trends in Mountain Livestock Farming research: a text mining approach. *Animal*, 15(1). <https://doi.org/10.1016/j.animal.2020.100058>

LAMPIRAN

A-1 Lampiran *Script Python Preprocessing Data*

```
In [1]: import pandas as pd
import numpy as np
import re
import string
import nltk
```

```
In [2]: data = pd.read_csv('DATASKRIPSI_FIX.csv')
data
```

```
Out[2]:
```

	Nama	Rating	Date	Ulasan
0	HasanYogyakarta	40	November 10, 2022	Gudegnya enak dan harga terjangkau. Tempat cuk...
1	tariatya	40	October 8, 2021	Rasa bumbu gudegnya istimewa, apalagi telur pi...
2	Oemaristan	30	May 30, 2021	Gudeg yang mainstream sebenarnya selain karena...
3	ayun860	50	December 27, 2020	gudeg nya enak, harga nya affordable, lokasi n...
4	restiw2020	50	December 24, 2020	Gudegnya enak bgt, khas jogja bgt, aslik buat ...
...
581	Avan1175	50	November 24, 2012	Nama ini selalu mengusik telinga, penasaran de...
582	Hastyo	50	November 19, 2012	rasa gudegnya yang manis sangat pas dengan ras...
583	PuriMesariHotel	20	October 3, 2012	Kami makan gudeg Yu Djum yang dekat Whiz Hotel...
584	MasMarjana	40	August 28, 2012	Jogja identik dengan kota gudeg. Belum ke Jogi...
585	Ruslimaem	30	July 10, 2012	salah satu gudeg tua yang masih bertahan di Jo...

586 rows x 4 columns

Case Folding

```
In [6]: def preprocess_text (kalimat) :
# Mengubah kalimat menjadi huruf kecil
lower_case = kalimat.lower()

# Menghapus angka dari kalimat
hasil = re.sub(r"/d+", "", lower_case)

# Menghapus tanda baca dari kalimat
hasil = hasil.translate (str.maketrans("", "", string.punctuation))

# Menghapus spasi di awal dan akhir kalimat
hasil = hasil.strip()

return hasil
```

```
In [7]: df['Ulasan_clean'] = df['Ulasan'].apply(preprocess_text)
df.head()
```

Tokenization ¶

```
In [8]: from nltk.tokenize import word_tokenize

def tokenize_text (kalimat) :
tokens = nltk.tokenize.word_tokenize(kalimat)
return tokens
```

```
In [9]: df['token'] = df['Ulasan_clean'].apply(tokenize_text)
df.head()
```

Filtering

```
In [15]: from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()+more_stopword
more_stopword = ['gudeg','makan','krecek','yg','yu','djum','jogja','yogyakarta','kangen','beberapa','aja','sangat','bungkus','pai']
print (stopwords)
```

```
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingg
a', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'se
kitar', 'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa',
'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'and
a', 'begitu', 'menga', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'dimana', 'keman
a', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai', 'sedangkan', 'selagi', 'sementara', 'teta
pi', 'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst',
'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nanti', 'melainkan', 'oh', 'ok', 'seh
arusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'a
palagi', 'bagaimanapun', 'gudeg', 'makan', 'krecek', 'yg', 'yu', 'djum', 'jogja', 'yogyakarta', 'kangen', 'beberapa', 'aja', 's
angat', 'bungkus', 'paling', 'sini', 'bgt', 'banget', 'hari', 'bawa', 'jadi', 'salah', 'satu', 'kalo', 'tp', 'hari', 'nama', 'm
aupun', 'buat', 'deh', 'nya', 'kalau', 'mungkin', 'kata', 'kota', 'buat', 'jakarta', 'beberapa', 'sendiri', 'asa', 'coba', 'ma
u', 'lebih', 'buat', 'jalan', 'malioboro', '2x', 'aja', 'hari', 'klo', 'kenal', 'kali', 'sy', 'lain', 'cari', 'lalu', 'hotel',
'gak', 'rp', 'sini', 'sendiri', 'More', 'krang', 'afdol', 'entah', 'dgn', 'apa', 'karna', 'dongapalagi', 'jum', 'disini']
```

```
In [16]: def stopword_text (tokens) :
cleaned_tokens = []
for token in tokens:
if token not in stopwords:
cleaned_tokens.append(token)
return cleaned_tokens
```

Stemming

```
In [18]: from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
stem_factory = StemmerFactory()
stemmer = stem_factory.create_stemmer()
```

```
In [19]: def stemming_text (tokens):
hasil = [stemmer.stem(token) for token in tokens]
return hasil
```

```
In [20]: df['stemmed'] = df['stopword'].apply(stemming_text)
df.head(60)
```

A-2 Lampiran Script Python Analisis Data

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [2]: data = pd.read_csv('DataSkripsi_BERSIH.csv')
data
```

```
Out[2]:
```

	Ulasan	Ulasan_clean	token	stopword	stemmed
0	Gudegnya enak dan harga terjangkau. Tempat cuk...	gudegnya enak dan harga terjangkau tempat cuku...	['gudegnya', 'enak', 'dan', 'harga', 'terjangkau']	['gudegnya', 'enak', 'harga', 'terjangkau', 't...']	['gudeg', 'enak', 'harga', 'jangkau', 'tempat...']
1	Rasa bumbu gudegnya istimewa, apalagi telur pi...	rasa bumbu gudegnya istimewa apalagi telur pin...	['rasa', 'bumbu', 'gudegnya', 'istimewa', 'apa...']	['rasa', 'bumbu', 'gudegnya', 'istimewa', 'tel...']	['rasa', 'bumbu', 'gudeg', 'istimewa', 'telur...']
2	Gudeg yang mainstream sebenarnya selain karena...	gudeg yang mainstream sebenarnya selain karena...	['gudeg', 'yang', 'mainstream', 'sebenarnya', '...']	['gudeg', 'mainstream', 'sebenarnya', 'letakny...']	['gudeg', 'mainstream', 'benar', 'letak', 'pin...']
3	gudeg nya enak, harga nya affordable, lokasi n...	gudeg nya enak harga nya affordable lokasi nya...	['gudeg', 'nya', 'enak', 'harga', 'nya', 'affo...']	['gudeg', 'enak', 'harga', 'affordable', 'loka...']	['gudeg', 'enak', 'harga', 'affordable', 'loka...']
4	Gudegnya enak bgt, khas jogja bgt, aslik buat ka...	gudegnya enak bgt khas jogja bgt aslik buat ka...	['gudegnya', 'enak', 'bgt', 'khas', 'jogja', '...']	['gudegnya', 'enak', 'khas', 'aslik', 'kalian', '...']	['gudeg', 'enak', 'khas', 'aslik', 'kalian', '...']
...
581	Nama ini selalu mengusik telinga, penasaran de...	nama ini selalu mengusik telinga penasaran den...	['nama', 'ini', 'selalu', 'mengusik', 'telinga', '...']	['selalu', 'mengusik', 'telinga', 'penasaran', '...']	['selalu', 'usik', 'telinga', 'penasaran', 'ra...']
582	rasa gudegnya yang manis sangat pas dengan ras...	rasa gudegnya yang manis sangat pas dengan ras...	['rasa', 'gudegnya', 'yang', 'manis', 'sangat', '...']	['rasa', 'gudegnya', 'manis', 'pas', 'rasa', '...']	['rasa', 'gudeg', 'manis', 'pas', 'rasa', 'gud...']
583	Kami makan gudeg Yu Djum yang dekat Whiz Hotel...	kami makan gudeg yu djum yang dekat whiz hotel...	['kami', 'makan', 'gudeg', 'yu', 'djum', 'yang', '...']	['makan', 'gudeg', 'dekat', 'whiz', 'kecewa', '...']	['makan', 'gudeg', 'dekat', 'whiz', 'kecewa', '...']
584	Jogja identik dengan kota gudeg. Belum ke Jogj...	jogja identik dengan kota gudeg belum ke jogja...	['jogja', 'identik', 'dengan', 'kota', 'gudeg', '...']	['identik', 'gudeg', 'makan', 'gudeg', 'baik', '...']	['identik', 'gudeg', 'makan', 'gudeg', 'baik', '...']
585	salah satu gudeg tua yang masih bertahan di Jo...	salah satu gudeg tua yang masih bertahan di jo...	['salah', 'satu', 'gudeg', 'tua', 'yang', 'mas...']	['gudeg', 'tua', 'bertahan', 'menjaga', 'cita', '...']	['gudeg', 'tua', 'tahan', 'jaga', 'cita', 'ras...']

Feature Extraction

```
In [4]: from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer()
datavector = vectorizer.fit_transform(data['stemmed'])
```

```
In [5]: datavector.shape
```

K-Means Clustering (k = using elbow method)

```
In [6]: from sklearn.cluster import KMeans

wcss = []

for i in range(1, 11):
    model = KMeans(n_clusters=i, max_iter=100, init='k-means++', random_state = 0)
    model.fit(datavector)
    wcss.append(model.inertia_)
plt.plot(range(1, 11), wcss, 'bx-')
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
print(i, wcss)
```

```
In [7]: from sklearn.cluster import KMeans
model = KMeans(n_clusters=3, init='k-means++', random_state = 0)
model.fit(datavector)
hasil = model.labels_
```

```
In [26]: model.cluster_centers_
```

```
In [8]: histogram = pd.value_counts(hasil)
histogram.plot(kind='bar')
plt.xlabel('Jumlah Cluster')
plt.ylabel('Jumlah Ulasan')
```

```
In [9]: data['cluster'] = hasil
```

```
In [10]: Cluster_0 = data[data.Cluster == 0]
Cluster_1 = data[data.Cluster == 1]
Cluster_2 = data[data.Cluster == 2]
```

Visualization WordCloud

```
In [19]: order_centroids = model.cluster_centers_.argsort()[:, :-1]
terms = vectorizer.get_feature_names()

print("Top terms per cluster:")
for i in range(0, 3):
    print("Cluster %d:" %i)
    for j in order_centroids[i, :10]:
        print(' %s' % terms[j])
```

Association Rule

```
In [1]: import pandas as pd
import numpy as np
from mlxtend.frequent_patterns import apriori
from mlxtend.preprocessing import TransactionEncoder
import ast
from mlxtend.frequent_patterns import fpgrowth, association_rules
```

```
In [2]: df = pd.read_csv('k1aster0.csv')
df
```

```
In [5]: df['stemmed'] = df['stemmed'].apply(ast.literal_eval)
```

```
In [6]: list = []  
list = df['stemmed'].values.tolist()  
list[:5]
```

```
In [7]: te = TransactionEncoder()  
te_data = te.fit_transform(list)  
df_encode = pd.DataFrame(te_data, columns=te.columns_)  
df_encode.head()
```

```
In [8]: frequent_itemsets = fpgrowth(df_encode,  
                                     min_support=0.02,  
                                     use_colnames=True)  
ar_itemset = association_rules(frequent_itemsets,  
                               metric='confidence',  
                               min_threshold=0.1).sort_values(by='confidence',  
                                                             ascending=False)
```

```
In [11]: ar_itemset.sort_values(by='confidence', ascending=False)
```

