

***ASPECT BASED SENTIMENT ANALYSIS TERHADAP ULASAN RUMAH MAKAN
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
BORDERLINE SMOTE-TOMEK LINKS***

SKRIPSI

ALVINO VIANDO PUTRA

181402121



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA**

MEDAN

2023

***ASPECT BASED SENTIMENT ANALYSIS TERHADAP ULASAN RUMAH MAKAN
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
BORDERLINE SMOTE-TOMEK LINKS***

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah Sarjana
Teknologi Informasi

ALVINO VIANDO PUTRA

181402121



PROGRAM STUDI S1 TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS SUMATERA UTARA

MEDAN

2023

PERSETUJUAN

Judul : *ASPECT BASED SENTIMENT ANALYSIS*
TERHADAPULASAN RUMAH MAKAN
MENGUNAKAN *SUPPORT VECTOR MACHINE*
DAN BORDERLINE SMOTE-TOMFK LINKS.

Kategori : SKRIPSI

Nama : ALVINO VIANDO PUTRA

Nomor induk mahasiswa : 181420101

Program Studi : TEKNOLOGI INFORMASI

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

Medan, 10 Januari 2019

Komisi Pembimbing

Pembimbing 2

Pembimbing 1




Rusy Nurhasanah, S.Kom., M.Kom.
NIP. 198707012019032016



Irena Jaya, S.T., M.Kom.
NIP. 198407072015041001

Diketahui/disetujui oleh
Program Studi S1 Teknologi Informasi

Ketua,


Dedy Arisandi S.T., M.Kom.
NIP. 197908312009121002

PERNYATAAN

ASPECT BASED SENTIMENT ANALYSIS TERHADAP ULASAN RUMAH
MAKAN MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN
BORDERLINE SMOTE-TOMEK LINKS

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 10 Januari 2024

Alvino Viando Putra

181402121

UCAPAN TERIMA KASIH

Puji syukur kehadiran Allah *Subhanahu wa Ta'ala* yang telah memberikan rahmat, nikmat, serta karunia yang sangat besar kepada penulis sehingga akhirnya dapat menyelesaikan tugas akhir ini sebagai salah satu syarat untuk lulus dan mendapatkan gelar Sarjana Komputer dari Program Studi S1 Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.

Tugas akhir ini penulis persembahkan untuk kedua orang tua penulis yang telah mendukung penulis sepenuhnya dalam masa perkuliahan ini. Penulisan tugas akhir ini tentunya tidak luput dari dukungan, bantuan, serta doa tanpa henti dari banyak pihak. Berbagai macam rintangan dan kesulitan yang penulis hadapi dan temukan dalam proses penulisan tugas akhir ini. Tetapi, semuanya dapat terlewati berkat dukungan berbagai pihak yang senantiasa berada di sisi penulis. Maka pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Muryanto Amin, S.Sos., M.Si. selaku Rektor Universitas Sumatera Utara.
2. Ibu Dr. Maya Silvi Lidya, M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
3. Bapak Dedy Arisandi, ST., M.Kom. selaku Pelaksana Tugas Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara
4. Bapak Ivan Jaya S.Si., M.Kom.selaku Dosen Pembimbing I dan Ibu Rossy Nurhasanah, S.Kom., M.Kom.. selaku Dosen Pembimbing II yang selalu membantu, meluangkan waktu, pemikiran, serta dukungan.
5. Bapak Indra Aulia S.TI., M. Kom. selaku Dosen Pembimbing yang juga telah membantu, meluangkan waktu, pemikiran, serta dukungan dari awal masa penulisan tugas akhir ini sampai dengan masa kepindahan.
6. Bapak dan Ibu Dosen Program Studi S1 Teknologi Informasi Universitas Sumatera Utara yang telah banyak berkontribusi dalam proses pembelajaran saya menjadi Sarjana Komputer.

7. Keluarga penulis, Papa Firman S.H, Mama Viandini Rachmansjah serta Adik penulis Aldo Celviano Ananda dan Olivia Febiola yang senantiasa mendoakan, memberikan dukungan, dan motivasi untuk bisa fokus dan tetap semangat dalam menyelesaikan tugas akhir ini.
8. Sahabat dekat penulis, M. Ali Akbar Sinaga S.Kom yang menjadi sahabat seperjuangan sedari masa SMA, sampai masa perkuliahan yang memberikan banyak sekali dukungan dan bantuan dari awal masa sekolah sampai proses penyelesaian tugas akhir ini.
9. Teman-teman seperjuangan penulis, Muhammad Ridwan Syah Putra Lubis S.Kom. , Suci Khairiah S.Kom. , Handika Riyandar S.Kom. , Indah S.Kom., Indriyani Sembiring S.Kom., Rini Haryati, dan teman-teman “Sukses Bersama” lainnya yang menemani penulis pada saat masa perkuliahan.
10. Sahabat dekat penulis sedari dulu, Alvin Arrazy S.T., Izza Bagaskara Hardi A.Md., dan Tengku Muhammad Fajri S.Ds. yang menemani penulis dikala kejenuhan, dan selalu memberi dukungan.
11. Seluruh teman-teman Angkatan 2018 Teknologi Informasi yang telah menemani masa-masa perkuliahan penulis.
12. Semua pihak yang berperan langsung maupun tidak langsung yang tidak dapat ditulis satu persatu yang telah membantu penulis dalam penyelesaian skripsi ini.
13. Kepada penulis sendiri, karena telah berjuang sampai saat sini. Karena telah sabar dari segala hal yang mengejar, tetap semangat walau kadangkala merasa kalah dan gundah.

Penulis menyadari bahwa masih terdapat banyak kelemahan dalam penyusunan tugas akhir ini. Kritik dan saran yang membangun sangat diharapkan guna memperbaiki dan menyempurnakan penelitian di masa yang akan datang.

Medan, 10 Januari 2024

Alvino Viando Putra

ABSTRAK

Pertumbuhan pesat industri kuliner di Indonesia mendorong munculnya berbagai restoran baru dengan ragam kuliner menarik. Meskipun Google Maps sudah menyediakan informasi terkait kualitas rumah makan, namun belum mendetail pada aspek-aspek fundamental seperti kualitas makanan, minuman, pelayanan, tempat dan harga. Platform lain yang menyediakan informasi lebih rinci memiliki keterbatasan pada jumlah restoran dan pengguna. Untuk mengatasi hal ini, penelitian ini mengusulkan penggunaan Analisis Sentimen Berbasis Aspek untuk memprediksi sentimen ulasan rumah makan secara real-time di Google Maps.

Dalam penelitian ini, 1005 data ulasan digunakan dalam proses training dengan Algoritma *Support Vector Machine* dengan K-Fold Cross Validation 10 Fold. Digunakan proses Resampling guna menyeimbangkan jumlah data. Resampling dilakukan dalam lima skenario, termasuk tanpa resampling, *Borderline* SMOTE, Tomek Links, *Borderline* SMOTE-Tomek Links, dan Tomek Links-*Borderline* SMOTE. Hasil training menunjukkan bahwa model *Support Vector Machine* dengan skenario Tomek Links-*Borderline* SMOTE memiliki performa terbaik, dengan rata-rata akurasi 87,4% dan rata-rata f1-score 85,6%. Penelitian ini dapat memberikan informasi lebih lengkap dan akurat terkait kualitas rumah makan, membantu pengguna membuat keputusan yang lebih baik saat memilih tempat makan. Dengan demikian, solusi ini dapat meningkatkan pengalaman makan pelanggan dan memberikan informasi yang lebih komprehensif dalam industri kuliner yang berkembang pesat di Indonesia.

Kata Kunci: Rumah Makan, *Support Vector Machine*, Analisis Sentimen Berbasis Aspek, *Resampling* .

ASPECT-BASED SENTIMENT ANALYSIS OF REVIEWS ON RESTAURANTS USING SUPPORT VECTOR MACHINE AND BORDERLINE SMOTE-TOMEK LINKS

ABSTRACT

The rapid growth of the culinary industry in Indonesia has led to the emergence of various new restaurants offering a diverse range of cuisines. Although Google Maps provides information about the quality of restaurants, it lacks detailed aspects such as food quality, beverages, service, ambiance, and pricing. Other platforms offering more detailed information face limitations in terms of the number of registered restaurants and users. To address this, this study proposes the use of Aspect-Based Sentiment Analysis to predict real-time restaurant review sentiments on Google Maps.

In this research, 1005 review data were used for training with the Support Vector Machine algorithm using 10-fold Cross-Validation. The process of resampling is used to balance the number of data. Resampling was conducted in five scenarios, including without resampling, Borderline SMOTE, Tomek Links, Borderline SMOTE-Tomek Links, and Tomek Links-Borderline SMOTE. The training results indicate that the Support Vector Machine model with the Tomek Links-Borderline SMOTE scenario performs the best, achieving an average accuracy of 87.4% and an average F1-score of 85.6%. This study can provide more comprehensive and accurate information regarding the quality of restaurants, assisting users in making better decisions when choosing dining venues. Thus, this solution has the potential to enhance the dining experience for customers and provide more comprehensive information in the rapidly growing culinary industry in Indonesia.

Keywords: Restaurants, Support Vector Machine, Aspect-Based Sentiment Analysis, Resampling.

DAFTAR ISI

PERSETUJUAN	Error! Bookmark not defined.
PERNYATAAN	ii
UCAPAN TERIMA KASIH.....	iii
ABSTRAK.....	v
ASPECT-BASED SENTIMENT ANALYSIS OF REVIEWS ON RESTAURANTS USING SUPPORT VECTOR MACHINE AND BORDERLINE SMOTE-TOMEK LINKS	vi
ABSTRACT.....	vi
Daftar Isi	vii
Daftar Tabel	x
Daftar Gambar.....	xii
BAB 1 Pendahuluan.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan.....	6
BAB 2 Landasan Teori	8
2.1 Natural Language Processing.....	8
2.2 Sentiment Analysis.....	8
2.3 Aspect Based Sentiment Analysis.....	8

2.4	Text Processing	9
2.5	Support Vector Machine	9
2.6	Scraping.....	12
2.7	Bag of Words (BOW)	13
2.8	K-Fold Cross Validation	14
2.9	Ketidakseimbangan Data.....	15
2.10	Resampling	16
2.10.1	Tomek Links (T-Links).....	17
2.10.2	Borderline SMOTE	18
2.11	Restoran	22
2.12	Penelitian Terdahulu	22
2.13	Perbedaan Penelitian.....	29
BAB 3	Analisis dan perancangan sistem	30
3.1	Dataset	30
3.2	Arsitektur Umum.....	31
3.2.1	Exploratory Data Analysis	32
3.2.2	Pre-processing.....	34
3.2.3	Feature Extraction	37
3.2.4	Resampling	40
3.2.5	Classification.....	48
3.2.6	Validation.....	52
3.2.7	Model	52
3.2.8	Scraping	55
3.2.9	Output	56
3.3	Perancangan Sistem.....	57
3.4	Metode Evaluasi	64
BAB 4	Implementasi dan Pengujian Sistem.....	66

4.1	Implementasi Sistem	66
4.1.1	Spesifikasi Perangkat Keras dan Perangkat Lunak.....	66
4.1.2	Implementasi Perancangan Antarmuka Sistem	66
4.2	Implementasi Model.....	73
4.2.1	Pelatihan Model	73
4.2.2	Pengujian Model	81
4.2.3	Hasil Pengujian Model.....	84
4.2.4	Evaluasi Model	93
BAB 5	Kesimpulan dan saran.....	107
5.1	Kesimpulan.....	107
5.2	Saran.....	108
	Daftar Pustaka.....	109

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	26
Tabel 3. 1 <i>Input</i> dan <i>output</i> proses <i>Punctual Removal</i>	34
Tabel 3. 2 <i>Input</i> dan <i>output</i> proses <i>Casefolding</i>	35
Tabel 3. 3 <i>Output</i> proses <i>Tokenization</i>	35
Tabel 3. 4 <i>Input</i> dan <i>output</i> proses <i>Stopword Removal</i>	36
Tabel 3. 5 <i>Input</i> dan <i>output</i> proses <i>Stemming</i>	36
Tabel 3. 6 <i>Input</i> dan <i>output</i> proses <i>Data Normalization</i>	36
Tabel 3. 7 Gambaran variabel <i>Dictionary</i>	37
Tabel 3. 8 Gambaran variabel <i>Dictionary</i> dan Vektor (setelah dihilangkan <i>stop words</i>)	38
Tabel 3. 9 Gambaran variabel <i>Vektor</i> (setelah dihilangkan <i>stop words</i>)	38
Tabel 3. 10 Contoh Dataset sebelum diubah menjadi bentuk Vektor.....	38
Tabel 3. 11 Contoh Dataset setelah diubah menjadi bentuk Vektor	39
Tabel 3. 12 Tabel koordinat studi kasus	49
Tabel 3. 13 Tabel <i>Confusion Matrix</i>	64
Tabel 4. 1 Hasil Pelatihan Sistem dengan beberapa skenario.....	74
Tabel 4. 2 Tabel Ulasan	82
Tabel 4. 3 Tabel Prediksi ulasan	85
Tabel 4. 4 Keterangan Sentimen Positif Aspek Makanan Confussion Matrix	97
Tabel 4. 5 Keterangan Sentimen Netral Aspek Makanan Confussion Matrix.....	97
Tabel 4. 6 Keterangan Sentimen Negatif Aspek Makanan Confussion Matrix.....	97
Tabel 4. 7 Keterangan Sentimen Positif Aspek Minuman Confussion Matrix.....	98
Tabel 4. 8 Keterangan Sentimen Netral Aspek Minuman Confussion Matrix	98
Tabel 4. 9 Keterangan Sentimen Negatif Aspek Minuman Confussion Matrix	98
Tabel 4. 10 Keterangan Sentimen Positif Aspek Pelayanan Confussion Matrix.....	98
Tabel 4. 11 Keterangan Sentimen Netral Aspek Pelayanan Confussion Matrix	99
Tabel 4. 12 Keterangan Sentimen Netral Aspek Pelayanan Confussion Matrix	99
Tabel 4. 13 Keterangan Sentimen Positif Aspek Tempat Confussion Matrix	99
Tabel 4. 14 Keterangan Sentimen Netral Aspek Tempat Confussion Matrix	100
Tabel 4. 15 Keterangan Sentimen Negatif Aspek Tempat Confussion Matrix	100
Tabel 4. 16 Keterangan Sentimen Positif Aspek Harga Confussion Matrix	100
Tabel 4. 17 Keterangan Sentimen Netral Aspek Harga Confussion Matrix	100

Tabel 4. 18 Keterangan Sentimen Negatif Aspek Harga Confussion Matrix.....	101
--	-----

DAFTAR GAMBAR

Gambar 2. 1 Contoh Data Ulasan Rumah Makan.....	9
Gambar 2. 2 <i>Hyperplane</i> kelas -1 dan kelas +1 (Cahyono et al., 2020)	10
Gambar 2. 3 Parameter Gamma	12
Gambar 2. 4 Ilustrasi Cross Validation dengan 5 Fold	15
Gambar 2. 5 Ilustrasi Proses <i>Undersampling</i> Tomek Link.....	17
Gambar 2. 6 <i>Flowchart</i> Tomek Links.....	18
Gambar 2. 7 Ilustrasi Pembagian Himpunan m-tetangga terdekat (Onur Durahim, 2016)	20
Gambar 2. 8 Alur Kerja Borderline SMOTE.....	21
Gambar 3. 1 Gambar Arsitektur Umum	32
Gambar 3. 2 Jumlah Dataset Aspek Makanan	32
Gambar 3. 3 Jumlah Dataset Aspek Minuman	33
Gambar 3. 4 Jumlah Dataset Aspek Pelayanan.....	33
Gambar 3. 5 Jumlah Dataset Aspek Tempat.....	33
Gambar 3. 6 Jumlah Dataset Aspek Harga	34
Gambar 3. 7 Jumlah data sebelum dan sesudah proses Tomek Links dan <i>Borderline</i> SMOTE pada aspek Makanan	40
Gambar 3. 8 Persebaran Jumlah Data Aspek Makanan sebelum <i>resampling</i> dengan visualisasi t-SNE.....	41
Gambar 3. 9 Persebaran Jumlah Data Aspek Makanan sesudah <i>resampling</i> dengan visualisasi t-SNE.....	41
Gambar 3. 10 Jumlah data sebelum dan sesudah proses Tomek Links dan <i>Borderline</i> SMOTE pada aspek Minuman.....	42
Gambar 3. 11 Persebaran Jumlah Data Aspek Minuman sebelum <i>resampling</i> dengan visualisasi t-SNE.....	42
Gambar 3. 12 Persebaran Jumlah Data Aspek Minuman sesudah <i>resampling</i> dengan visualisasi t-SNE.....	43
Gambar 3. 13 Jumlah data sebelum dan sesudah proses Tomek Links dan <i>Borderline</i> SMOTE pada aspek Pelayanan	43

Gambar 3. 14 Persebaran Jumlah Data Aspek Pelayanan sebelum <i>resampling</i> dengan visualisasi t-SNE.....	44
Gambar 3. 15 Persebaran Jumlah Data Aspek Pelayanan sesudah <i>resampling</i> dengan visualisasi t-SNE.....	44
Gambar 3. 16 Jumlah data sebelum dan sesudah proses Tomek Links dan <i>Borderline</i> SMOTE pada aspek Tempat	45
Gambar 3. 17 Persebaran Jumlah Data Aspek Tempat sebelum <i>resampling</i> dengan visualisasi t-SNE.....	45
Gambar 3. 18 Persebaran Jumlah Data Aspek Tempat sesudah <i>resampling</i> dengan visualisasi t-SNE.....	46
Gambar 3. 19 Jumlah data sebelum dan sesudah proses Tomek Links dan <i>Borderline</i> SMOTE pada aspek Harga.....	46
Gambar 3. 20 Persebaran Jumlah Data Aspek Harga sebelum <i>resampling</i> dengan visualisasi t-SNE.....	47
Gambar 3. 21 Persebaran Jumlah Data Aspek Harga sesudah <i>resampling</i> dengan visualisasi t-SNE.....	47
Gambar 3. 22 Diagram Koordinat Studi Kasus	49
Gambar 3. 23 <i>Flowchart</i> proses <i>Scraping</i>	55
Gambar 3. 24 Rancangan halaman Beranda	58
Gambar 3. 25 Rancangan halaman Login Admin.....	59
Gambar 3. 26 Rancangan halaman Training Model.	60
Gambar 3. 27 Rancangan halaman lanjutan Training Model.	61
Gambar 3. 28 Rancangan halaman Home bagian Sentiment.....	62
Gambar 3. 29 Rancangan halaman Home bagian Sentiment.....	63
Gambar 4. 1 Implementasi Tampilan Beranda	67
Gambar 4. 2 Lanjutan Implementasi Tampilan Beranda	68
Gambar 4. 3 Implementasi Tampilan Login Admin	69
Gambar 4. 4 Implementasi Tampilan Training Data	69
Gambar 4. 5 Lanjutan Implementasi Tampilan Training Data	70
Gambar 4. 6 Lanjutan Implementasi Tampilan Training Data	70
Gambar 4. 7 Lanjutan Implementasi Tampilan Training Data	71
Gambar 4. 8 Implementasi Tampilan User Input.....	72
Gambar 4. 9 Output tabel pada Implementasi Tampilan User Input	72

Gambar 4. 10 Output plot perbandingan jumlah sentimen pada Implementasi Tampilan User Input.....	73
Gambar 4. 11 Gambar Tabel F1-Score	76
Gambar 4. 12 Diagram kurva pelatihan model aspek makanan	77
Gambar 4. 13 Diagram kurva pelatihan model aspek minuman.....	78
Gambar 4. 14 Diagram kurva pelatihan model aspek pelayanan.....	79
Gambar 4. 15 Diagram kurva pelatihan model aspek tempat	80
Gambar 4. 16 Diagram kurva pelatihan model aspek harga	81
Gambar 4. 17 Tampilan <i>Testing</i>	88
Gambar 4. 18 Bagan <i>Pie Chart</i> Tampilan <i>Testing</i>	88
Gambar 4. 19 <i>Pie Chart</i> Aspek Makanan	89
Gambar 4. 20 <i>Pie Chart</i> Aspek Minuman	90
Gambar 4. 21 <i>Pie Chart</i> Aspek Pelayanan	91
Gambar 4. 22 <i>Pie Chart</i> Aspek Tempat.....	92
Gambar 4. 23 <i>Pie Chart</i> Aspek Harga	93
Gambar 4. 24 <i>Confusion Matrix</i> model aspek makanan.....	94
Gambar 4. 25 <i>Confusion Matrix</i> model aspek minuman	95
Gambar 4. 26 <i>Confusion Matrix</i> model aspek pelayanan	95
Gambar 4. 27 <i>Confusion Matrix</i> model aspek tempat	96
Gambar 4. 28 <i>Confusion Matrix</i> model aspek harga	96

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Industri kuliner saat ini sedang mengalami perkembangan pesat. Menurut data yang dikeluarkan Badan Ekonomi Kreatif (Bekraf) Republik Indonesia tahun 2016, subsektor kuliner menyumbang 41,4% dari total kontribusi perekonomian kreatif sebesar Rp922 triliun pada tahun tersebut. Angka ini merupakan yang tertinggi dibandingkan dengan 16 subsektor lain di Bekraf RI. Bahkan di tengah kontraksi pertumbuhan ekonomi saat ini, industri makanan dan minuman berhasil tetap tumbuh sebesar 0,22% dibanding tahun sebelumnya. Kementerian Perindustrian juga mencatat bahwa kinerja industri makanan dan minuman selama periode 2015-2019 rata-rata tumbuh 8,16%, melebihi rata-rata pertumbuhan industri pengolahan nonmigas yang sebesar 4,69%. Meskipun terjadi kontraksi pertumbuhan industri nonmigas sebesar 2,52% selama triwulan IV tahun 2020 akibat dampak pandemi, industri makanan dan minuman masih berhasil tumbuh positif sebesar 1,58% pada tahun 2020.

Kemajuan Teknologi Informasi yang berkembang sangat pesat pada saat ini juga sangat memberi dampak positif untuk berbagai aspek kehidupan manusia. Sektor bisnis di bidang kuliner atau restoran menjadi salah satu yang mengalami manfaat positif dari perkembangan tersebut. Internet berfungsi sebagai sarana atau *platform* yang digunakan dalam mendapatkan suatu informasi, menjadi bagi individu untuk mengekspresikan informasi berbentuk opini atau ulasan terkait sebuah restoran secara bebas (Nurifan et al.2019). Ulasan yang diberikan oleh individu di internet terhadap suatu restoran mengandung kelebihan serta

kekurangan dari suatu restoran dapat membantu calon pelanggan lain dalam mengevaluasi kualitas usaha atau layanan yang diberikan sehingga mereka dapat membuat keputusan dalam mengunjungi restoran (Govindarajan, 2014).

Dengan perkembangan teknologi saat ini, menyebabkan adanya kemudahan pelaku usaha di bidang kuliner dalam mempromosikan atau menyebarluaskan informasi umum terkait kuliner yang mereka dagangkan. Tetapi, terdapat satu permasalahan dimana konsumen mengalami kesulitan mendapatkan informasi terkait suatu restoran secara detail, sampai ke aspek-aspek fundamentalnya seperti kualitas makanan, minuman, bangunan fisik restoran, layanan, dan harga, sehingga mempengaruhi pada saat pengambilan keputusan pada saat akan sebuah restoran. Saat ini, memang sudah ada website yang menyediakan ulasan mengenai kualitas restoran, Contohnya ulasan pada Google Maps, namun ulasan yang ditampilkan hanya bersifat ulasan secara keseluruhan, apakah positif atau negatif, belum secara spesifik ulasan itu ditujukan kepada aspek fundamentalnya. Hal ini dikarenakan setiap orang lebih banyak memberikan penilaian subjektif kepada konteks yang umum. Kondisi ini dapat menyebabkan kebingungan bagi setiap orang untuk mendapatkan rekomendasi dari sudut pandang konteks fundamentalnya.

Sistem rekomendasi dapat membantu konsumen dalam menentukan restoran yang akan dikunjungi. Sistem rekomendasi yang mampu mengekstraksi konteks fundamental dari sebuah ulasan sangat dibutuhkan untuk membantu konsumen dalam mendapatkan informasi tentang restoran dengan lebih spesifik. Sistem ini diharapkan tidak hanya memberikan penilaian apakah suatu restoran memiliki kualitas yang baik ataupun buruk, tetapi juga memberikan penilaian berdasarkan aspek fundamentalnya.

Metode *Aspect Based Sentiment Analysis* atau ABSA dapat digunakan dalam kasus seperti ini, dimana dilakukan Analisis Sentimen berbasis aspek-aspek yang sudah ditentukan terhadap suatu kalimat. Berbagai Studi terkait ABSA telah

dilakukan dalam memperoleh sentimen terkait topik restoran. Seperti penelitian oleh Amalia, P. R., & Winarko, E. pada tahun 2021 yang melakukan analisis sentiment berbasis aspek terhadap ulasan restoran di Indonesia dengan menggunakan kombinasi *Convolutional Neural Network* dan *Contextual Word Embedding*. Hasil pengujian terbaik didapatkan menggunakan model *ELMo-CNN*, dengan nilai *micro-average f1-score* sebesar 0.86. Untuk klasifikasi sentimen terbaik dihasilkan model *BERT-CNN* dengan nilai *f1-score* sebesar 0.91. Penelitian lainnya dilakukan oleh Yustihan, S. et al. pada tahun 2021 yang melakukan analisis sentiment berbasis aspek terhadap ulasan restoran menggunakan metode *Support Vector Machine* atau SVM. Aspek yang dianalisis antara lain *Ambience*, *Drink*, *Food*, *Restaurant*, serta *Service*. Hasil pengujian metode *Support Vector Machine* dalam mengklasifikasikan aspek dengan menggunakan pendekatan *macro averaging* menghasilkan *f-measure* sebesar 0,73. Sedangkan dalam mengklasifikasikan polaritas sentimen menghasilkan *f-measure* sebesar 0,92. Selain itu, penelitian yang dilakukan oleh Hairani et al., pada tahun 2023 mengemukakan bahwa penggunaan SMOTE dan Tomek Links pada data penelitian terkait Klasifikasi Penyakit Diabetes menggunakan algoritma *Random Forest* yang bersifat tidak seimbang, dapat diatasi. Penggunaan algoritma *Random Forest* dikombinasikan dengan SMOTE dan Tomek Links menghasilkan akurasi sebesar 86,4%, sensitivitas sebesar 88,2%, presisi sebesar 82,3%, dan F1-score sebesar 85,1%. Penggunaan Smote-TomekLink dapat meningkatkan kinerja metode random forest berdasarkan akurasi, sensitivitas, presisi, dan F1-score.

Berdasarkan kondisi saat ini dan penelitian terdahulu yang relevan dengan penelitian ini, penulis melakukan penelitian dengan judul “***Aspect Based Sentiment Analysis Terhadap Ulasan Rumah Makan menggunakan Support Vector Machine dan Borderline SMOTE-Tomek Links***”.

1.2 Rumusan Masalah

Pertumbuhan industri kuliner di Indonesia yang semakin cepat saat ini menyebabkan semakin beragamnya jenis serta kualitas makanan, minuman serta pelayan yang diberikan oleh masing-masing restoran. Kondisi ini membuat

pengguna mengalami kesulitan mendapatkan informasi mengenai suatu restoran secara detail, sampai ke aspek-aspek fundamentalnya seperti kualitas makanan, harga, dan lainnya, yang mempengaruhi pada saat pengambilan keputusan dalam mengunjungi sebuah restoran. Namun pada sistem yang ada pada saat ini hanya menampilkan informasi berupa rating secara keseluruhan tanpa adanya informasi apakah restoran tersebut mendapat tanggapan positif ataupun negatif dari pengguna. Oleh karena itu, analisis ulasan dari pelanggan berdasarkan aspek yang telah ditentukan perlu dilakukan guna mendapatkan informasi terkait restoran yang selanjutnya digunakan sebagai tolak ukur dalam memilih restoran sesuai dengan preferensi yang pengguna inginkan.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk mengukur kinerja algoritma *Support Vector Machine* dengan *feature extraction Bag of words* dan teknik *resampling Borderline SMOTE* dan *Tomek Links* dalam memperoleh sentimen pelanggan rumah makan berdasarkan ulasan rumah makan yang diberikan berdasarkan aspek yang telah ditentukan.

1.4 Batasan Masalah

Adapun Batasan masalah dalam penelitian ini sebagai berikut :

1. Data yang digunakan berupa data teks yang diperoleh melalui *website* Google Maps pada bagian *review* atau ulasan.
2. Data yang digunakan pada penelitian ini menggunakan Bahasa Indonesia.
3. Kelas/Aspek yang digunakan untuk klasifikasi aspek antara lain adalah makanan, minuman, pelayanan, tempat atau bangunan, dan harga.
4. Kata Kunci yang diinputkan pada tahap uji coba hanya berdasarkan 1 nama rumah makan.
5. Nama rumah makan yang digunakan sebagai *input* uji coba harus sesuai dan spesifik dengan yang terdapat pada *website* Google Maps

1.5 Manfaat Penelitian

Manfaat penelitian ini antara lain:

1. Dengan adanya penelitian ini, diharapkan dapat mempermudah pengguna dalam hal ini calon pelanggan dalam menentukan restoran yang ingin dikunjungi sesuai dengan preferensinya masing-masing, berdasarkan sentiment dari aspek-aspek yang ditampilkan.
2. Mengevaluasi kemampuan Algoritma Support Vector Machine dengan pendekatan *feature extraction Bag of Words* dan teknik *resampling* Tomek Links + SMOTE pada bidang analisis sentiment berbasis aspek.

1.6 Metodologi Penelitian

Tahapan–tahapan yang akan dilakukan pada penelitian ini adalah :

1. Studi literatur

Pada tahap ini, dilakukan tahap pengumpulan informasi yang dibutuhkan dalam proses penelitian antara lain referensi jurnal, buku, maupun penelitian lain yang sudah dilakukan sebelumnya terkait *sentiment analysis*, *aspect-based sentiment analysis*, algoritma *Support Vector Machine*, metode *Feature Extraction Bag of words*, teknik *resampling Borderline SMOTE-Tomek Links* serta *library* yang akan digunakan.

2. Analisis permasalahan

Pada tahap ini, dilakukan analisis terhadap referensi yang sudah dikumpulkan sebelumnya untuk mendapatkan pemahaman tentang teknik *aspect-based sentiment analysis* menggunakan algoritma *Support Vector Machine* dengan *Bag of words* dan *Borderline SMOTE-Tomek Links* untuk mengatasi permasalahan dalam penelitian ini.

3. Pengumpulan Data

Dilakukan pengumpulan data berupa teks ulasan yang diperoleh dari platform Google Maps pada bagian ulasan restoran.

4. Perancangan Sistem

Pada tahap ini, penulis melakukan perancangan arsitektur umum, mengumpulkan data ulasan restoran, dan merancang tampilan antarmuka sistem.

5. Implementasi

Pada tahap ini, penulis menerapkan rancangan yang ada dan membangun sistem yang sesuai dengan tujuan penelitian.

6. Pengujian

Pada tahap ini, penulis akan melakukan proses uji performa untuk memastikan sistem yang dibangun berjalan dengan baik.

7. Penyusunan laporan

Pada tahap ini, penulis melakukan penyusunan laporan berdasarkan keseluruhan penelitian yang sudah dikerjakan.

1.7 Sistematika Penulisan

Untuk memberikan gambaran dan kerangka yang jelas dalam penulisan ini, maka penulis memerlukan sistematika penulisan masing masing bab yang dijelaskan sebagai berikut :

Bab 1 : Pendahuluan

Bagian ini menjelaskan tentang pendahulu yang mencakup latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metode penelitian, dan sistematika penulisan.

Bab 2 : Landasan Teori

Bagian ini menjelaskan landasan pengetahuan berupa teori-teori yang berhubungan dengan permasalahan pada penelitian ini. Teori yang akan dibahas mengenai *sentiment analysis*, *aspect-based sentiment analysis*, algoritma *Support Vector Machine*, metode *Bag of Word*, teknik *oversampling Borderline SMOTE* dan *Tomek Links* serta library yang akan digunakan.

Bab 3 : Analisis dan Perancangan

Bagian ini menjelaskan terkait analisis permasalahan penelitian dan perancangan metode *Support Vector Machine* dalam analisis sentimen dari ulasan makanan,

dari arsitektur umum, sampai tahapan dan proses perancangan pembuatan aplikasi penelitian.

Bab 4 : Implementasi dan Pengujian

Pada bagian ini, dijelaskan terkait implementasi rancangan aplikasi dan hasil dari pengujian dari aplikasi yang dibuat.

Bab 5 : Kesimpulan dan Saran

Dalam bagian ini menjelaskan tentang kesimpulan dan ringkasan dari penelitian yang dibuat. Saran dari penelitian yang telah dilakukan dan dapat digunakan untuk pengembangan penelitian berikutnya.

BAB 2

LANDASAN TEORI

2.1 Natural Language Processing

Natural Language Processing atau Pemrosesan Bahasa Alami merupakan satu dari beberapa cabang ilmu komputer pada bidang *artificial intelligence* (AI) atau kecerdasan buatan dan *computational linguistic*. *Natural Language Processing* berfokus pada interaksi antara Bahasa Alami dan komputer. *Natural Language Processing* memungkinkan sebuah interaksi yang halus, lancar, dan terpadu antara komputer dan manusia. *Natural Language Processing* memberikan kemampuan komputer untuk memahami perkataan manusia, baik itu dalam bentuk teks ataupun suara, dengan bantuan *machine learning* atau pemrograman mesin (Chopra et al, 2016).

2.2 Sentiment Analysis

Sentiment Analysis atau bisa disebut *opinion mining* adalah sebuah bidang studi yang menganalisa terkait pendapat orang lain, sentiment, evaluasi, penilaian, perilaku dan emosi terhadap suatu entitas antara lain, produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atribut-atributnya . *Sentiment Analysis* dan *opinion mining* berfokus dalam menilai opini dimana mengekspresikan atau mengimplikasi positif atau negative sebuah sentiment (Liu, 2020).

2.3 Aspect Based Sentiment Analysis

Aspect Based Sentiment Analysis merupakan pengembangan dari *Sentiment Analysis*, dimana dalam *Aspect Based Sentiment Analysis*, penilaian opini yang dilakukan tidak hanya secara general atau umum, melainkan ke beberapa aspek yang terkandung dalam suatu kalimat. Contoh, dalam kalimat “Rasa makanannya secara keseluruhan enak, tetapi untuk segi harga agak mahal”, dapat dinilai bahwa pada aspek makanan terkandung opini positif, tetapi pada aspek harga terkandung opini negatif.

2.4 Text Processing

Text Processing atau Pemrosesan Teks adalah serangkaian proses yang dilakukan terhadap data teks yang bertujuan agar data tersebut menjadi lebih terstruktur. Dengan data yang terstruktur, memungkinkan proses pengolahan data lebih optimal sehingga tujuan dari penelitian dapat tercapai dengan baik. Tahap-tahap pada *text processing* menghasilkan kata yang bersih, yang siap diolah. Tahap-tahapnya antara lain adalah *cleaning*, *data normalization*, *case folding*, *punctual removal*, *tokenization*, *stopword removal*, dan *stemming*.



Iwan Sulistiyo

Local Guide · 128 ulasan · 246 foto

★★★★★ 2 bulan lalu

Selagi dikota medan, sy sempatkan mencicipi kembali mie aceh yg terkenal di kota medan, sy memesan 1 porsi mie aceh tumis, rasanya lezat, bumbunya terasa, tapi mie kurang kenyal. Acar bawang dan empingnya dengan porsi seadanya. Tersedia banyak meja, ada musholah dan playground untuk anak2 bermain.



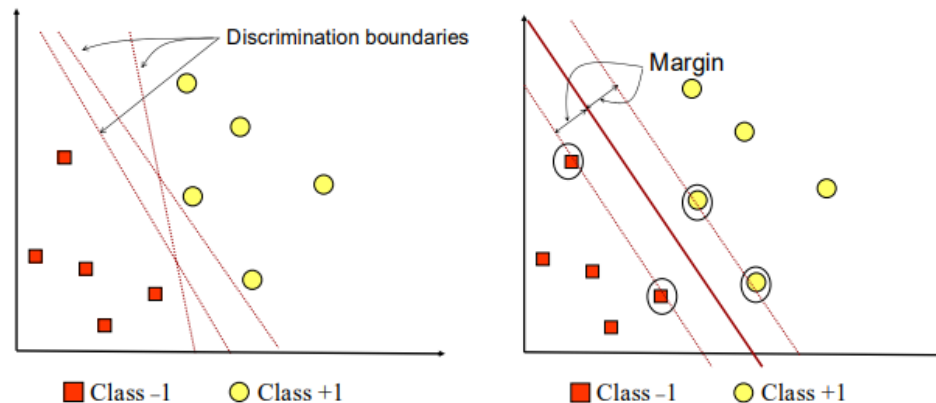
Gambar 2. 1 Contoh Data Ulasan Rumah Makan

Sumber : www.maps.google.com

2.5 Support Vector Machine

Support Vector Machine atau SVM merupakan sebuah teknik klasifikasi dalam pembelajaran mesin dimana menggunakan sebuah fungsi yang dapat disebut *kernel* untuk memetakan sebuah ruang berisikan *data points* dimana *data points* tersebut tidak secara linear dipisahkan kedalam ruang baru dimana ia berada (Mullen & Collier, 2004) . Pada tahun 1992, Boser, Guyon, dan Vapnik mengembangkan SVM. SVM merupakan kombinasi dari teori-teori komputasi yang sebelumnya telah dikembangkan beberapa tahun kebelakang (Nugroho et al., 2003). Umumnya, tujuan dari jaringan saraf

adalah untuk menemukan fungsi *hyperplane* yang memisahkan antar kelas. Sedangkan pada SVM, berusaha mendapatkan fungsi *hyperplane* terbaik yang terdapat pada *input space*. SVM memiliki prinsip dasar *linear classifier*, yang selanjutnya dikembangkan, sehingga dapat bekerja pada permasalahan *non-linear* (Nugroho et al., 2003).



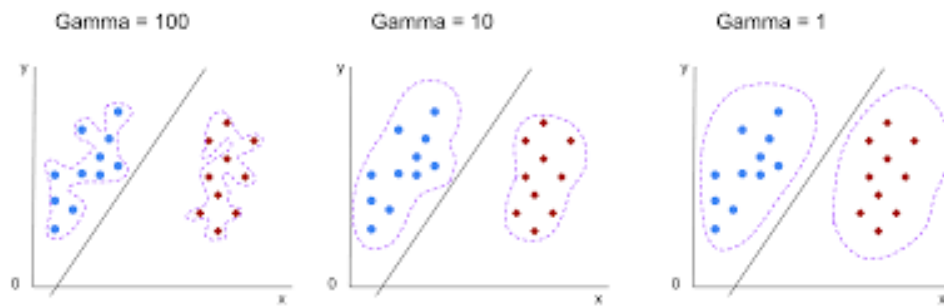
Gambar 2. 2 *Hyperplane* kelas -1 dan kelas +1 (Cahyono et al., 2020)

Pada gambar 2.2, ditunjukkan beberapa pola atau *pattern*. Pola tersebut merupakan anggota yang terdiri dari dua buah kelas, yaitu kelas +1 dan kelas -1. Pola yang tergabung pada kelas -1 disimbolkan dengan warna merah (kotak), sedangkan pola pada kelas +1, disimbolkan dengan warna kuning (lingkaran). Dengan mengukur margin *hyperplane* dan mendapatkan titik maksimal dari *hyperplane* tersebut, kita dapat menemukan *hyperplane* paling optimal yang memisahkannya antara kedua kelas. *Margin* merupakan jarak antara *node* atau titik pola terdekat dari masing-masing kelas dan *hyperplane* (Pamungkas & Romadhony, 2021). Berdasarkan ilustrasi yang ditampilkan pada Gambar 2.2, letak *hyperplane* berada di tengah antara dua kelas. Hal ini menunjukkan, bahwa jarak antara *hyperplane* dengan objek-objek data berbeda dengan kelas terluar atau yang paling dekat (ditandai dengan objek berbentuk kotak merah dan lingkaran kuning) dengan garis *hyperplane*. Dalam SVM, *support vector* merujuk pada objek data yang paling terluar, terletak paling dekat dengan *hyperplane*. Objek ini paling sulit dikenali atau diklasifikasikan disebabkan posisi yang tumpang tindih (*overlap*) dengan kelas lain. Maka dari itu, dalam menemukan *hyperplane* terbaik, hanya *support vector* yang menjadi perhitungan penting oleh SVM.

Pada SVM, terdapat Parameter C atau Parameter *Cost*. Parameter C adalah parameter yang digunakan dalam pengoptimalan SVM dalam menghindari misklasifikasi data yang terdapat pada dataset *training* dan sebagai nilai batas ambang antara margin dan error klasifikasi. Penentuan nilai Parameter C dilakukan dengan cara melakukan percobaan pada beberapa nilai dan dilihat atau dievaluasi efeknya pada nilai akurasi yang diperoleh pada SVM, contohnya dengan menggunakan metode Cross Validation dalam pengvalidasian. Penalti yang diberikan terhadap error pada pengklasifikasian akan semakin jika nilai C juga besar.(Al Farobi, 2021). Dengan kata lain, dengan nilai C yang lebih besar, kemungkinan kesalahan dalam proses klasifikasi akan lebih rendah. Sebaliknya, jika nilai C lebih kecil, proporsi kesalahan dalam menentukan solusi akan lebih tinggi.. Maka, penentuan nilai C yang paling baik dan paling optimal sangatlah penting.

Beberapa masalah lain yang dihadapi adalah tidak dapatnya data dipisahkan secara linear dalam ruang *input*, *soft margin* SVM tidak dapat menemukan pemisah atau *hyperplane* yang menyebabkan akurasi yang dihasilkan kecil dan data tidak tergeneralisasi dengan baik. Masalah ini dapat diatasi dengan kernel. Kernel dapat mentransformasikan data menjadi data ruang dimensi yang lebih tinggi atau dapat disebut ruang kernel yang berfungsi untuk memisahkan data secara linear. Beberapa fungsi kernel lainnya yang umum digunakan adalah kernel *Linear*, *Polynomial*, *Radial Basis Function* atau RBF dan Sigmoid(Alfarobi, 2021).

Parameter lainnya yang diperhitungkan dalam Algoritma SVM adalah nilai parameter gamma, khususnya pada kernel RBF, *Polynomial* dan *Sigmoid*. Parameter gamma mendefinisikan "kelebaran" atau "kemiringan" fungsi kernel. Kurva *decision boundary* menjadi rendah apabila nilai gamma rendah, membuat wilayah keputusan menjadi luas. Sebaliknya, jika nilai gamma tinggi, kurva dari *decision boundary* akan lebih rapat dengan data (Al-Mejibli et al., 2020). Ini menyebabkan pengaruh perlakuan model terhadap data. Semakin besar nilai gamma, model akan cenderung hanya mempertimbangkan data yang dekat *support vector*, yang dapat menyebabkan model menjadi terlalu spesifik dan *overfitting*. Sebaliknya, jika nilai gamma semakin kecil, maka model akan menjadi terlalu umum dan tidak dapat menangkap detail ataupun pola di dalam data. Ilustrasi dari konsep nilai parameter gamma ditunjukkan pada gambar 2.3



Gambar 2. 3 Parameter Gamma

Sumber : <https://stackabuse.com/understanding-svm-hyperparameters/>

Model Persamaan SVM dalam menghitung nilai *hyperplane* antara lain :

$$w \cdot x + b = 0 \quad (2.1)$$

Penjelasan:

w = parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik support vector)

x = data input SVM (x_1 = index kata, x_2 = bobot kata)

b = parameter *hyperplane* yang dicari (nilai bias)

2.6 Scraping

Scraping atau *Web Scraping* merupakan sebuah metode yang digunakan untuk mengkonversi data yang tidak terstruktur di web menjadi data yang terstruktur yang dapat disimpan dan dapat dianalisa di database lokal ataupun *spreadsheet* (De & Sirisuriya, 2015). *Web Scraping* juga didefinisikan sebagai proses ekstraksi dan penggabungan konten dari web secara semantik (Glez-Peña et al., 2013). *Web Scraping* diketahui sebagai teknik paling efisien dan terbaik untuk proses pengambilan data dalam skala besar atau *bigdata* (Zhao, 2017). Biasanya, Proses *Scraping* memanfaatkan *Hyper-text Transfer Protocol* (HTTP) ataupun melalui *browser web*, bisa dilakukan baik secara manual oleh pengguna maupun secara otomatis dengan bot atau *web crawler*. Proses dalam melakukan *Web Scraping* sejatinya dibagi kedalam dua tahap utama, yaitu proses memperoleh data dari sumber daya web dan proses ekstraksi

informasi yang diinginkan yang telah diperoleh ke berbagai macam format (Zhao, 2017).

Proses *Scraping* dimulai dari membuat sebuah permintaan HTTP untuk memperoleh sumber daya dari situs web yang dituju. Permintaan tersebut bisa berbentuk URL berisikan sebuah *GET query* ataupun perintah *POST query*. Setelah situs web menerima dan mengolah permintaan, sumber daya *web* dapat diperoleh dari website, lalu dikirim kembali ke program yang melakukan *web scraping*. Sumber daya dapat berbentuk dalam beberapa format, antara lain berbentuk XML JSON, maupun dalam bentuk multimedia seperti gambar, audio atau video. Setelah data web di download, proses ekstraksi dalam berlanjut ke proses menguraikan, memformat ulang, dan mengorganisir data ke dalam bentuk yang lebih terstruktur.

Ada dua modul utama yang digunakan untuk melakukan proses *scraping*, yaitu modul untuk meminta permintaan HTTP, seperti Urllib2 atau Selenium. Dan modul lainnya untuk mengurai dan mengekstraksi informasi dari kode HTML mentah, seperti BeautifulSoup dan Pyquery. Pada Penelitian ini, digunakan modul Selenium dan BeautifulSoup. Selenium adalah sebuah *web browser wrapper* yang membangun *web browser* seperti Google Chrome atau Microsoft Edge, yang memungkinkan pengguna untuk melakukan otomatisasi proses penjelajahan web dengan *programming*. Untuk proses ekstraksi data, digunakan BeautifulSoup, dimana di desain untuk proses *scraping* HTML ataupun XML dokumen lainnya. BeautifulSoup juga menyediakan menyediakan fungsi Pythonic yang memudahkan seperti fungsi navigasi, *searching* dan memodifikasi *parse tree*, sebuah *toolkit* untuk mendekompos sebuah file HTML dan mengekstraksi informasi yang diinginkan via lxml atau html5lib (Zhao, 2017). .

2.7 Bag of Words (BOW)

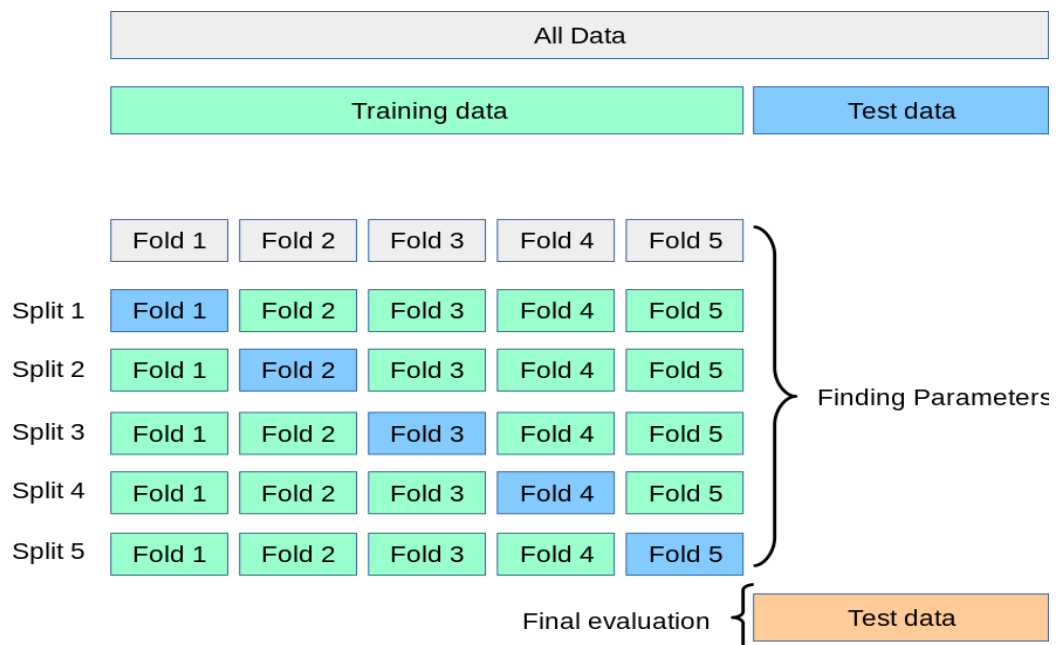
Bag of words merupakan salah satu metode yang dapat digunakan untuk merepresentasikan serangkaian teks panjang yang telah menjadi fitur tingkat kata. Konsep pada metode *Bag of words* adalah menganggap sebuah dokumen (paragraf ataupun kalimat) sebagai sebuah kumpulan kata-kata atau dapat digambarkan seperti tas karung berisikan kumpulan kata. Tetapi, pengaplikasiannya hanya mempertimbangkan duplikasi kata atau banyaknya jumlah kemunculan kata tersebut dan mengabaikan tata bahasa urutan dari kata-kata tersebut (Qader et al., 2019). Berdasarkan konsep tersebut, dokumen dapat direpresentasikan kedalam vektor dengan

ukuran *vocabulary*, dimana setiap kata yang muncul di dalam dokumen berkorespondansi dengan dimensi yang unik dan bernilai tidak nol (0). Selanjutnya, skor dapat dikomputasi ke dalam setiap kata (berdasarkan jumlah kemunculan) untuk menilai bobot kata dari kata di dalam dokumen. Bag of Words (BoW) umumnya digunakan dalam teknik klasifikasi dokumen, di mana berfungsi sebagai fitur untuk melatih algoritma klasifikasi berdasarkan frekuensi kemunculan kata-kata dalam dokumen. Metode *Bag of Words* ini juga berjalan dengan baik dalam pengaplikasian untuk *spam filtering*, klasifikasi teks dan *information retrieval* atau pengambilan informasi.

2.8 K-Fold Cross Validation

Cross Validation adalah metode pengambilan data ulang dengan tujuan menilai kemampuan suatu model prediksi dan untuk mencegah *overfitting*, yang memiliki kualitas "optimal" atau "terbaik", model bekerja dengan baik dengan data training yang "terbaik" tetapi berjalan dengan buruk ketika dilakukan percobaan dengan data baru. (Berrar, 2018).

K-Fold Cross Validation merupakan salah satu metode cross validation yang sangat umum digunakan. Prinsip kerja *K-Fold Cross Validation* Prinsip kerja Cross Validation K-Fold melibatkan pembagian data menjadi K atau N bagian dengan jumlah yang setara.



Gambar 2. 4 Ilustrasi Cross Validation dengan 5 Fold

Sumber : <https://vitalflux.com/k-fold-cross-validation-python-example/>

Ilustrasi pada gambar 2.4, dataset dibagi menjadi 5 bagian. Setelah itu, model dilatih menggunakan $K - 1$ bagian, dan selanjutnya dilakukan pengujian prediksi model pada bagian lainnya (sebagai data validasi). Hal ini bertujuan untuk mengestimasi kinerja model pada data uji. Proses ini dilakukan sebanyak K kali atau sebanyak jumlah fold yang ditentukan, dimana setiap K atau fold digunakan sebagai testing set sebanyak sekali secara bergilir. Selain itu, setiap subsampel sebaiknya yang serupa dengan sampel aslinya, seperti proporsi warna biru dan merah yang tetap sama di setiap partisi pada setiap iterasi. Konsep ini dikenal dengan istilah *stratified sampling*.

Dengan berdasarkan metode dan aturan yang telah disebutkan sebelumnya, stabilitas model diukur dengan menghitung varian kinerja dari K model. Nilai rata-rata kinerja yang dihasilkan memberikan estimasi terhadap kinerja model pada pengujian data. Model yang stabil adalah yang diinginkan, Ketika varians memiliki nilai kecil, itu menunjukkan bahwa eksperimen dapat direplikasi, dan hasil kinerja yang konsisten dapat diperoleh dengan pengulangan eksperimen. Sebaliknya, varians yang tinggi dapat menunjukkan bahwa nilai kinerja diperoleh secara acak atau kebetulan. (Wira & Putra, 2020.). Persamaan dalam menghitung nilai performa model sebagai berikut :

$$\text{Performa} = \frac{1}{K} \sum_{i=1}^K \text{Kinerja Model} \quad (2.2)$$

2.9 Ketidakseimbangan Data

Kasus ketidakseimbangan data adalah suatu hal yang sering dijumpai dan menjadi masalah utama dan penting dalam bidang pembelajaran mesin maupun *data mining* pada dasawarsa ini. Ketidakseimbangan data atau *data imbalanced* adalah suatu kondisi dimana data pada suatu kelas tidak terdistribusi dengan baik dan merata sehingga menyebabkan data di kelas tersebut lebih banyak dibanding kelas lainnya (Pulungan & Selvida, 2022). Kelas data yang memiliki data lebih banyak disebut kelas mayoritas, sedangkan kelas yang memiliki data lebih sedikit disebut kelas minoritas. Kesenjangan jumlah data antara kelas mayoritas dan kelas minoritas menyebabkan rasio ketidakseimbangan atau imbalanced rasio. Keadaan ini mengakibatkan bias terhadap data yang akan mempengaruhi kinerja pada pembelajaran mesin dalam memprediksi. Mesin akan kesulitan dalam memprediksi kemunculan kelas minoritas dikarenakan

kurang cukup dan banyaknya kelas minoritas untuk dipelajari oleh model sehingga model kurang terlatih dalam memprediksi kelas minoritas dengan akurat. Ketidakseimbangan data ini menyebabkan representasi yang dihasilkan cenderung rendah pada data kelas minor, sehingga data tersebut cenderung diabaikan oleh model. Hal ini sering terjadi terutama pada proses pengklasifikasian data *multiclass*.

Penggunaan dataset yang tidak seimbang ini juga menyebabkan sebagian besar pembelajaran atau algoritma klasifikasi cenderung mengklasifikasikan kelas yang menjadi mayoritas memiliki akurasi yang tinggi, sedangkan kelas yang menjadi minoritas memiliki akurasi yang rendah (Gu et al., 2016). Walaupun menghasilkan akurasi yang tinggi pada kelas mayoritas, tentu model tidak dapat mengidentifikasi dan mengklasifikasi kelas mayoritas dengan efektif. Selain metrik akurasi, metrik yang dapat digunakan sebagai metrik akurasi adalah f1-score. F1-score adalah nilai rata-rata dari *precision* dan *recall*, dimana *precision* mengukur sebaik mana model dapat memprediksi nilai benar untuk kelas positif, sementara *recall* mengukur sebaik mana model dapat mengidentifikasi secara akurat anggota kelas positif. *F1-score* adalah metrik akurasi yang membutuhkan kedua nilai dari *false positives* dan *false negatives* dalam proses penghitungan. Nilai f1-score biasanya lebih berguna dan lebih memberikan informasi terutama dalam klasifikasi terhadap data yang tidak seimbang (Seo et al., 2021).

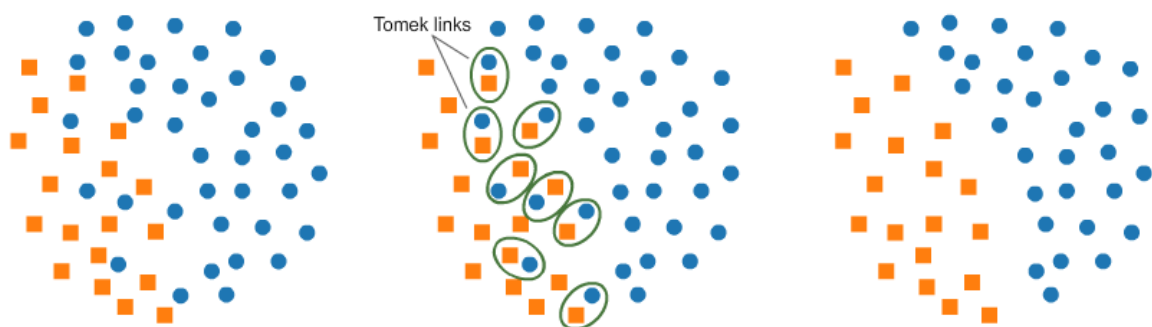
2.10 Resampling

Beberapa teknik dapat dilakukan dalam mengatasi ketidakseimbangan data. Salah satunya adalah menggunakan teknik *Resampling*. *Resampling* adalah sebuah teknik yang digunakan dalam menangani ketidakseimbangan data dengan cara menyesuaikan jumlah antara kelas mayoritas dan kelas minoritas (Sasada et al., 2020). Terdapat dua jenis metode *resampling*, yaitu *undersampling* dan *oversampling*. *Undersampling* bekerja dengan mengurangi beberapa jumlah data kelas mayoritas data latih. Dalam proses pengurangan ini, ada kemungkinan data yang dapat dikatakan baik terhapus dan meninggalkan data yang kurang baik, yang berpengaruh dalam proses pelatihan model. Keadaan ini menyebabkan ketidakmampuan model dalam membentuk aturan diskriminan yang diperlukan dalam klasifikasi, yang berimbas buruk terhadap keakuratan klasifikasi. Selain itu, Di sisi lain, *Oversampling* bekerja berkebalikan dengan *undersampling*, yaitu dengan menambahkan data yang memiliki karakteristik

yang mirip dengan data kelas minoritas, sehingga rasio antara kelas mayoritas dan minoritas tidak terlalu jauh perbedaannya. Tetapi, jika data minoritas yang digunakan untuk *oversampling* terlampaui sedikit, *overfitting* dapat terjadi karena data baru yang ditambahkan memiliki karakteristik yang mirip, sehingga tidak adanya keberagaman data untuk dilatih. Dalam penelitian ini, digunakan masing-masing 1 teknik dari *undersampling* dan *oversampling*, antara lain:

2.10.1 Tomek Links (T-Links)

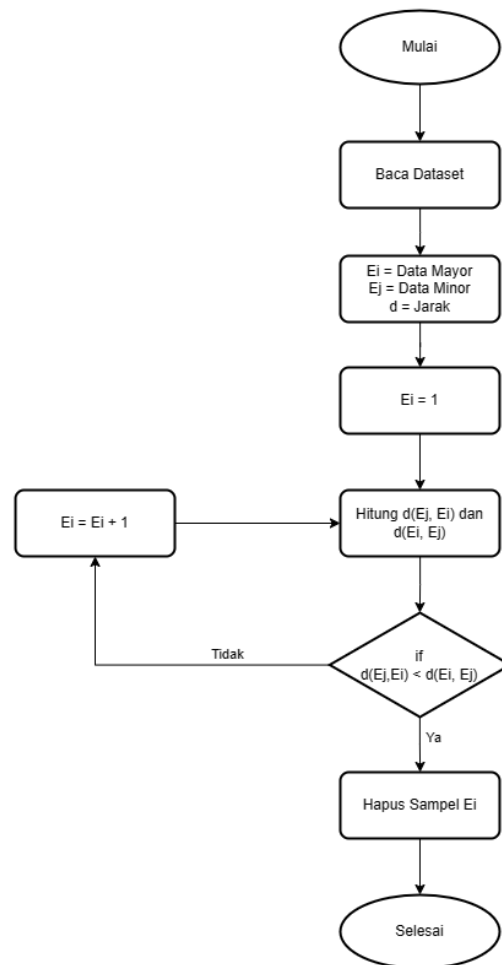
Tomek Links atau T-Links merupakan sebuah metode *undersampling*, dengan fungsi untuk menghapus atau mengurangi data pada kelas mayoritas. Tujuan Tomek Links adalah menghapus *noise* dan *borderline*, yang dapat mempersulit proses pengklasifikasian data yang tidak seimbang. Cara kerja Tomek Links adalah dengan mengidentifikasi karakteristik yang memiliki kesamaan antara kelas mayor dan minor (Sastrawan et al., 2010). Kelas mayoritas dan minoritas tersebut didefinisikan sebagai pasangan sehingga tidak ada lagi data diantara 2 kelas yang mirip dan berdekatan tersebut. Contoh, terdapat 2 kelas, diilustrasikan sebagai kelas E_i dan kelas E_j . Jarak antara E_i dan E_j disebut dengan $d(E_i, E_j)$. Sepasang E_i, E_j disebut Tomek Links apabila tidak terdapat kelas lain pada E_k , ada dengan kata lain $d(E_i, E_k) < d(E_i, E_j)$ atau $d(E_j, E_k) < d(E_i, E_k)$. Apabila terbentuk Tomek Links pada 2 sample tersebut, maka salah satu dari 2 sample tersebut adalah *noise* atau keduanya adalah *borderline*. Dapat disimpulkan, Tomek Links dapat adalah metode *undersampling* ketika sample dari kedua kelas dihapus. Implementasi dari Tomek Links dapat dilihat pada gambar 2.5 berikut ini



Gambar 2. 5 Ilustrasi Proses *Undersampling* Tomek Link

Sumber : <https://vijay-anandan.medium.com/how-to-resample-an-imbalanced-datasets-8e413dabbc21>

Flowchart dari proses cara kerja Tomek Links dapat dilihat pada gambar 2.6 berikut ini :



Gambar 2. 6 *Flowchart* Tomek Links

Sumber : Dokumentasi Pribadi

2.10.2 Borderline SMOTE

Borderline SMOTE adalah sebuah metode *oversampling* yang merupakan pengembangan dari metode SMOTE. SMOTE atau *Synthetic Minority Oversampling Technique* merupakan sebuah metode *oversampling* yang digunakan untuk menambahkan data pada kelas minoritas. Sesuai namanya, teknik *oversampling* ini bekerja dengan membuat sample kelas minoritas sintetis atau buatan berdasarkan kelas minoritas yang ada. Cara kerja metode SMOTE diawali dengan memilih satu sampel kelas minoritas secara acak. Sampel acak tersebut digunakan sebagai referensi dan dilakukan perhitungan *K-Nearest Neighbors* atau KNN dari sampel tersebut. Selanjutnya, akan dipilih salah satu tetangga atau *neighbor*. Dari kedua sampel tersebut

dihitung selisih atribut yang dimiliki oleh sampel referensi dan sampel tetangga. Selisih tersebut selanjutnya dikalikan dengan angka yang dipilih secara acak antara 0 sampai 1. Hasil perkalian yang diperoleh dimasukkan ke dalam sampel referensi kembali dan menjadi data sampel baru. Sampel baru tersebut digunakan kembali menjadi sampel referensi baru untuk mendapatkan sampel sintesis baru. Persamaan 2.3 Menampilkan perhitungan pembobotan dari sampel baru

$$X_{new} = X_i + (\hat{X}_i - x_i \times \delta) \quad (2.3)$$

Dimana X_i adalah atribut kelas minoritas, \hat{X}_i adalah nilai *k-nearest neighbor* untuk sampel X_i , dan δ adalah bilangan acak antara 0 sampai 1. Dengan mengimplementasikan SMOTE, data kelas minoritas dapat menjadi lebih beragam dengan menambahkan sampel sintesis atau buatan baru, yang diperoleh dari proses interpolasi sampel referensi dengan tetangga terdekat. Dengan bertambahnya jumlah data kelas minoritas, kesenjangan rasio antara kelas mayoritas dan minoritas akan berkurang serta representasi akan meningkat. Dengan begitu, tentu performa model akan meningkat dalam memprediksi kelas minoritas.

Perbedaan antara SMOTE dan *Borderline* SMOTE terletak pada fokus *Borderline* SMOTE yang terbatas pada duplikasi atau penguatan kelas minoritas saja (Trianto et al., 2022). Pengembangan *Borderline* SMOTE didasari dari keinginan mencapai hasil prediksi yang lebih baik. Dilakukan percobaan pembelajaran data yang berada di bagian *borderline* (garis batas) pada setiap kelas selama proses *training*. Dari hasil percobaan tersebut, seringkali terjadi misidentifikasi pada data contoh di daerah *borderline* dan sekitarnya, dibandingkan data yang terletak jauh dari *borderline*. Dengan alasan ini pula, disimpulkan bahwa bagian *borderline* perlu mendapat perhatian lebih, agar hasil prediksi menjadi lebih optimal (Han et al., 2005). Berdasarkan pertimbangan tersebut, Han et al. mengajukan pendekatan baru. Pendekatan tersebut dapat dikenal dengan *Borderline* SMOTE. Pendekatan ini secara khusus melakukan *oversampling* pada data contoh yang terletak di sekitar wilayah *borderline* pada kelas minoritas. Keseluruhan metode ini fokus pada identifikasi wilayah *borderline* pada kelas minoritas, diikuti dengan pembuatan sampel sintesis dari data di wilayah tersebut, yang kemudian digabungkan kembali ke dalam *dataset* asli.

Langkah-langkah dari Borderline SMOTE yang dijelaskan oleh Han et al. (2005) adalah sebagai berikut: Awalnya, sesuai dengan Persamaan 2.4, diasumsikan bahwa T mewakili dataset asli, P merupakan kelas minor, dan N adalah kelas mayor, di mana p_num dan n_num mewakili nilai total dari kelas minor dan mayor.

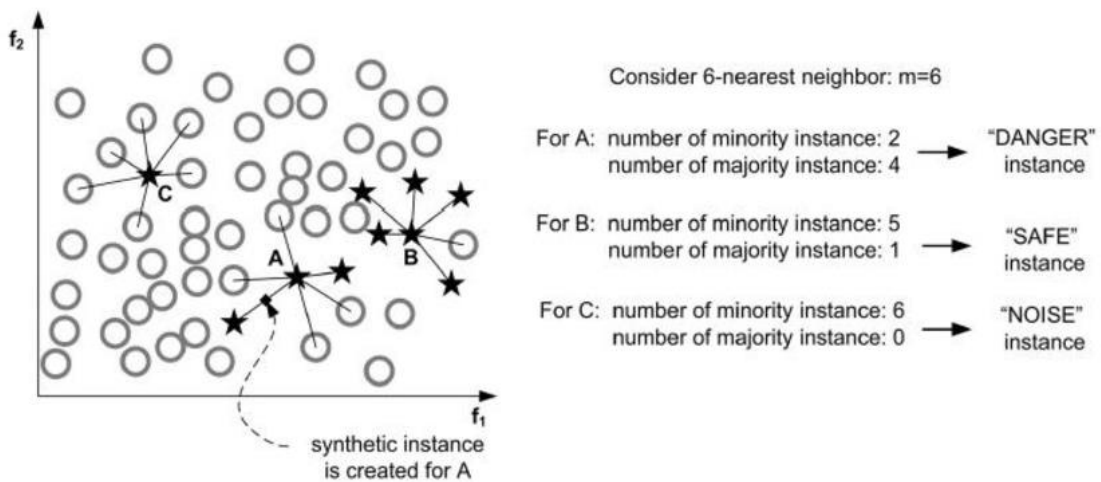
$$P = \{p_1, p_2, \dots, p_{num}\}, N = \{n_1, n_2, \dots, n_{num}\} \quad (2.4)$$

Pada setiap p_i ($i=1,2,\dots,p_num$) dalam kelas minor P, langkah selanjutnya adalah menghitung tetangga terdekat (m) dari dataset T. Semua sampel dari kelas minor yang terletak dalam jarak tetangga terdekat (m) ditandai sebagai m' ($0 \leq m' \leq m$). Sampel-sampel dalam tetangga terdekat (m) kemudian dikelompokkan ke dalam tiga himpunan, yaitu DANGER, SAFE, dan NOISE. Prosedur pembagian ke dalam masing-masing himpunan dijelaskan pada Persamaan 2.5, 2.6, dan 2.7. Ilustrasi dari pembagian himpunan dapat dilihat pada Gambar 2.7.

$$\text{If } m' = m \quad \text{then } p_i \text{ is "NOISE"} \quad (2.5)$$

$$\text{If } \frac{m}{2} \leq m' < m \quad \text{then } p_i \text{ is "DANGER"} \quad (2.6)$$

$$\text{If } 0 \leq m' < \frac{m}{2} \quad \text{then } p_i \text{ is "SAFE"} \quad (2.7)$$



Gambar 2. 7 Ilustrasi Pembagian Himpunan m -tetangga terdekat (Onur Durahim, 2016)

Sampel yang termasuk dalam himpunan DANGER merupakan sampel yang berada di wilayah *borderline* pada kelas minor P, sehingga dapat disimpulkan bahwa DANGER

$\subseteq P$. Oleh karena itu, dapat dijelaskan dengan Persamaan 2.8, yaitu $0 \leq d_{num} \leq p_{num}$. Selanjutnya, nilai tetangga terdekat (k) dari data P akan kembali dilakukan penentuan.

$$DANGER = \{p'_1, p'_2, \dots, p'_{dnum}\} \quad (2.8)$$

Selanjutnya, sampel sintetis di daerah DANGER akan dibuat. Pertama, akan dipilih nilai acak s antara 1 dan k . Lalu, jarak dif_j ($j = 1, 2, \dots, s$), antara p'_i dan s akan dihitung. Setelah itu, lakukan perkalian terhadap dif_j dengan bilangan acak r_j ($j = 1, 2, \dots, s$) yang memiliki rentang 0 – 1. Sampel sintetis baru s sekarang telah berhasil dibuat di antara p'_i dan tetangga terdekatnya. Rincian prosedur ini dapat ditemukan pada persamaan 2.9. Untuk mendapatkan sampel sintetis di seluruh wilayah borderline, langkah-langkah tersebut diulang untuk setiap p'_i dalam himpunan DANGER.

$$G = (N_{num} - P_{num}) \quad (2.9)$$

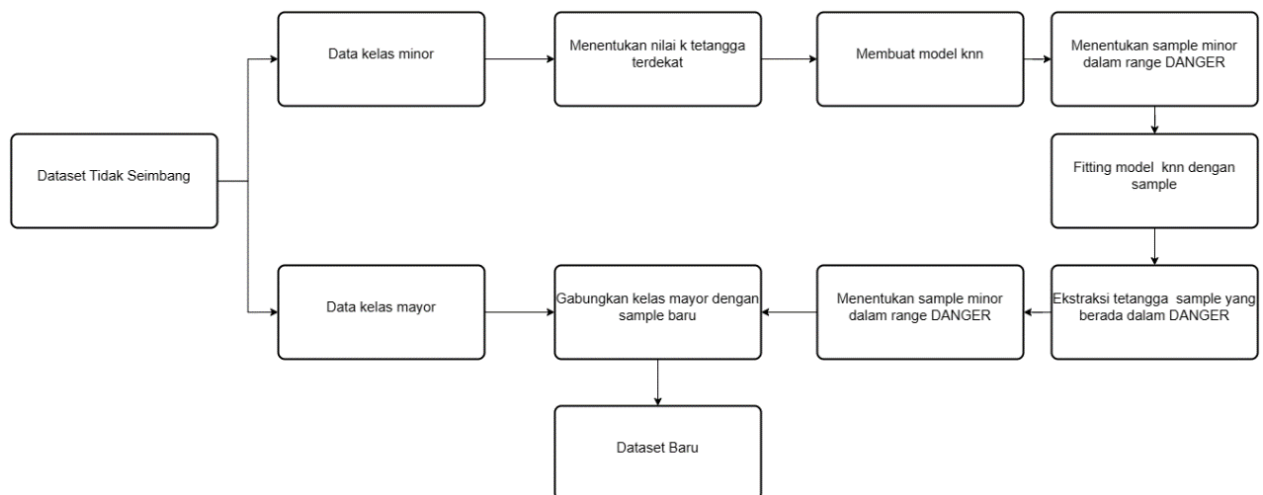
Penjelasan:

G = jumlah sampel sintetis yang dapat dibuat

N_{num} = jumlah data mayoritas

P_{num} = jumlah data minoritas

Berikut ini merupakan *flow chart* dari *Borderline SMOTE* yang dapat dilihat pada gambar 2.8



Gambar 2. 8 Alur Kerja Borderline SMOTE

Sumber : Dokumentasi Pribadi

2.11 Restoran

Restoran merupakan suatu tempat atau bangunan yang diorganisir secara komersil, yang menyelenggarakan pelayanan dengan baik kepada semua konsumennya baik berupa makanan maupun minuman (Atmodjo, 2005). Menurut Prof. Vanco Christian dari School Hotel Administration di Cornell University, tujuan operasional restoran adalah mencapai keuntungan atau *profit*. Sementara itu, selain aspek bisnis dan pencapaian keuntungan, kepuasan konsumen juga dianggap sebagai tujuan operasional utama yang harus dicapai oleh restoran. (Atmodjo, 2005).

Berdasarkan dari studi literatur oleh Gagic et.al pada tahun 2013 yang membahas terkait ekspektasi dan tingkat kepuasan konsumen serta persepsi kualitas pelayanan di industri kuliner, kualitas dari sebuah restoran dikategorikan ke empat dimensi atau aspek antara lain Kualitas Makanan, Kualitas Pelayanan, Lingkungan Fisik atau Bangunan, serta Tingkat Harga (Nurifan et al.,2019).

2.12 Penelitian Terdahulu

Beberapa penelitian terdahulu antara lain penelitian yang oleh Sari dan Wibowo (2019). Pada penelitian tersebut, peneliti melakukan Analisis Sentimen terhadap Toko Online JD.ID. Tujuan dari penelitian ini ialah untuk mendapatkan informasi berbentuk penilaian opini dari pengguna Toko Online JD.ID. Informasi yang didapatkan ini sangatlah berguna bagi pelanggan lain maupun pihak toko. Dataset yang digunakan pada penelitian ini adalah berupa opini yang diberikan pengguna JD.ID pada platform Twitter yang didapatkan melalui proses *crawling*. Dataset yang digunakan sebanyak 900 tweets, dibagi menjadi 300 *tweets* positif, 300 *tweets* negative dan 300 *tweets* netral. Pada penelitian ini digunakan metode sampling *Non Random Sample* atau Sampel Tidak Acak. Pada Metode *Non Random Sample*, setiap elemen atau anggota populasi tidak memiliki peluang yang sama untuk dijadikan sampel karena pada proses pengambilan sampel berdasarkan kriteria tertentu. Pada penelitian ini, Penulis menggunakan algoritma *Naïve Bayes Classifier* atau NBC. Nilai Akurasi yang dihasilkan pada penelitian Analisis Sentimen menggunakan algoritma *Naïve Bayes Classifier* adalah sebesar 96,44% dengan tanpa pembobotan TF-IDF. Sedangkan dengan pembobotan TF-IDF menghasilkan sebesar 98%. Peningkatan akurasi dengan pembobotan TF-IDF dibandingkan dengan tanpa pembobotan TF-IDF adalah sebesar 1,56%

Selanjutnya, penelitian yang dilakukan oleh Amalia dan Winarko (2021). Pada penelitian tersebut, peneliti yakni melakukan Analisis Sentimen berbasis aspek terhadap review restoran berbahasa Indonesia menggunakan kombinasi dari *Convolutional Neural Network* atau CNN dan *Contextual Word Embedding*. Penelitian ini bertujuan mendapatkan opini yang dihasilkan dari *review* atau ulasan yang diberikan terhadap suatu restoran. Informasi yang terkandung dalam ulasan ini sangat berguna bagi calon pengunjung/pelanggan atau *customer* untuk mengidentifikasi produk makanan ataupun layanan yang diberikan atau untuk pemilik restoran sebagai sarana evaluasi untuk memenuhi kebutuhan pengunjung/pelanggan atau *customer*. Pada penelitian ini, Peneliti menggunakan *dataset* berupa *review* atau ulasan yang diperoleh melalui situs Zomato, Tripadvisor, Traveloka, serta Qraved. Pada penelitian ini, digunakan kombinasi model *Convolutional Neural Network* dan *Contextualized Word Embedding* yang selanjutnya akan dilakukan perbandingan ke 3 model yaitu *BERT-CNN*, *ELMo-CNN*, dan *Word2vec-CNN*. Hasil yang didapatkan dari proses pengujian, model *ELMo-CNN* mendapatkan hasil terbaik dengan nilai *micro-average precision* sebesar 0.88, *micro-average recall* sebesar 0.84, dan *micro-average f1-score* sebesar 0.86. Sedangkan untuk klasifikasi sentimen memberikan hasil terbaik pada model *BERT-CNN* dengan nilai *precision* sebesar 0.89, *recall* sebesar 0.89, dan *f1-score* sebesar 0.91.

Penelitian lainnya yaitu penelitian yang dilakukan oleh Nurifan et al.(2019). Pada penelitian tersebut, peneliti melakukan Analisis Sentimen berbasis aspek terhadap review restoran, dimana pada tahap kategorisasi aspek menggunakan pendekatan *Hybrid ELMo-Wikipedia* (HEW) untuk *Aspect Categorization* (AC) dan pada tahap sentiment analisis (SA) nya menggunakan pendekatan *Hybrid Expanded Opinion Lexicon-SentiCircle* (HEOLS). Pada penelitian ini, dataset yang digunakan diperoleh dari *International Workshop on Semantic Evaluation 2016* atau SemEval 2016 Task 5 [27] Subtask 1 tentang *restaurant reviews*. Pada penelitian ini, dilakukan modifikasi pada *Double Propagation* (DP) dan *Topic Modelling* (TM) pada tahap *Aspect Term Extraction* (ATE). Dengan dilakukannya modifikasi tersebut, hasiln menunjukkan peningkatan nilai f1 dari tahap kategorisasi aspek atau *Aspect Categorization* (AC) sebesar rata-rata 2%. Lalu dengan digunakannya *Hybrid ELMo-Wikipedia* (HEW)

menunjukkan peningkatan nilai f1 dibandingkan metode sejenis sebesar rata-rata 6%. Selain itu, HEOLS dapat mengembangkan dan menentukan polaritas dari *Opinion Lexicon* dan meningkatkan nilai f1 pada tahap sentiment analisis sebesar 6%.

Penelitian lainnya yaitu yang dilakukan oleh Juwianto et al(2020). Pada penelitian tersebut, peneliti melakukan Analisis Sentimen terhadap opini dari suatu perusahaan berbahasa Indonesia. Tujuan dari Analisis sentimen ini guna mengetahui tingkat kepuasan pelanggan terhadap suatu perusahaan yang selanjutnya dapat digunakan untuk meningkatkan kualitas pelayanan serta menaikkan *brand* perusahaan . Pada penelitian ini, digunakan algoritma *Convolutional Neural Network* atau CNN, dengan menggunakan pendekatan Word2Vec. Pada penelitian ini, digunakan *Dataset* opini dalam bentuk *tweets* yang diperoleh melalui proses crawling dari situs *Twitter*. Total Data *tweets* yang diperoleh sebanyak 3.129 *tweets* yang selanjutnya dilakukan proses penghilangan *noise* sehingga menghasilkan sebanyak 999 *tweets* yang dapat digunakan digunakan agar proses sentiment menjadi lebih akurat. Data *tweets* ini kemudian di *labelling* atau dikelompokkan menjadi 3 jenis yaitu opini positif, negative, serta netral. Kemudian, data *tweets* tersebut dibagi menjadi pembagian 75:25 dimana 75% data digunakan untuk *training* data dan 25% untuk *testing* data. Hasil yang diperoleh dari proses pengujian dengan menggunakan model Word2Vec dengan dimensi 100 menghasilkan tingkat akurasi terbaik. Selain itu, hal lain yang memengaruhi tingkat akurasi adalah pemilihan jumlah filter, dimana jumlah filter dengan ukuran 50 menghasilkan nilai rata-rata akurasi sebesar 65,43% pada 3 I label dan 73,10% pada *sentiment* 2 label.

Penelitian selanjutnya dilakukan oleh Yang et al. (2020). Pada penelitian ini, peneliti melakukan sentiment analisis terhadap teks berbahasa Cina. Pada penelitian ini, digunakan metode pendekatan Elmo-RNN. Pendekatan Elmo-RNN ini digunakan dengan tujuan agar nilai keakuratan yang didapatkan lebih tinggi atau lebih akurat dalam mengartikan kata bermakna ganda dibandingkan teknologi *Word Embeddings* tradisional seperti *Word2Vec* maupun GloVe. Pada penelitian ini, digunakan *Dataset* yang diperoleh menggunakan metode *crawling* terhadap komentar dari pengguna *shampoo* pada situs JingDong atau JD. *Dataset* ini kemudian dikelompokkan menjadi dua jenis berdasarkan warna emosi pada komentar. Komentar positif ditandai dengan 1 dan negative ditandai dengan 0. Terdapat 10000 data, termasuk 5000 komentar positif

dan 5000 komentar negative.. Hasil dari pengujian terhadap *Dataset* tersebut menampilkan bahwa penggunaan Elmo-RNN dapat menghasilkan tingkat akurasi sebesar 0.8891. Nilai akurasi yang dihasilkan Elmo-RNN mendapatkan tingkat akurasi tertinggi dibandingkan CNN yang menghasilkan sebesar 0.8285, BiLSTM sebesar 0.8614, *Baseline* sebesar 0.8375, dan CNN+BiGRU sebesar 0.8756.

Penelitian lainnya yaitu penelitian yang dilakukan oleh Yustihan et al, (2021). Pada penelitian ini, peneliti melakukan Analisis Sentimen berbasis Aspek terhadap *review* atau ulasan rumah makan yang diberikan oleh pelanggan rumah makan. Pada penelitian ini, digunakan algoritma *Support Vector Machine* atau SVM. *Dataset* yang digunakan pada penelitian ini berupa data *review* atau ulasan rumah makan yang diperoleh dari *SemEval-2016 Task 5*, dengan sebanyak 300 data latih dan 100 data uji. Data tersebut selanjutnya akan dilakukan proses sentiment analisis berbasis aspek dengan aspek-aspeknya antara lain *Ambience*, *Drink*, *Food*, *Restaurant*, serta *Service*. Hasil dari pengujian terhadap *Dataset* tersebut menggunakan algoritma *Support Vector Machine* dengan pendekatan macro averaging menghasilkan precision sebesar 0,94, recall sebesar 0,6, accuracy sebesar 0,88, dan f-measure sebesar 0,73. Hasil tersebut menunjukkan bahwa sistem sudah cukup tepat dalam mengidentifikasi kelas yang bukan termasuk aspeknya dengan benar namun sistem belum cukup berhasil untuk mengidentifikasi semua kelas yang termasuk aspeknya dengan benar dikarenakan jumlah data yang tidak seimbang saat proses training pada setiap kelas aspek data. Hasil tersebut juga menunjukkan bahwa sistem sudah cukup mampu dalam mengidentifikasi kelas yang mempunyai sentimen positif dengan benar, namun belum berhasil mengidentifikasi kelas yang mempunyai sentimen negatif.

Selanjutnya, penelitian yang dilakukan oleh Hairani et al. pada tahun 2023. Pada penelitian ini, peneliti melakukan Klasifikasi Data Penderita Diabetes menggunakan metode *Random Forest*. Data yang digunakan untuk pelatihan bersifat tidak seimbang, dimana data kelas negatif berjumlah lebih banyak, tepatnya 500 data, dibanding data kelas positif, yaitu sebanyak 268 data. Digunakan 2 skenario dalam mengatasi ketidakseimbangan ini dengan cara *meresampling* data, yaitu hanya menggunakan teknik SMOTE, dan menggunakan Teknik SMOTE+Tomek Links. Hasil uji dengan menggunakan teknik *K-Fold Cross Validation* dengan *fold* sebanyak 10 Fold menunjukkan bahwa algoritma Random Forest dengan SMOTE-Tomek Links

mendapatkan akurasi, sensitivitas, presisi, dan F1-score tertinggi dibandingkan dengan *Random Forest* dengan SMOTE, dengan dengan akurasi sebesar 86,4%, sensitivitas sebesar 88,2%, presisi sebesar 82,3%, dan F1-score sebesar 85,1%. Penggunaan Smote-TomekLink dapat meningkatkan kinerja metode random forest berdasarkan akurasi, sensitivitas, presisi, dan F1-score.

Tabel 2. 1 Penelitian Terdahulu

No.	Peneliti	Judul	Tahun	Keterangan
1.	Fransiska Vina Sari dan Arief Wibowo	Analisis Sentimen Pelanggan Toko Online JD.ID menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi	2019	Menggunakan Metode Naïve Bayes serta metode Non Random Sampling untuk proses samplingnya. Menghasilkan akurasi sebesar 96,44% dengan tanpa pembobotan TF-IDF dan 98% dengan pembobotan TF-IDF
2.	Putri Rizki Amalia dan Edi Winarko	Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding	2021	Menggunakan metode gabungan dari CNN dan Contextual Word Embedding. <i>ELMo-CNN</i> mendapatkan hasil terbaik, dengan nilai <i>micro-average</i> <i>precision</i> sebesar 0.88, <i>micro-average recall</i> sebesar 0.84, dan <i>micro-average f1-</i> <i>score</i> sebesar 0.86. Hasil klasifikasi sentimen terbaik dihasilkan model <i>BERT-</i> <i>CNN</i> dengan nilai <i>precision</i> sebesar 0.89, <i>recall</i> sebesar

					0.89, dan <i>f1-score</i> sebesar 0.91.
3.	Farza Nurifan, Riyanarto Sarno, dan Kelly Rossa Sungkono	<i>Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle</i>	2019	Menggunakan metode gabungan Double Propagation dan Topic Modelling untuk Ekstraksi Aspek Istilah, TF-ICF untuk Ekstraksi Aspek <i>Keyword</i> , Hybrid ELMo-Wikipedia untuk Pengkategorian Aspek, dan HEOLS untuk Sentimen Analisis. Menghasilkan peningkatan nilai <i>f1</i> pada pengkategorian aspek sebesar rata-rata 2% dan	
4.	Hans Juwiantho, Esther Irawati Setiawan, Joan Santoso, dan Mauridhi Hery Purnomo	<i>Sentiment Analysis Twitter Bahasa Indonesia berbasis Word2Vec menggunakan Deep Convolutional Neural Network</i>	2020	Menggunakan metode Convolutional Neural Network dengan pendekatan Word Embedding Word2Vec. Hasil terbaik didapatkan dengan model Word2Vec dengan dimensi 100. Pemilihan jumlah filter mempengaruhi nilai akurasi , dimana jumlah filter dengan ukuran 50 menghasilkan nilai rata-rata akurasi sebesar 65,43% pada 3 I label dan 73,10% pada <i>sentiment</i> 2 label.	

5.	Min Yang, Junwu Xu, Kai Luo, dan Yiping Zhang	Sentiment analysis of Chinese text based on Elmo-RNN model	2021	Sentiment Analisis teks berbahasa Cina dengan menggunakan model ELMo-RNN. ELMo-RNN menghasilkan akurasi sebesar 0.8891, lebih tinggi dibandingkan CNN BiLSTM , <i>Baseline</i> , dan CNN+BiGRU.
6.	Salsabila Rahma Yustihan, Putra Pandu Adikara, dan Indriati	Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM)	2021	Sentimen Analisis Berbasis Aspek menggunakan Metode SVM dengan aspek-aspeknya antara lain <i>Ambience</i> , <i>Drink</i> , <i>Food</i> , <i>Restaurant</i> , dan <i>Service</i> . Menghasilkan precision sebesar 0,94, recall sebesar 0,6, accuracy sebesar 0,88, dan f-measure sebesar 0,73
7.	Hairani Hairani, Anthony Anggrawan, Dadang Priyanto	Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link	2023	Peningkatan performa klasifikasi data diabetes menggunakan metode <i>Random Forest</i> yang datanya bersifat tidak seimbang. Penggunaan SMOTE-TomekLinks menghasilkan nilai tertinggi dengan akurasi sebesar 86,4%, sensitivitas sebesar 88,2%, presisi sebesar 82,3%, dan F1-score sebesar 85,1%.

2.13 Perbedaan Penelitian

Terdapat perbedaan penelitian ini dengan penelitian-penelitian yang telah disebutkan sebelumnya. Pada penelitian sebelumnya, beberapa peneliti masih melakukan penelitian pada tahap *sentiment analysis*. Adapun pada penelitian yang melakukan sampai tahap *Aspect Based Sentiment Analysis*, menggunakan dataset berbahasa Inggris, tepatnya dataset SemEval 2016 Task 5. Perbedaan lainnya yang membedakan penelitian ini dengan penelitian lainnya ialah perbedaan objek yang digunakan atau dipilih, perbedaan algoritma, serta penggunaan teknik *resampling Borderline SMOTE+TomekLinks* untuk mengatasi ketidakseimbangan data pada dataset, tepatnya dataset ulasan rumah makan. Sehingga, pada penelitian ini, peneliti melakukan penelitian analisis sentiment berbasis aspek dengan menggunakan algoritma Support Vector Machine dengan *feature extraction Bag of Words* dan teknik *resampling Borderline SMOTE+TomekLinks*, yang menghasilkan sebuah sistem yang dapat memberikan informasi berdasarkan aspek makanan, minuman, pelayanan, tempat, harga, dari rumah makan yang di *inputkan* oleh user secara *realtime*.

BAB 3

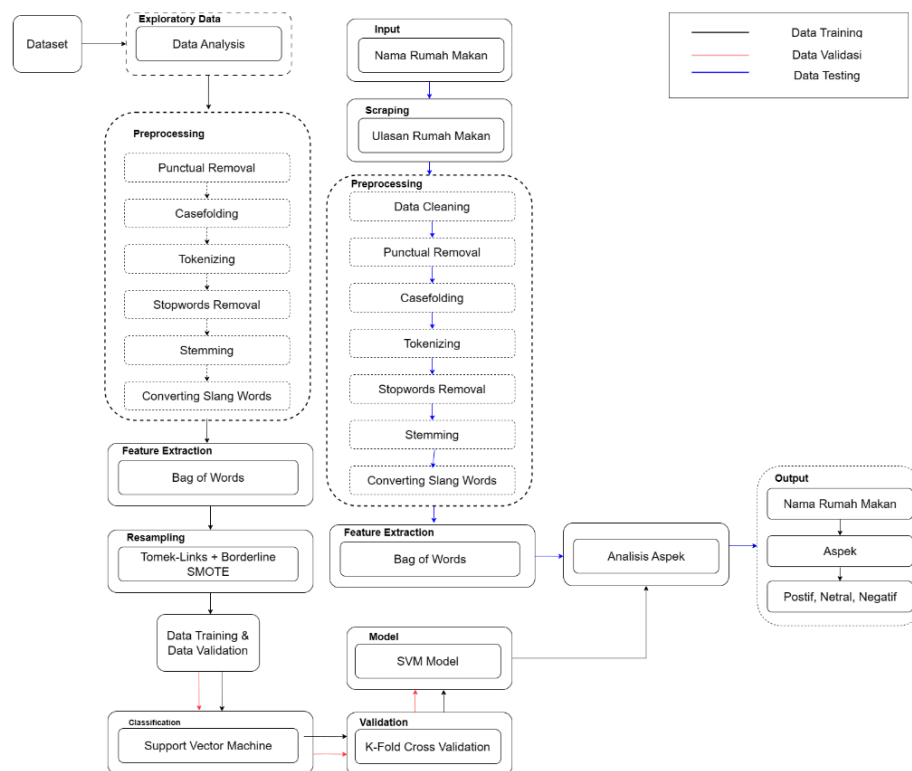
ANALISIS DAN PERANCANGAN SISTEM

3.1 Dataset

Data yang digunakan pada penelitian ini merupakan kumpulan *review* rumah makan yang diperoleh melalui proses crawling dari platform Google Maps, dan disimpan kedalam format *csv*. Review yang diperoleh mengandung penilaian dari aspek Makanan, Minuman, Tempat/Bangunan, Layanan, serta Harga. Ulasan yang diperoleh berasal dari 3 rumah makan berbeda menggunakan proses crawling. Jumlah keseluruhan data dari proses crawling berjumlah 1005 data, dengan 500 data dari Warkop Agem Senyum, 250 data dari Warkop Iwan, dan 255 data untuk rumah makan Soto Bathok. Data berjumlah 1005 tersebut nantinya akan digunakan sebagai data latih. Proses penentuan label suatu kalimat harus melewati proses validasi terlebih dahulu dengan melibatkan pandangan dari berbagai pihak (William et al, 2020). Proses validasi dengan cara me *label* dengan beberapa orang *annotator* dapat dilakukan. Data latih dilakukan pemberian label secara manual melalui 39 orang *annotator* yang merupakan mahasiswa yang memiliki kemampuan untuk menilai suatu sebuah ulasan rumah makan berdasarkan informasi tekstual. Data tersebut dilabel kedalam 5 aspek yaitu aspek makanan, minuman, pelayanan, tempat dan haraga, dan di label kedalam tiga kelompok sentiment yaitu ulasan dengan sentiment positif, netral, serta negatif. Skenario validasi dilakukan dengan cara membagi data review ke dalam 13 form kuisioner *Google Form*, dengan setiap form berisi 77 pertanyaan penilaian validasi data. Penulis kemudian memberikan form kuisioner tersebut ke 39 orang responden tersebut, yang dikelompokkan menjadi kelompok-kelompok beranggotakan 3 orang dan setiap kelompok responden melakukan penilaian terhadap data yang berbeda. Responden memiliki latar belakang yaitu memiliki tingkat pemahaman yang sama dalam menilai suatu ulasan rumah makan, dengan umur 19 tahun.

3.2 Arsitektur Umum

Pada penelitian ini, dilakukan beberapa tahapan, antara lain pengumpulan data berupa ulasan suatu rumah makan yang diambil melalui platform Google Maps pada bagian Ulasan Restoran melalui proses crawling. Ulasan tersebut selanjutnya disimpan kedalam format .csv. Setelah data dikumpulkan, Data yang telah disimpan memasuki proses *Pre Processing*. Tahapan *Pre Processing* antara lain terdiri dari *Data Cleaning*, *Punctual Removal*, *Casefolding*, *Tokenization*, *Stopword Removal*, *Stemming*, dan *Data Normalization*. Data yang sudah diproses selanjutnya memasuki tahap *feature extraction*. *Bag of words* yang digunakan sebagai metode *feature extraction* mengubah data kedalam bentuk vektor di dalam matrix. Matrix yang berisi vektor selanjutnya melewati tahap *resampling* Tomek-Links dan *Borderline SMOTE*. Setelah diberlakukan proses resampling, dilakukan proses *data splitting*. Data dibagi menjadi Data *Training* dan Data *Testing*. Kedua data tersebut menjadi *input* dalam proses *training*. Setelah proses *training* selesai, selanjutnya memasuki tahap validasi data untuk mengevaluasi performa model. Setelah di evaluasi, dihasilkan sebuah model yang akan digunakan dalam tahap analisis aspek. Setelah model dihasilkan, model akan diuji kemampuannya dalam melakukan sentimen berbasis aspek terhadap kumpulan ulasan yang di scraping berdasarkan nama restoran yang di input. Selanjutnya akan dilakukan penyimpulan terhadap sentimen berbasis aspek tersebut.



Gambar 3. 1 Gambar Arsitektur Umum

Sumber : Dokumentasi Pribadi

3.2.1 Exploratory Data Analysis

Dataset yang telah dikumpulkan sebelumnya, yang telah dijelaskan pada subbab 3.1, selanjutnya memasuki tahap Exploratory Data Analysis. Pada tahap ini, dilakukan eksplorasi terhadap dataset *training* yang sudah dikumpulkan. Tujuan dari tahap ini ialah untuk mengidentifikasi serta memahami karakteristik dan hubungan antar variabel dalam dataset. Dilakukan penghitungan terhadap sentimen masing-masing kelas aspek yang terdapat pada dataset. Dengan melakukan penghitungan, dapat ditentukan apakah dataset seimbang atau tidak seimbang. Hasil penghitungan dalam dilihat pada gambar 3.1, 3.2, 3.3, 3.4, 3.5, dan 3.6 berikut:

```

Positif      633
Netral       347
Negatif      25
Name: makanan, dtype: int64
  
```

Gambar 3. 2 Jumlah Dataset Aspek Makanan

Sumber : Dokumentasi Pribadi

Dari eksplorasi yang dilakukan, ditemukan bahwa pada terdapat 633 data “Positif” , 347 data “Netral”, dan 25 data “Negatif” pada aspek kelas Makanan.

```
Netral      866
Positif     128
Negatif      11
Name: minuman, dtype: int64
```

Gambar 3. 3 Jumlah Dataset Aspek Minuman

Sumber : Dokumentasi Pribadi

Pada aspek kelas Minuman, ditemukan sebanyak Ditemukan sebanyak 128 data “Positif” , 866 data “Netral”, dan 11 data “Negatif”.

```
Netral      804
Positif     145
Negatif      56
Name: pelayanan, dtype: int64
```

Gambar 3. 4 Jumlah Dataset Aspek Pelayanan

Sumber : Dokumentasi Pribadi

Pada aspek kelas Pelayanan, ditemukan sebanyak 145 data “Positif” , 804 data “Netral”, dan 56 data “Negatif”

```
Netral      660
Positif     271
Negatif      74
Name: tempat, dtype: int64
```

Gambar 3. 5 Jumlah Dataset Aspek Tempat

Sumber : Dokumentasi Pribadi

Pada aspek kelas Tempat, ditemukan sebanyak 271 data “Positif” , 660 data “Netral”, dan 74 data “Negatif”.

```

Netral      798
Positif     191
Negatif     16
Name: harga, dtype: int64

```

Gambar 3. 6 Jumlah Dataset Aspek Harga

Sumber : Dokumentasi Pribadi

Dan pada aspek kelas Harga, ditemukan sebanyak 191 data “Positif” , 798 data “Netral”, 16 data “Negatif”.

Dapat disimpulkan, dataset yang digunakan bersifat tidak seimbang, dimana mayoritas kelas aspek memiliki data berlabel “Netral”. Selain itu, data berlabel “Negatif” memiliki jumlah yang terpaut jauh dibandingkan data berlabel ”Positif”.

3.2.2 Pre-processing

Di tahap ini dataset akan di dilakukan pre-processing dalam berbagai tahapan antara lain *Data Cleansing*, *Punctual Removal*, *Casefolding*, *Stopword Removal*, *Stemming*, *Tokenization*, dan *Data Normalization*.

3.2.2.1 Data Cleansing


Pada tahap Data Cleansing, dilakukan pembersihan data pada dataset yang kurang baik serta tidak memberikan informasi. Contohnya seperti ulasan “bagus”, “oke”, “mantap”. Pada tahap ini, diberlakukan ketentuan untuk menghapus ulasan dibawah 10 karakter.

3.2.2.2 Punctual Removal

Punctual Removal berfungsi penghilangan tanda baca, angka, maupun karakter special seperti simbol dan *emoticon*. Tujuan dari tahap Punctual Removal adalah agar data terbebas dari karakter-karakter selain huruf, sehingga menghasilkan data yang bersih dan siap diolah ke tahap selanjutnya.

Tabel 3. 1 *Input dan output proses Punctual Removal*

Sebelum <i>Punctual Removal</i>	Sesudah <i>Punctual Removal</i>
Makanannya enak, Harganya murah, Pegawainya juga ramah... Kurangnya tempatnya agak sempit aja...	Makanannya enak Harganya murah Pegawainya juga ramah Kurangnya

Keseluruhan restoran ini udah mantap lah 	tempatnya agak sempit aja Keseluruhan restoran ini udah mantap lah
---	--

3.2.2.3 Casefolding

Tahap Casefolding mengubah semua kata berawalan huruf kapital menjadi huruf kecil. Tujuan dari *casefolding* adalah meningkatkan keakuratan data dan menghindari data yang *redundant* atau berulang. Contohnya pada kata “Masak” menjadi “masak”

Tabel 3. 2 *Input dan output proses Casefolding*

Sebelum <i>Casefolding</i>	Sesudah <i>Casefolding</i>
Makanannya enak Harganya murah Pegawainya juga ramah Kurangnya tempatnya agak sempit aja Keseluruhan restoran ini udah mantap lah	makanannya enak harganya murah pegawainya juga ramah kurangnya tempatnya agak sempit aja keseluruhan restoran ini udah mantap lah

3.2.2.4 Tokenization

Pada tahap Tokenization, dilakukan pemecahan kalimat atau teks mentah menjadi sebuah kata tunggal. Kata tunggal ini disebut dengan token. Tujuan tahap *tokenization* ini adalah mempermudah dalam perhitungan kemunculan setiap kata.

Tabel 3. 3 *Output proses Tokenization*

Setelah dilakukan <i>Tokenization</i>
['makanannya', 'enak', 'harganya', 'murah', 'pegawainya', 'juga', 'ramah', 'kurangnya', 'tempatnya', 'agak', 'sempit', 'aja', 'keseluruhan', 'restoran', 'ini', 'udah', 'mantap']

3.2.2.5 Stopword Removal

Stopword Removal berfungsi untuk menghilangkan kata *stopword* atau kata umum yang tidak memberikan informasi penting yang terdapat pada kalimat. Contoh dari *stopword* antara lain Saya, dan, di, dan lainnya. Kumpulan *Stopword* yang akan dihilangkan diperoleh dari library NLTK, lalu dilakukan penyesuaian berdasarkan kondisi *dataset*.

Tabel 3. 4 *Input dan output proses Stopword Removal*

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['makanannya', 'enak', 'harganya', 'murah', 'pegawainya', 'juga', 'ramah', 'kurangnya', 'tempat', 'agak', 'sempit', 'aja', 'keseluruhan', 'restoran', 'ini', 'udah', 'mantap', 'lah']	['makanannya', 'enak', 'harganya', 'murah', 'pegawainya', 'ramah', 'kurangnya', 'tempat', 'agak', 'sempit', 'keseluruhan', 'restoran', 'udah', 'mantap',]

3.2.2.6 Stemming

Pada tahap Stemming, dilakukan pemangkasan imbuhan pada kata. Kata berimbuhan dipangkas dan diubah menjadi kata dasarnya. Sebagai contoh, kata “menggoreng” menjadi “goreng”, berhitung menjadi hitung. Tujuan dari tahap ini adalah untuk mengurangi pengulangan kata, kata yang jamak, dan meningkatkan akurasi dalam proses penelitian. Tahap stemming ini menggunakan stemmer berbahasa Indonesia oleh library PySastrawi.

Tabel 3. 5 *Input dan output proses Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['makanannya', 'enak', 'harganya', 'murah', 'pegawainya', 'juga', 'ramah', 'kurangnya', 'tempat', 'agak', 'sempit',	['makanan', 'enak', 'harga', 'murah', 'pegawai', 'ramah', 'kurang', 'tempat', 'agak', 'sempit
Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
'keseluruhan', 'restoran', 'udah', 'mantap']	'sempit', 'seluruh', 'restoran', 'udah', 'mantap']

3.2.2.7 Data Normalization

Pada tahap Data Normalization, dilakukan proses dimana data teks di normalisasi, dengan cara mengubah teks kata typo atau salah ketik ataupun singkatan menjadi kata aslinya. Tujuan tahap normalisasi ini agar keakuratan dari data meningkat.

Tabel 3. 6 *Input dan output proses Data Normalization*

Sebelum <i>Data Normalization</i>	Sesudah <i>Data Normalization</i>
['makanan', 'enak', 'harga', 'murah', 'pegawai', 'ramah', 'kur	['makanan', 'enak', 'harga', 'murah', 'pegawai', 'ramah', 'kur

ang', 'tempat' , 'sempit', 'seluruh', 'restoran', 'udah', 'mantap']	ang', 'tempat', 'agak', 'sempit', 'seluruh', 'restoran', 'sudah', 'mantap',]
---	--

3.2.3 Feature Extraction

Pada tahap *Feature Extraction*, data yang sudah di proses sebelumnya di ekstrak fitur-fitur yang ada di dalam nya untuk mendapatkan representasi fitur. Representasi fitur ini selanjutnya digunakan sebagai *input* oleh model.

3.2.3.1 Bag of Words

Bag of Words berfungsi untuk merepresentasikan teks berbentuk data teks kedalam bentuk vektor. Bentuk vektor ini yang selanjutnya akan digunakan oleh model dalam proses. *Bag of Words* digunakan karena penggunaannya simpel, fleksibel, komputasi nya ringan serta output yang dihasilkan sesuai dengan kebutuhan model. Fungsi *Bag of Words* yang digunakan pada penelitian ini menggunakan library `CountVectorizer()` yang disediakan oleh Scikit-learn. `CountVectorizer` mengimplementasikan fungsi *Bag of Words* untuk mengubah kumpulan kata ataupun teks kedalam matrix dari token atau penggalan kata. `CountVectorizer` mengambil *corpus* dari teks dokumen sebagai input dan menghasilkan matrix dimana berisi baris-baris yang berkorespondensi dengan kata unik di corpus.

`CountVectorizer()` memiliki fungsi bawaan untuk mentokenisasi kalimat menjadi bentuk pecahan kata atau token. Token yang sudah dihasilkan yang bersifat unik (hanya satu dan berbeda satu sama lain) dari corpus disimpan kedalam bentuk *dictionary*. Pengimplementasiannya dapat dilihat di tabel 3.7.

Tabel 3. 7 Gambaran variabel *Dictionary*

Kalimat	<i>Dictionary</i>
Makanan disini gurih dan enak!	{‘makanan’. ‘disini’, ‘gurih’, ‘dan’, ‘enak’, ‘biasa’, ‘saja’, ‘hambar’, ‘tidak’}
Makanan disini biasa saja.	
Makanan disini hambar dan tidak enak!	

`CountVectorizer` selanjutnya melakukan *encode* terhadap *corpus* di dictionary tersebut sesuai dengan jumlah kemunculan kata. Setiap dokumen direpresentasikan kedalam

baris untuk matrix yang di hasilkan, dan direpresentasikan ke dalam kolom untuk kata yang unik di dalam *dictionary*. Pengimplementasiannya dapat dilihat di tabel 3.8.

Tabel 3. 8 Gambaran variabel *Dictionary* dan Vektor (setelah dihilangkan *stop words*)

Kalimat	<i>Corpus</i>								
	makanan	disini	gurih	dan	enak	biasa	saja	hambar	tidak
Makanan disini gurih dan enak!	1	1	1	1	1	0	0	0	0
Makanan disini biasa saja.	1	1	0	0	0	1	1	0	0
Makanan disini hambar dan tidak enak!	1	1	0	1	1	0	0	1	1

Berdasarkan tabel 3.8 , diperoleh vektor kalimat sebagai berikut yang dapat dilihat di tabel 3.9.

Tabel 3. 9 Gambaran variabel *Vektor* (setelah dihilangkan *stop words*)

Kalimat	Vektor
Makanan disini gurih dan enak!	[1, 1, 1, 1, 1, 0, 0, 0, 0]
Makanan disini biasa saja.	[1, 1, 0, 0, 0, 1, 1, 0, 0]
Makanan disini hambar dan tidak enak!	[1, 1, 0, 1, 1, 0, 0, 1, 1]

Tabel 3. 10 Contoh Dataset sebelum diubah menjadi bentuk Vektor

No	Dataset
1	"Menu paling rekomended yaitu mie Bangladesh nya. Enak banget banget, selalu rame pengunjung, tiap hari ga pernah sepi. Tempat nya juga luas. Harga

	nya juga bersahabat. Sampe2 kami nambah 1 porsi lagi karna saking enak nya. Top deh kapan2 makan disini lagi "
2	"Penasaran dengan Mie Bangladeshnya, untuk rasa masih okelah, teh susunya enak, hari biasa tetap ramai, pelayanan cepat. toilet bersih"
3	"Buat kalangan orang Medan khususnya anak muda yang hobi kulineran tidak Asing lagi kalau di sebut Agem. Menu favorit di Agem itu Indomie Banglades becek. Sering padat dan antre tempatnya tapi banyak juga yang nungguin karena seenak itu."
4	"Fotonya mungkin rada unappetizing, tapi ini asli enak banget mie banglades nya. Rasanya ada rasa karinya gitu. Beli lewat go food gara2 sampe hotel uda tepar, tapi tetep endeus rasanya. Harga juga terjangkau."
5	"Kalau teringatnya cari makan malam Indomie, salah satunya warkop agem, apalagi buat ketemu kawan lama buat ngobrol Asik sih."

Tabel 3. 11 Contoh Dataset setelah diubah menjadi bentuk Vektor

No	Vektor
1	[000000002010000100000010012000010001100000 200110010000200110001100100400 010011100100 11001000110000010000001001000010]
2	[000000000000100001100100100100000000010000 0000000000000000000010010000001000110000101 0000000000000010100000100010001000]
3	[02110010010011000010010020001000000001010 211100100001110000001101000100110000000000 000000011001001001000000100000002]
4	[100000011100001000000000000110110111100101 010000010000000100000010100010000000001001 200010000000002000101010000110000]

5	[0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 1 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0]
---	---

3.2.4 Resampling

Pada tahap *exploratory data analysis*, ditemukan bahwa jumlah kelas pada dataset tidak seimbang. Maka dari itu, perlu dilakukan *resampling* guna mengatasi ketidakseimbangan yang terjadi dimana jumlah kelas disesuaikan rasionya, antara data mayoritas dengan minoritas, agar seimbang. Dilakukan *oversampling*, penambahan jumlah kelas minoritas serta *undersampling*, pengidentifikasian kelas mayoritas dan minoritas yang berdekatan (*noise* dan *borderline*) lalu menghapus pasangan kelas tersebut, terhadap dataset. Dengan teknik SMOTE tepatnya *Borderline SMOTE* sebagai teknik *oversampling* dan teknik Tomek Links sebagai teknik *undersampling*.

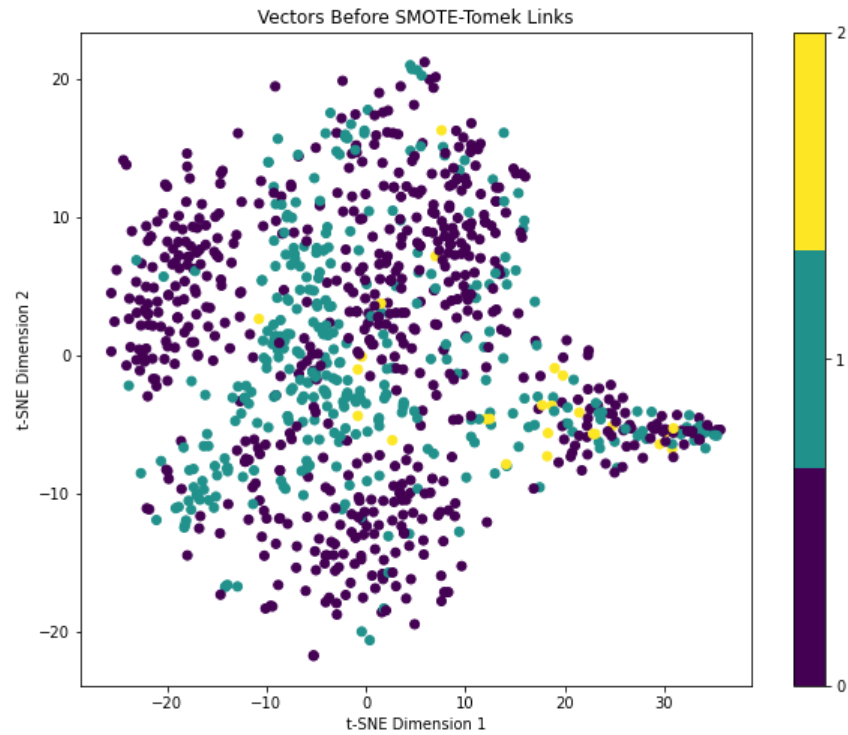
Skenario yang dilakukan terhadap dataset adalah dengan melakukan Tomek Links terlebih dahulu, dilanjutkan dengan *Borderline SMOTE*. Hal ini dilakukan untuk mendapatkan data yang bersifat “bersih” dan baik terlebih dahulu. Penerapan Tomek Links di awal dilanjutkan dengan *Borderline SMOTE* dilakukan untuk mendapatkan data minoritas lebih banyak dengan menggunakan teknik sintetis. Dengan ini, diharapkan jumlah dataset seimbang dengan kondisi yang bersih.

Hasil dari skenario yang dilakukan menghasilkan dataset yang ditampilkan pada gambar

```
Bentuk x_resampled setelah Tomek Links: (986, 1440)
Bentuk y_resampled setelah Tomek Links: (986,)
Bentuk x_resampled setelah SMOTE: (1872, 1440)
Bentuk y_resampled setelah SMOTE: (1872,)
Positif      624
Netral       624
Negatif      624
Name: makanan, dtype: int64
```

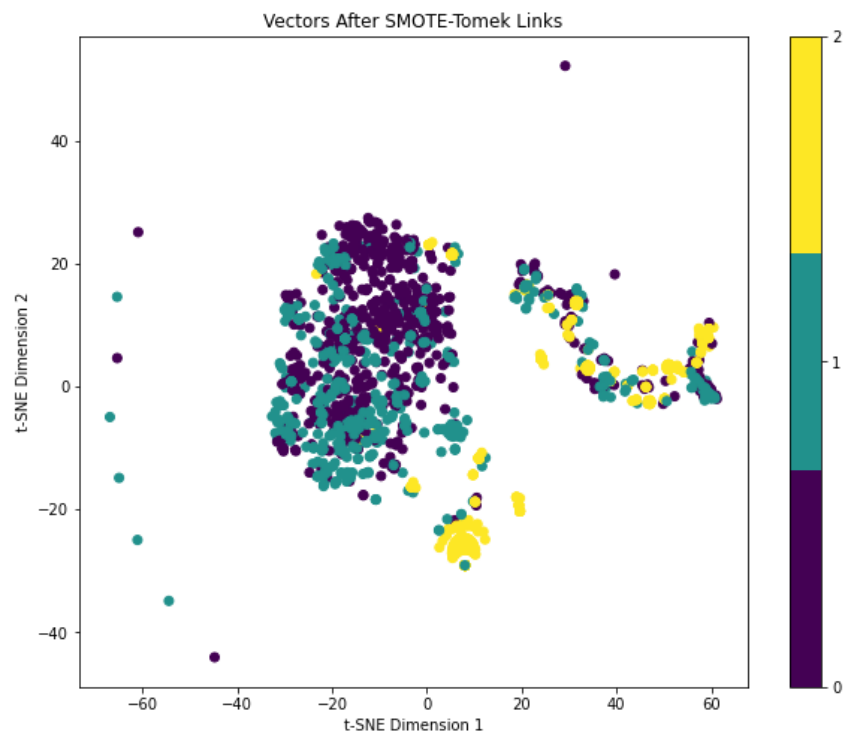
Gambar 3. 7 Jumlah data sebelum dan sesudah proses Tomek Links dan *Borderline SMOTE* pada aspek Makanan

Sumber : Dokumentasi Pribadi



Gambar 3. 8 Persebaran Jumlah Data Aspek Makanan sebelum *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi



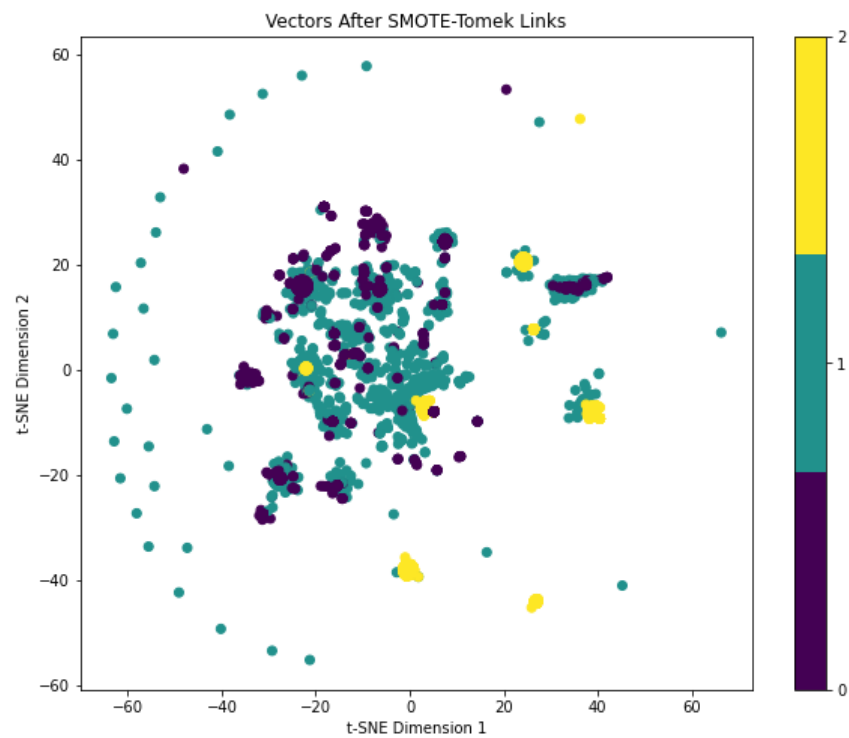
Gambar 3. 9 Persebaran Jumlah Data Aspek Makanan sesudah *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi

```
Bentuk x_resampled setelah Tomek Links: (990, 1440)
Bentuk y_resampled setelah Tomek Links: (990,)
Bentuk x_resampled setelah SMOTE: (2577, 1440)
Bentuk y_resampled setelah SMOTE: (2577,)
Netral      859
Positif     859
Negatif     859
Name: minuman, dtype: int64
```

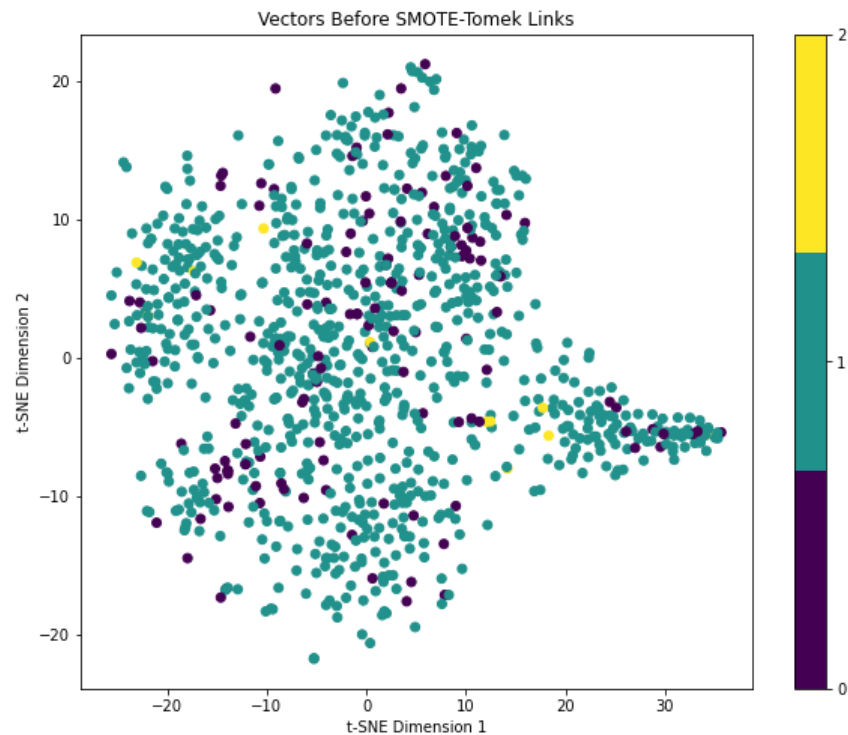
Gambar 3. 10 Jumlah data sebelum dan sesudah proses Tomek Links dan *Borderline* SMOTE pada aspek Minuman

Sumber : Dokumentasi Pribadi



Gambar 3. 11 Persebaran Jumlah Data Aspek Minuman sebelum *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi



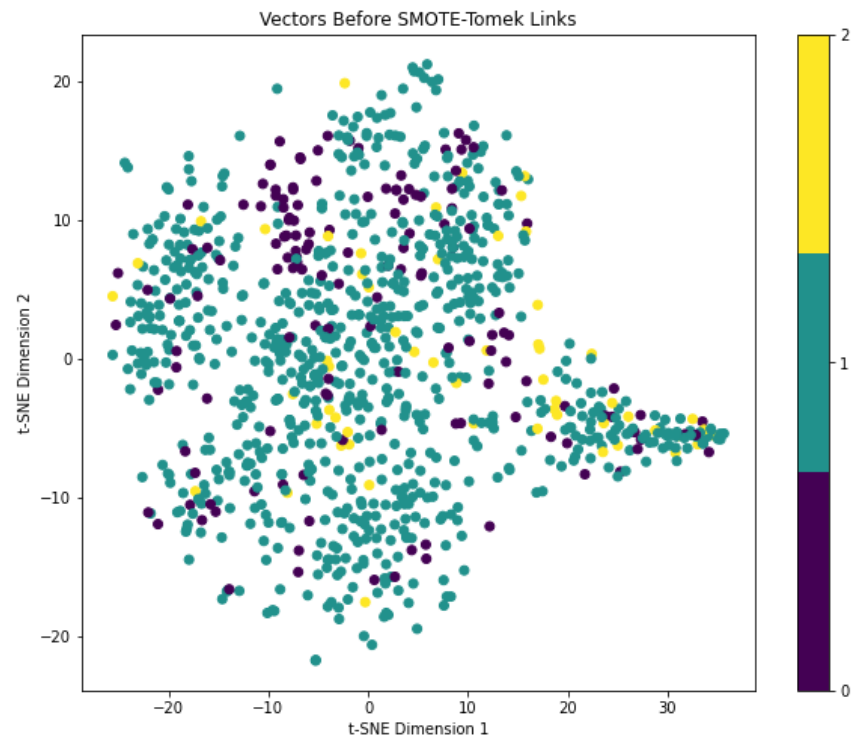
Gambar 3. 12 Persebaran Jumlah Data Aspek Minuman sesudah *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi

```
Bentuk x_resampled setelah Tomek Links: (995, 1440)
Bentuk y_resampled setelah Tomek Links: (995,)
Bentuk x_resampled setelah SMOTE: (2397, 1440)
Bentuk y_resampled setelah SMOTE: (2397,)
Netral      799
Positif     799
Negatif     799
Name: pelayanan, dtype: int64
```

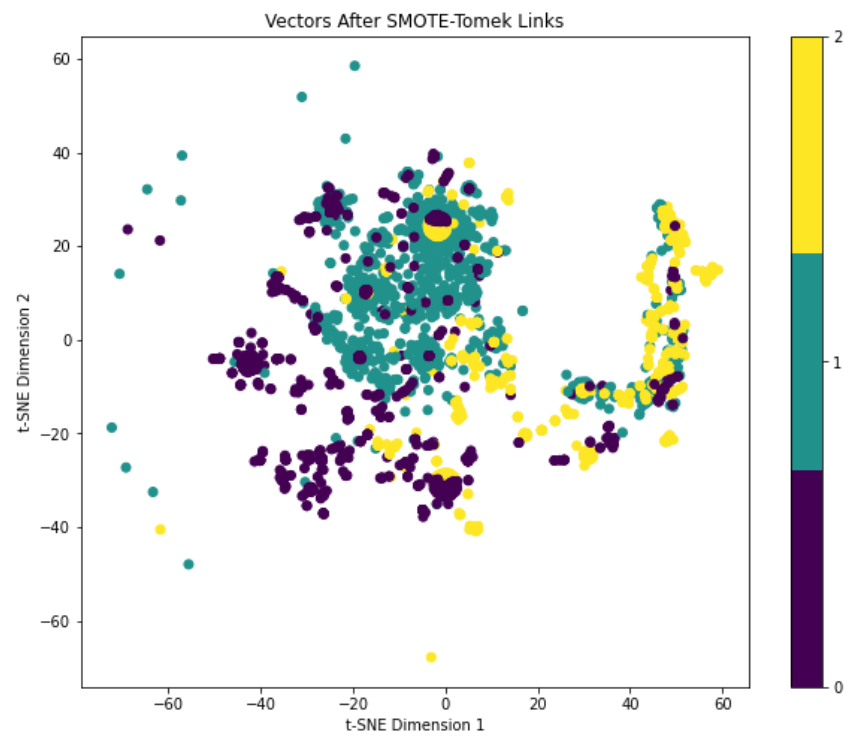
Gambar 3. 13 Jumlah data sebelum dan sesudah proses Tomek Links dan *Borderline* SMOTE pada aspek Pelayanan

Sumber : Dokumentasi Pribadi



Gambar 3. 14 Persebaran Jumlah Data Aspek Pelayanan sebelum *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi



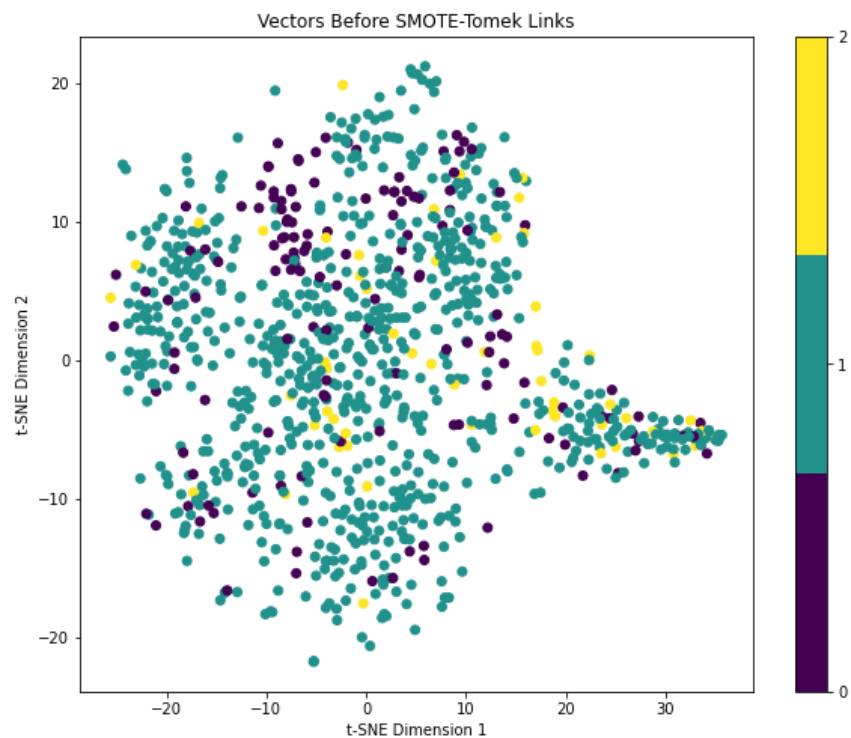
Gambar 3. 15 Persebaran Jumlah Data Aspek Pelayanan sesudah *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi

```
Bentuk x_resampled setelah Tomek Links: (982, 1440)
Bentuk y_resampled setelah Tomek Links: (982,)
Bentuk x_resampled setelah SMOTE: (1950, 1440)
Bentuk y_resampled setelah SMOTE: (1950,)
Positif      650
Negatif      650
Netral       650
Name: tempat, dtype: int64
```

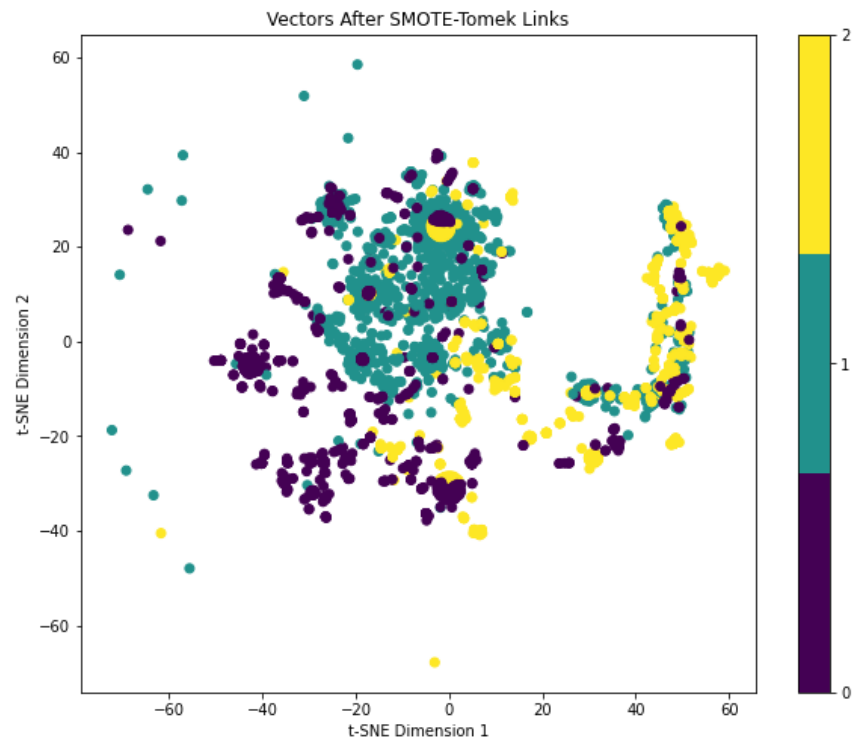
Gambar 3. 16 Jumlah data sebelum dan sesudah proses Tomek Links dan *Borderline* SMOTE pada aspek Tempat

Sumber : Dokumentasi Pribadi



Gambar 3. 17 Persebaran Jumlah Data Aspek Tempat sebelum *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi



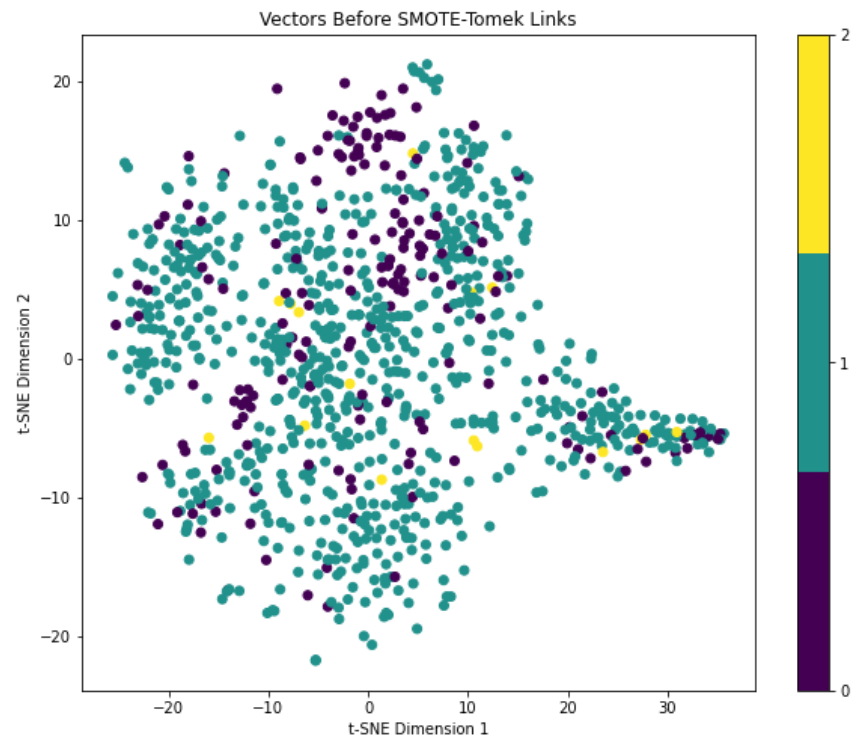
Gambar 3. 18 Persebaran Jumlah Data Aspek Tempat sesudah *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi

```
Bentuk x_resampled setelah Tomek Links: (994, 1440)
Bentuk y_resampled setelah Tomek Links: (994,)
Bentuk x_resampled setelah SMOTE: (2376, 1440)
Bentuk y_resampled setelah SMOTE: (2376,)
Positif      792
Netral       792
Negatif      792
Name: harga, dtype: int64
```

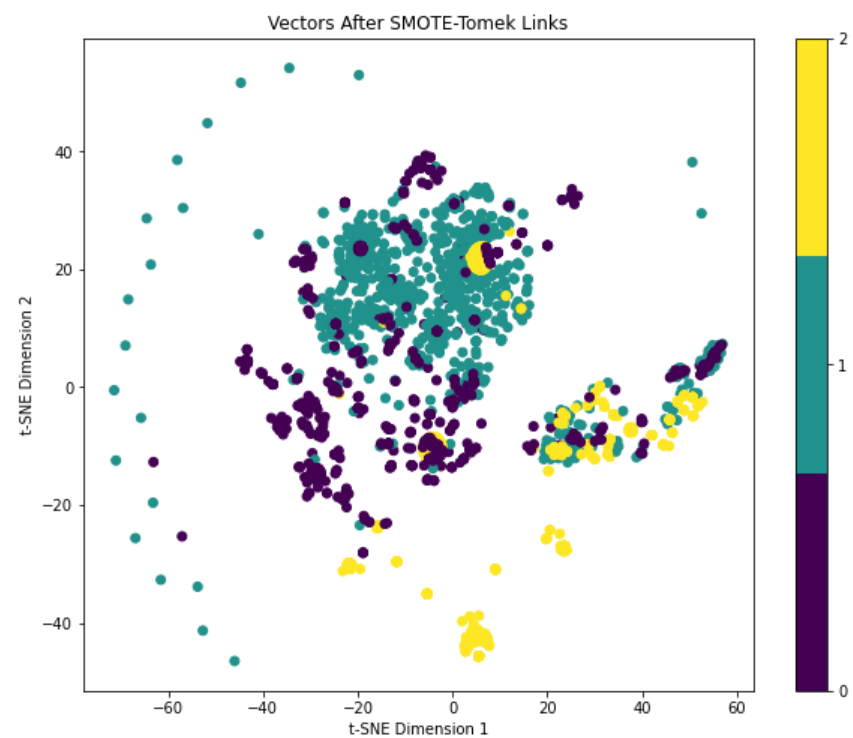
Gambar 3. 19 Jumlah data sebelum dan sesudah proses Tomek Links dan *Borderline* SMOTE pada aspek Harga

Sumber : Dokumentasi Pribadi



Gambar 3. 20 Persebaran Jumlah Data Aspek Harga sebelum *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi



Gambar 3. 21 Persebaran Jumlah Data Aspek Harga sesudah *resampling* dengan visualisasi t-SNE

Sumber : Dokumentasi Pribadi

Berdasarkan plot yang ditampilkan pada gambar 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14, 3.15, 3.16, 3.17, 3.18, 3.19, 3.20, dan 3.21, dapat dilihat perbandingan kerapatan atau kepadatan antar satu data dengan data lainnya yang ditampilkan menggunakan metode t-SNE atau *t-distributed stochastic neighbor embedding*.

t-SNE (*t-distributed Stochastic Neighbor Embedding*) merupakan teknik visualisasi yang digunakan untuk merepresentasikan data berdimensi tinggi dalam plot dua dimensi. Setiap titik dalam plot tersebut menggambarkan satu contoh (seperti dokumen teks atau gambar) pada data. Dalam plot t-SNE, dimensi pertama (Dimensi t-SNE 1) dan dimensi kedua (Dimensi t-SNE 2) diwakilkan oleh sumbu x dan y yang diperoleh melalui reduksi dimensi. Dimensi-dimensi ini dibentuk sedemikian rupa dengan tujuan mempertahankan informasi tentang jarak antar contoh dalam ruang aslinya sebanyak mungkin, sehingga contoh yang terpisah dalam ruang berdimensi tinggi kemungkinan besar akan tetap terpisah dalam plot t-SNE. Plot t-SNE memberikan gambaran visual tentang bagaimana contoh dalam data diatur dalam ruang berdimensi tinggi. t-SNE (*t-distributed stochastic neighbor embedding*) adalah metode yang sering juga digunakan untuk menganalisis kemiripan atau pola dalam data.

Contoh pada plot atau diagram aspek Makanan sebelum dilakukan *resampling* dapat dilihat bahwa kerapatan atau kepadatan antar data cukup renggang dengan berkisar antara -20 sampai 30 untuk t-SNE Dimensi 1 & -20 sampai 20 untuk t-SNE Dimensi 2.

Sedangkan pada plot atau diagram sesudah dilakukan *resampling*, kerapatan atau kepadatan antara data menjadi lebih rapat dan padat, dimana mengalami penambahan jumlah data, serta memiliki rentang yang lebih besar, antara -60 sampai 60 untuk t-SNE Dimensi 1 dan -40 sampai 40 untuk t-SNE Dimensi 2

Hal ini menunjukkan bahwa teknik SMOTE-Tomek Links berhasil melakukan *resampling* data dengan representasi yang serupa ditandai dengan rapat atau padatnya jarak antara titik satu data dengan data lainnya.

3.2.5 Classification

Pada tahap *Classification*, dilakukan proses klasifikasi terhadap *input* yang dimasukkan yaitu berupa kalimat beserta vektornya menggunakan metode *Support Vector Machine* atau *Support Vector Machine*.

3.2.5.1 Support Vector Machine

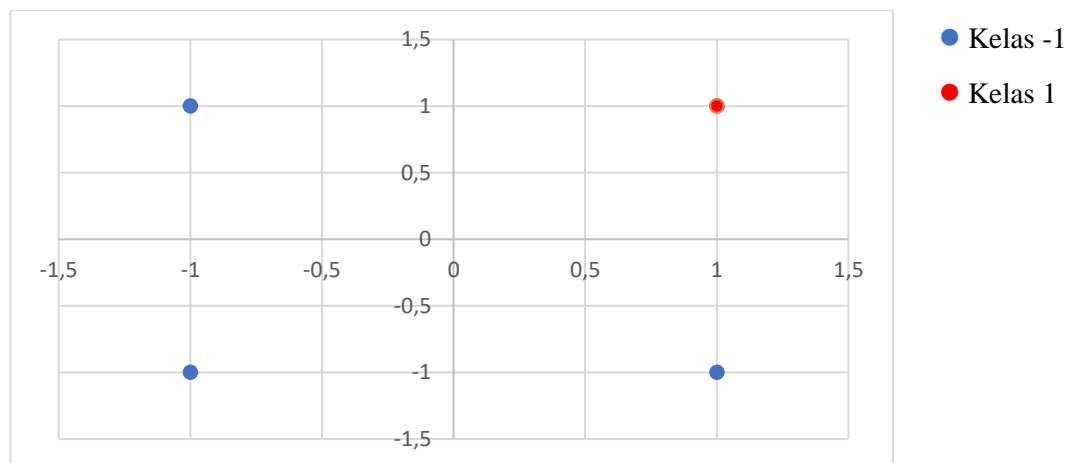
Input berupa vektor selanjutnya akan digunakan untuk mencari nilai kelas +1 (positif) dan kelas -1 (negatif) dengan menggunakan rumus persamaan (1). Setelah kelas diketahui, selanjutnya akan dicari dua data dari kelas yang berbeda yang jaraknya paling dekat atau *support vector*. Dari dua data tersebut, akan dicari nilai *hyperplane* atau garis pemisah dengan menggunakan rumus persamaan (2). Terakhir, akan dicari jarak maksimum dari *support vector* dan *hyperplane* atau *margin* menggunakan persamaan (5) untuk mendapatkan nilai *hyperplane* terbaik dari dua data tersebut.

Contoh Studi Kasus penerapan *Support Vector Machine* :

Tabel 3. 12 Tabel koordinat studi kasus

X1	X2	Y
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

Berdasarkan tabel diatas, didapatkan plot sebagai berikut



Gambar 3. 22 Diagram Koordinat Studi Kasus

Sumber : Dokumentasi Pribadi

Terdapat 2 kelas yaitu kelas -1 yang ditandai oleh titik biru dan kelas 1 yang ditandai dengan titik merah. Studi kasus ini berjenis *Support Vector Machine* Linear

dikarenakan ada dua fitur (x_1 dan x_2), maka w juga akan memiliki 2 fitur (w_1 dan w_2).
Formulasi penyelesaian kasus yang digunakan sebagai berikut :

Meminimalkan nilai margin :

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (3.1)$$

Syarat :

$$y_i(w \cdot x_i + b) \geq 1, \quad i=1,2,3 \dots, N$$

$$y_i(w_1 \cdot x_1 + w_2 \cdot x_2 + b) \geq 1$$

Sehingga didapatkan beberapa persamaan berikut :

1. $(w_1 + w_2 + b) \geq 1$ untuk $y_1 = 1, x_1 = 1, x_2 = 1$
2. $(-w_1 + w_2 + b) \geq 1$ untuk $y_2 = -1, x_1 = 1, x_2 = -1$
3. $(w_1 - w_2 - b) \geq 1$ untuk $y_3 = -1, x_1 = -1, x_2 = 1$
4. $(w_1 + w_2 - b) \geq 1$ untuk $y_4 = -1, x_1 = -1, x_2 = -1$

Didapatkan beberaa persamaan berikut :

1. $(w_1 + w_2 + b) \geq 1$
2. $(-w_1 + w_2 - b) \geq 1$
3. $(w_1 - w_2 - b) \geq 1$
4. $(w_1 + w_2 - b) \geq 1$

Menjumlahkan persamaan (1) dan (2)

$$(w_1 + w_2 + b) \geq 1$$

$$(-w_1 + w_2 - b) \geq 1$$

----- +

$$2w_2 = 2$$

$$\text{Maka } w_2 = 1$$

Menjumlahkan persamaan (1) dan (3)

$$(w_1 + w_2 + b) \geq 1$$

$$(w_1 - w_2 - b) \geq 1$$

----- +

$$2w_2 = 2$$

$$\text{Maka } w_2 = 1$$

Menjumlahkan persamaan (2) dan (3)

$$(-w_1 + w_2 - b) \geq 1$$

$$(w_1 - w_2 - b) \geq 1$$

----- +

$$-2b = 2$$

$$\text{Maka } b = -1$$

Sehingga didapatkan persamaan *hyperplane* atau persamaan 5 :

$$w_1x_1 + w_2x_2 = b = 0$$

$$x_1 + x_2 - 1 = 0$$

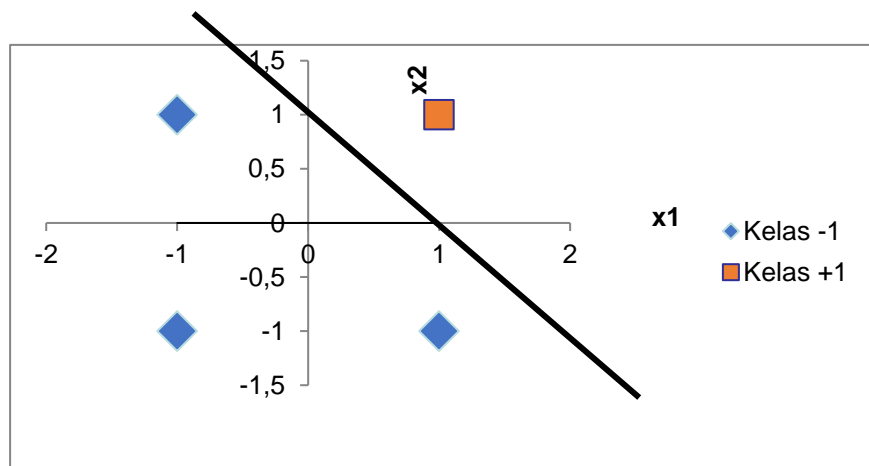
$$x_2 = 1 - x_1$$

Visualisasi garis *hyperplane* (sebagai fungsi klasifikasi) :

$$w_1x_1 + w_2x_2 + b = 0$$

$$x_1 + x_2 - 1 = 0$$

$$x_2 = 1 - x_1$$



Gambar 3. 13 Visualisasi Plot dalam mendapatkan Hyperplane
Sumber : Dokumentasi Pribadi

3.2.6 Validation

Pada tahap Validation, dilakukan proses validasi data menggunakan metode *K-Fold Cross Validation*

3.2.6.1 *K-Fold Cross Validation*

Dataset berjumlah 1000 berupa kalimat *review*/ulasan yang sudah mengalami proses *feature extraction* sehingga menghasilkan *output* berupa review itu sendiri beserta vektor nya akan dilakukan validasi kembali menggunakan metode *K-Fold Cross Validation* guna mencegah kondisi *overfitting* pada model. Proses validasi menggunakan metode K-Fold Cross Validation dilakukan dengan cara membagi dataset menjadi sekian bagian atau $K=N$ bagian. Pada sistem ini, digunakan $K=N=10$ atau 10 fold. Sehingga, dataset berjumlah 1000 tersebut dibagi menjadi 10 bagian atau masing-masing bagian berjumlah 100 data review. Dari 1000 data total yang terdapat pada dataset, 100 data review digunakan sebagai data testing dan 900 data lainnya menjadi data training. Proses dimulai dari K-1 atau Fold 1. Proses training menggunakan algoritma *Support Vector Machine* dilakukan dengan dengan K-1 berjumlah 100 data dari data 1 sampai dengan data ke 100 sebagai *data testing* dan 900 data lainnya digunakan sebagai *data training*. Dilanjutkan dengan K-2 atau Fold 2 sebagai *data testing* dengan jumlah yang sama yaitu 100 data, dengan ketentuan data ke 101 sampai dengan data ke 200 dengan 900 data lainnya digunakan sebagai data training. Lalu K-3 data atau Fold 3 sebagai data testing dengan jumlah yang sama yaitu 100 data, dengan ketentuan data ke 201 sampai dengan data ke 300 dengan 900 data lainnya digunakan sebagai data training. Proses ini diulangi sampai dengan K-10 atau Fold 10. *Output* yang dihasilkan dari proses validasi dengan metode K-Fold Cross Validation adalah berupa nilai skor akurasi dari algoritma yang digunakan, dimana pada sistem ini adalah *Support Vector Machine*, berjumlah $K=N$ yang pada sistem ini berjumlah 10. 10 skor ini di rata-ratakan sehingga menghasilkan rata-rata skor yang dapat dijadikan sebagai acuan apakah dataset yang digunakan baik atau tidak.

3.2.7 Model

3.2.7.1 *Pelatihan Model Support Vector Machine*

Setelah sekian tahap yang dilakukan mulai dari tahap *preprocessing* dari dataset yang digunakan sebagai data *training* dan *validation*, tahap *feature extraction* yang menghasilkan vektor dari dataset, tahapan dilanjutkan ke tahap *training* model Support

Vector Machine yang akan digunakan untuk proses klasifikasi dan *sentiment analysis*, serta tahap *validation*, dimana dilakukan proses validasi terhadap dataset apakah baik atau tidak. Pada proses *training* ini, digunakan kernel *Radial Basis Function* atau RBF serta menerapkan pendekatan “One vs One” atau “All vs All”. Penggunaan kernel RBF didasari dari penelitian oleh Cho et al. pada tahun 2017. Dalam penelitian tersebut, ditunjukkan bahwa kernel RBF menghasilkan nilai keseluruhan paling baik dan diikuti oleh kernel *Polynomial* dalam mengklasifikasikan data berbentuk *multiclass*. Dalam pendekatan ini, model dibuat secara terpisah berdasarkan jumlah setiap kelas yang ada pada dataset. Beberapa keuntungan dari pendekatan ini antara lain mudah dipahami diimplementasikan, dikarenakan hanya perlu melatih model berdasarkan kelas. Keuntungan lainnya adalah lebih akurat, karena setiap model hanya harus memutuskan antara dua kelas, sehingga model lebih mudah mencapai akurasi yang tinggi dalam pemisahan atau mengklasifikasikan kelas. Pada penelitian ini, terdapat 5 kelas yaitu kelas makanan, minuman, pelayanan, tempat, dan harga. Dengan demikian, proses training menghasilkan 5 buah model Support Vector Machine yang digunakan untuk melakukan proses sentimen ulasan rumah makan berdasarkan nama rumah makan yang diinput oleh user yang ulasannya diperoleh dari proses scraping pada platform Google Maps secara *real-time*. Proses pelatihan juga menggunakan beberapa nilai parameter C atau parameter *Cost* dan nilai parameter gamma guna mendapatkan model dengan kinerja paling optimal dalam melakukan klasifikasi, dengan nilai parameter C nya antara lain 0.01, 0.05, 0.25, 0.5, 0.75, 1, 10, 100, serta nilai parameter gamma (γ) nya antara lain 0.01, 0.1, 1, 10. Nilai-nilai tersebut diperoleh berdasarkan penelitian yang dilakukan Prangga pada 2017, dimana digunakan nilai-nilai dengan rentang tersebut, dengan sedikit penyesuaian yang penulis lakukan, yang memungkinkan untuk digunakan pada penelitian ini. Nilai-nilai tersebut didasarkan pada penelitian-penelitian sebelumnya yaitu pada penelitian Huang, Hung, Lee, Li, dan Jiang menggunakan $C = \{10, 50, 100\}$, $\gamma = \{2.4, 5, 10\}$ dan $C = \{5, 10, 50\}$, $\gamma = \{0.08, 4, 11\}$ kemudian pada penelitian Erfanifard, Behnia dan Moosavi menggunakan $C = \{100, 200, 300\}$ dan $\gamma = \{0.2, 0.3, 0.4\}$ dan yang terakhir oleh Rusydina menggunakan $C = \{0.25, 0.50, 0.75, 1, 2, 3, 4\}$ dan $\gamma = \{0.005, 0.05, 0.1, 0.15\}$. (Erfanifard et al., 2014; Huang et al., 2014; Rusydina, 2016). Teknik yang dapat mencari kombinasi terbaik dari parameter C dan γ disebut dengan Grid Search, yang mana digunakan library GridSearchCV dalam proses

eksekusinya. Hal ini diterapkan untuk melihat nilai parameter C dan gamma mana yang menghasilkan *f1-score* klasifikasi terbaik, yang nantinya nilai parameter C tersebut dipilih untuk model yang akan digunakan pada proses prediksi berdasarkan input user. Model dengan parameter C dan gamma yang menghasilkan *f1-score* terbaik selanjutnya akan disimpan dan digunakan pada proses prediksi berupa review rumah makan berdasarkan input nama rumah makan yang dimasukkan user. Berikut adalah parameter terbaik yang akan ditentukan oleh GridSearchCV :

```
# Define the parameter grid

param_grid = {
    'C': [0.01, 0.05, 0.25 , 0.5, 0.75, 1, 10, 100],
    'gamma': [0.01, 0.1, 1, 10 ],
}

# Define the scorers

scorers = {
    'Akurasi': make_scorer(accuracy_score),
    'Precision': make_scorer(precision_score,
average='macro'),
    'Recall': make_scorer(recall_score,
average='macro'),
    'F1-score': make_scorer(f1_score, average='macro')
}

# Create the GridSearchCV object

grid_search = GridSearchCV(
    SVC(kernel='rbf', class_weight='balanced',
decision_function_shape='ovo', random_state=None),
    param_grid,
    scoring=scorers,
    refit='F1-score',
```

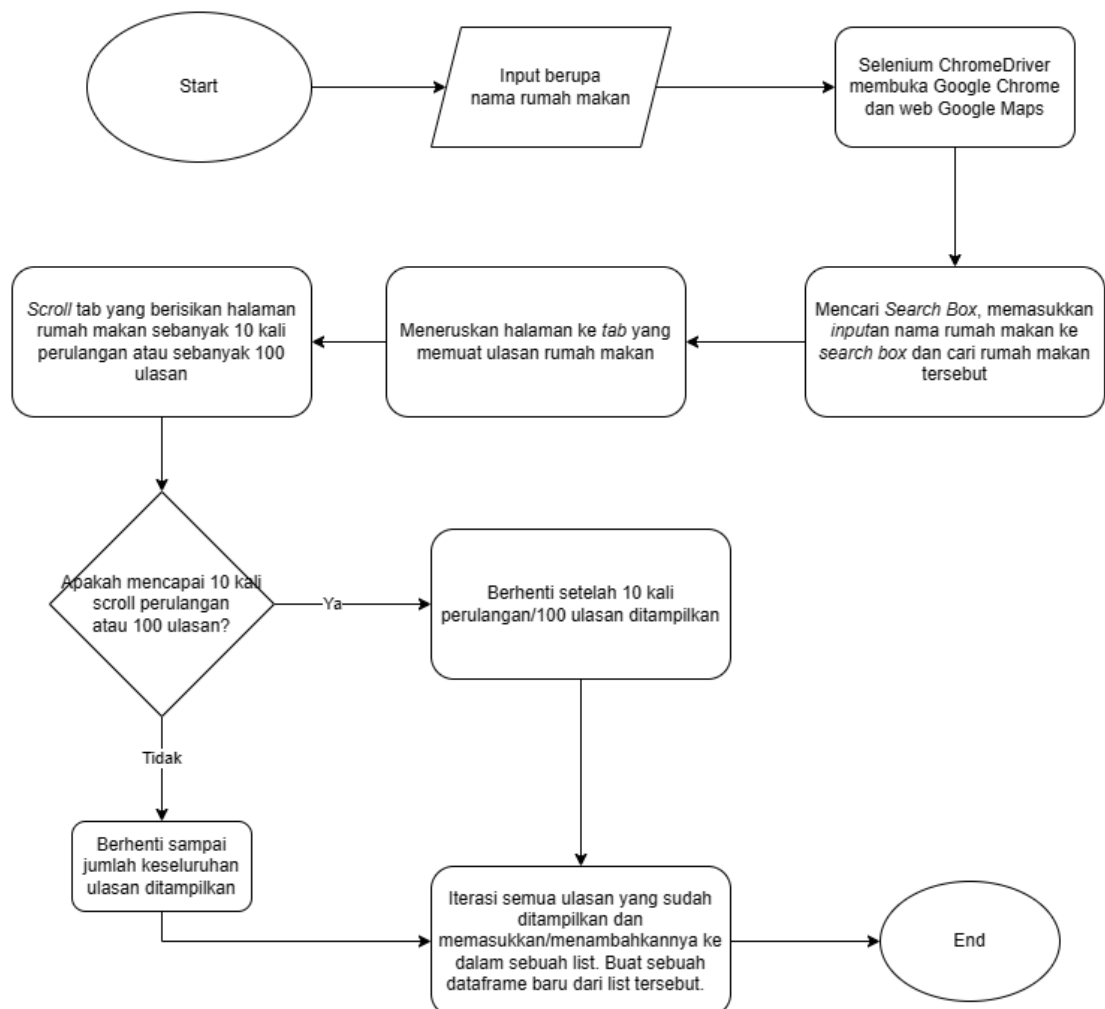
```

        return_train_score=True,
        cv=10
    )

```

3.2.8 Scraping

Pada sistem ini, dilakukan proses *scraping* dengan tujuan mendapatkan ulasan rumah makan berdasarkan nama rumah makan yang diinput oleh *user*. Proses *Scraping* dilakukan pada halaman web Google Maps pada *section* ulasan yang terdapat pada halaman rumah makan. Proses *scraping* dilakukan dengan menggunakan *library* Selenium sebagai *web browser automation* yang digunakan untuk meminta permintaan HTTP, permintaan *scrolling* halaman, dan lainnya.



Gambar 3. 23 Flowchart proses Scraping

Sumber : Dokumentasi Pribadi

Tahap Scraping dimulai ketika *input* berupa nama restoran dimasukkan oleh *user*. Setelah *input* diterima, *browser* dibuka, dalam sistem ini browser Google Chrome, dengan menggunakan library Selenium. *Browser* akan diarahkan ke halaman utama Google Maps. Selanjutnya dilakukan pencarian elemen *search bar* berdasarkan Xpath. Setelah elemen tersebut diperoleh, dilakukan *input* berupa nama rumah makan tersebut, dan melakukan *submit* untuk mencari halaman GoogleMaps rumah makan. Setelah halaman rumah makan tampil, selanjutnya halaman akan diarahkan ke *section* atau *tab* yang berisikan seluruh ulasan dari rumah makan tersebut. Ulasan yang ditampilkan adalah ulasan yang terbaru ditambahkan. Setelah halaman menampilkan ulasan tampilan terbaru, driver akan melakukan *scrolling*. *Scrolling* dilakukan dengan dengan melakukan perulangan. Perulangan ini dapat ditentukan sesuai dengan kebutuhan. Pada sistem ini, perulangan yang dilakukan sebanyak 10 kali. Dalam satu kali proses *scrolling*, *section* memuat sebanyak 10 ulasan. Dengan begitu, dalam 10 kali perulangan, ulasan yang di *scraping* berjumlah 100 ulasan. Setelah semua ulasan dimuat, semua ulasan yang sudah tampil akan di iterasi. Setelah semua ulasan di iterasi, ulasan tersebut akan ditambahkan kedalam sebuah list. *List* yang berisi ulasan tersebut selanjutnya akan diubah ke *data frame* baru yang nantinya akan digunakan dalam proses *sentimen analysis* rumah makan tersebut menggunakan model yang sudah ada.

3.2.9 Output

Output yang dihasilkan pada penelitian ini adalah berupa *dashboard* yang bertujuan untuk membantu *user* dalam mengetahui opini pengunjung lain terhadap suatu rumah yang nantinya dapat dijadikan acuan dalam memilih rumah makan yang akan dikunjungi. *Dashboard* yang ditampilkan terdiri dari visualisasi dalam bentuk *pie chart* yang dibuat berdasarkan *output* sentiment yang sudah dihasilkan pada tahap *sentiment analysis* oleh model sebelumnya, dimana output yang dihasilkan berupa sentiment beserta aspek dengan jumlah ulasan yang di *scraping*. Visualisasi yang ditampilkan adalah berupa diagram atau grafik yang dibuat menggunakan *library* bahasa pemrograman *python* yaitu *matplotlib*. Elemen - elemen yang ditampilkan pada visualisasi antara lain:

- Nama rumah makan yang ulasan nya sudah di *scraping*.
- Jumlah total ulasan rumah makan yang diperoleh.

- Grafik perbandingan sentimen per aspek, berisikan perbandingan sentimen positif, netral, serta negatif dari 5 (lima) aspek yang telah ditentukan yaitu aspek makanan, minuman, pelayanan, bangunan, serta harga berdasarkan dari seluruh ulasan yang sudah di *scraping*.

3.3 Perancangan Sistem

Dalam penelitian ini, dibuat sistem antarmuka dengan beberapa halaman dengan tujuan memudahkan pengguna dalam proses penggunaan sistem. Halaman-halaman tersebut antara lain halaman *home* atau beranda yang juga memuat terdiri dari halaman/bagian *testing* atau input dari pengguna, halaman *training*, serta halaman login yang diakses oleh admin pada saat sebelum proses *training*. Adapun penjelasan dari setiap halaman/bagian yang diterapkan pada sistem adalah sebagai berikut :

- Rancangan Halaman Beranda

Pada halaman beranda, ditampilkan informasi-informasi umum terkait sistem yang dibangun antara lain judul sistem yang menjelaskan sistem yang digunakan user, nama pengembang sistem ini yaitu penulis beserta dengan keterangan tambahan dari penulis, dan form untuk menginput nama rumah makan yang ingin dicari beserta tombol *submit* atau cari. Jika *user* belum melakukan *training* model, *user* akan diminta untuk melakukan training model terlebih dahulu.

Nama Website		Home	User Input	Training
		1	2	3
<input type="text" value="Nama Website"/>	<div>4</div> <input type="text" value="Teks Deskripsi"/>	<div>Gambar</div>		
<input type="button" value="Cari"/>				
<div>5</div>				
Cari Nama Rumah Makan				

Gambar 3. 24 Rancangan halaman Beranda

Sumber : Dokumentasi Pribadi

Keterangan :

1. Tombol Home berfungsi sebagai tombol navigasi atau menampilkan halaman *home* atau beranda.
 2. Tombol User Input berfungsi sebagai tombol navigasi atau mengarahkan ke *section* atau bagian untuk melakukan *input* nama rumah makan.
 3. Tombol Home berfungsi sebagai tombol navigasi atau menampilkan halaman *training*
 4. Deskripsi mengenai informasi website
 5. Tombol Cari berfungsi untuk mengarahkan mengarahkan ke *section* atau bagian untuk melakukan *input* nama rumah makan.
- Rancangan Halaman Login Admin

Halaman Login Admin hanya dapat diakses oleh admin. Halaman Login akan ditampilkan sebelum proses *training* setelah mengklik tombol navigasi *training data*, dimana proses training tersebut hanya dapat dilakukan oleh admin.

Login Admin

Deskripsi **1**

Username **2**

Password **3**

Login **4**

Gambar 3. 25 Rancangan halaman Login Admin

Sumber : Dokumentasi Pribadi

Keterangan :

1. Deskripsi atau penjelasan mengenai halaman Login.
2. Form Username Admin.
3. Form Password Admin.
4. Tombol Login

- Rancangan Training Model

Setelah admin melakukan Login, admin akan diarahkan ke halaman training model. Pada halaman training model, ditampilkan sebuah form untuk melakukan *input* dataset yang digunakan untuk training model beserta tombol submit.

Nama Website	Home User Input Training
<h2 style="margin: 0;">Training Model</h2> <div style="border: 1px solid black; padding: 10px; margin: 10px auto; width: 80%;"> <p style="text-align: center;">Instruksi</p> <p style="text-align: right; font-size: 24px; font-weight: bold;">2</p> </div> <div style="display: flex; justify-content: center; gap: 10px; margin: 10px auto; width: 60%;"> <div style="border: 1px solid black; padding: 5px 10px;">Upload File</div> <div style="font-size: 24px; font-weight: bold;">1</div> <div style="border: 1px solid black; padding: 5px 10px;">Submit</div> </div> <div style="border: 1px solid black; height: 150px; margin: 10px auto; width: 90%;"></div> <p style="text-align: center; margin-top: 20px;">Output Preprocessing</p>	

Gambar 3. 26 Rancangan halaman Training Model.

Sumber : Dokumentasi Pribadi

Keterangan :

1. Form untuk melakukan upload file.
2. Instruksi dan penjelasan mengenai proses upload file yang akan digunakan untuk melakukan proses training model.

Output yang dihasilkan pada proses training model ini adalah berupa model yang akan digunakan dalam proses sentiment, serta tabel berisi teks data latih model yang telah dilakukan proses *pre processing* dan diagram informasi hasil training.



Gambar 3. 27 Rancangan halaman lanjutan Training Model.

Sumber : Dokumentasi Pribadi

Keterangan :

1. Tabel output proses Preprocessing.
2. Diagram informasi hasil training seperti nilai C, akurasi . *precission*, *recall* dan *f1-score*

- Rancangan Halaman Sentiment

Pada halaman sentiment, ditampilkan beberapa informasi antara lain nama rumah makan yang telah dicari serta di *scraping* ulasannya, jumlah total ulasan yang sudah di *scraping*, tabel ulasan yang sudah di *scraping*, sertasentiment dari ulasan rumah makan yang sudah di *scraping* yang di bagi kedalam 5 (lima) aspek antara lain aspek makanan,

minuman, bangunan, pelayanan dan harga yang disajikan ke dalam bentuk diagram *pie chart*.

The wireframe shows a web page layout. At the top, there is a header bar with two sections: 'Nama Website' on the left and 'Home User Input Training' on the right. Below the header, the main content area is titled 'Cari Nama Rumah Makan'. In the center of this area, there is a large rectangular box labeled 'Panduan 1'. Below this box, there is a form labeled '2'. The form consists of a text input field with the placeholder text 'Masukkan Nama Rumah Makan' and a button labeled 'Cari'.

Gambar 3. 28 Rancangan halaman Home bagian Sentiment.

Sumber : Dokumentasi Pribadi

Keterangan :

1. Panduan cara melakukan proses pencarian berupa input nama rumah makan.
2. Form input nama rumah makan beserta tombol cari.

Nama Website

Home User Input Training

Masukkan Nama Rumah Makan

Cari

Tabel Review yang sudah di ambil

1

Diagram Pie Chart

2

Diagram Pie Chart

Gambar 3. 29 Rancangan halaman Home bagian Sentiment

Sumber : Dokumentasi Pribadi

Keterangan :

1. Tabel berisi ulasan yang telah di *scraping* atau di ambil.
2. Diagram persentase jumlah sentiment dari rumah makan tersebut berdasarkan aspek makanan, minuman, pelayanan, tempat dan harga. Serta jumlah total ulasan yang telah di *scraping* atau di ambil.

3.4 Metode Evaluasi

Untuk mengetahui performa suatu model yang sudah dilatih, apakah model tersebut sudah dapat bekerja dengan baik atau tidak, perlu dilakukan evaluasi. Metode evaluasi yang diterapkan pada penelitian kali ini yaitu menggunakan metode *confussion matrix*. Evaluasi yang diterapkan dalam metode ini ditentukan berdasarkan beberapa nilai antara lain *precision*, *recall*, *F-1 score*, dan *accuracy*. Penerapan metode evaluasi *confussion matrix* dapat dilihat melalui Tabel 3.11 berikut :

Tabel 3. 13 Tabel *Confusion Matrix*

		Label Secara Manual		
		Positif	Negatif	Netral
Hasil Identifikasi	Positif	TP	FP	FP
	Negatif	FN	TN	FN
	Netral	FNet	FNet	TNet

Dengan Keterangan sebagai berikut:

- True Positive (TP) : Data aktual dengan kelas **Positif** diprediksi sebagai kelas **Positif** (sesuai)
- False Positive (FP) : Data aktual dengan kelas **selain Positif** atau **tidak Positif**, tetapi diprediksi sebagai kelas **Positif** (sesuai)
- Positif (tidak sesuai)
- True Negative (TN): Data aktual dengan kelas **Negatif** yang diprediksi sebagai kelas **Negatif** (sesuai)
- False Negative (FN): Data aktual dengan kelas **selain Negatif** atau **tidak Negatif**, tetapi diprediksi sebagai kelas **Negatif** (tidak sesuai)
- True Netral (TNet): Data aktual dengan kelas **Netral**, yang diprediksi sebagai kelas **Netral** (sesuai)
- False Netral (FNet) : Data aktual dengan kelas **selain Netral** atau **tidak Netral**, tetapi diprediksi sebagai kelas **Netral** (tidak sesuai)

Digunakan persamaan yang digunakan dalam menghitung nilai *precision*, *recall*, *f1-score*, dan akurasi. Persamaan nya ditunjukkan pada persamaan 3.2, 3.3, 3.4 dan 3.5 berikut ini :

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (3.2)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3.3)$$

$$F1-score = \frac{2 \times recall \times precision}{recall+precision} = \frac{2TP}{2TP+FN+FP} \times 100 \quad (3.4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3.5)$$

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Implementasi Sistem

Rancangan yang telah di bahas bab 3 mengenai sistem *aspect based sentimen analysis* berikutnya akan diimplementasikan. Perangkat keras dan perangkat lunak yang saling mendukung digunakan dalam pengimplementasian guna menjalankan sistem dengan lancar. Adapun perangkat yang digunakan sebagai berikut :

4.1.1 Spesifikasi Perangkat Keras dan Perangkat Lunak

Pengimplementasian sistem pada penelitian ini menggunakan perangkat keras dengan spesifikasi sebagai berikut:

1. Processor Intel Core i5 7200U 7th Gen
2. Memory RAM sebesar 8 GB
3. SSD sebesar 240 GB

Adapun perangkat lunak yang digunakan dalam pengimplementasian sistem ini memiliki spesifikasi sebagai berikut :

1. Sistem Operasi Windows 10
2. Bahasa Pemrograman Python versi 3.9.0
3. *Library* dalam bahasa pemrograman Python antara lain Pandas, Matplotlib, Seaborn, NLTK, Sastrawi, scikit-learn, Selenium dan Flask.
4. Microsoft Visual Studio Code
5. Jupyter Notebook

4.1.2 Implementasi Perancangan Antarmuka Sistem

Rancangan antarmuka yang sebelumnya telah dibahas di bagian 3.2 diimplementasikan dan dijabarkan sebagai berikut :

1. Tampilan halaman Beranda/*Home*

Implementasi pada tampilan halaman home dibuat berdasarkan rancangan yang telah dijelaskan pada bab 3. Pada halaman *Home*, ditampilkan *header* yang berisi nama web beserta logonya, tombol navigasi menuju halaman *home*, halaman *training data*, dan bagian *user input* yang mengarahkan pengguna ke bagian “Cari Rumah Makan”. Lalu ditampilkan deskripsi terkait website, dan tombol Cari Rumah Makan. Tombol tersebut akan mengarahkan pengguna ke bagian pencarian *input* dari pengguna/*user input*.



Gambar 4. 1 Implementasi Tampilan Beranda

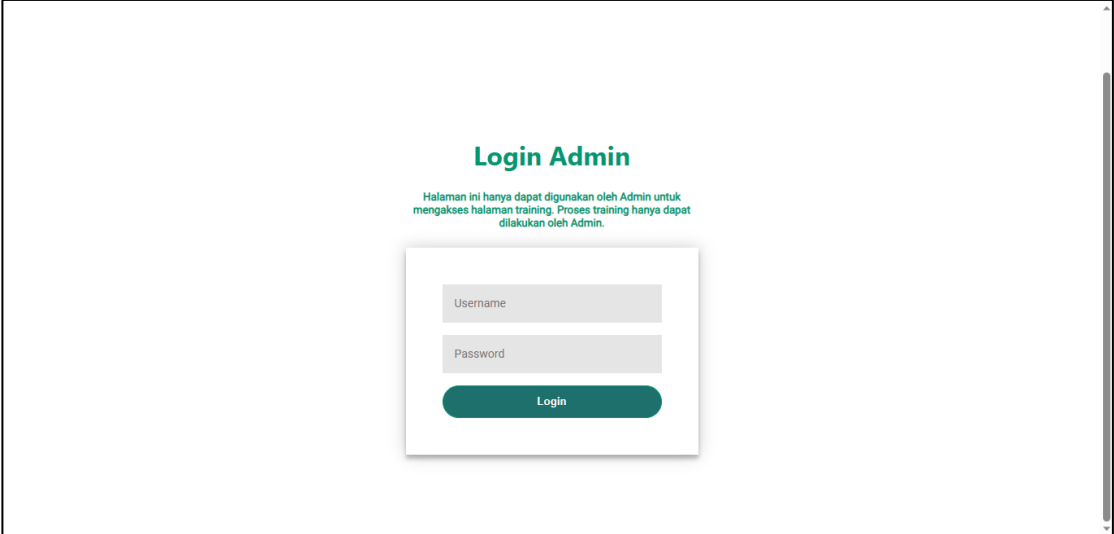
Sumber : Dokumentasi Pribad

Gambar 4. 2 Lanjutan Implementasi Tampilan Beranda

Sumber : Dokumentasi Pribadi

2. Tampilan Training Data

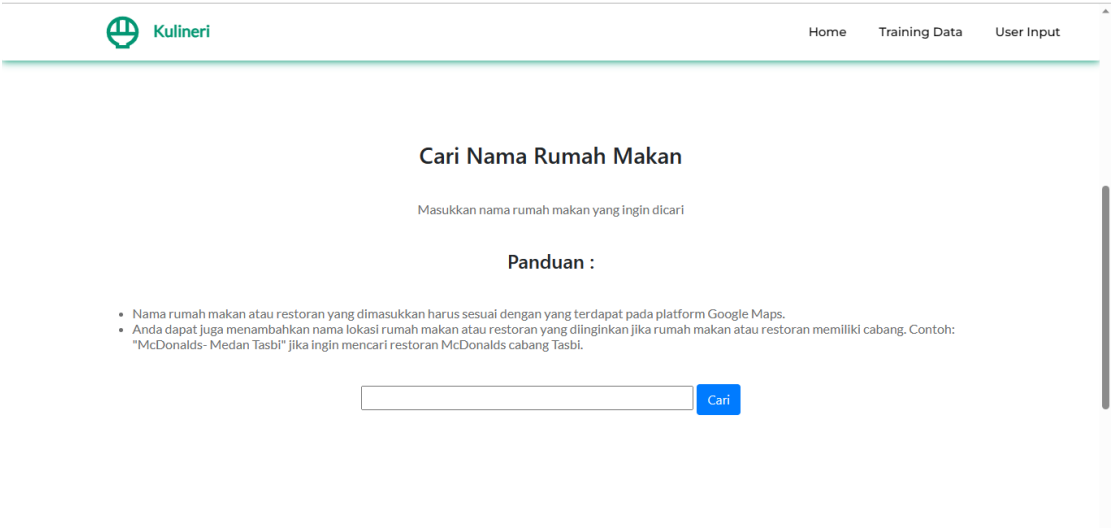
Implementasi pada tampilan halaman home dibuat berdasarkan rancangan yang telah dijelaskan pada bab 3. Pada halaman Training Data, tampilan yang ditampilkan tidak jauh berbeda seperti halaman *home*, yaitu memiliki *header* yang berisi nama web beserta logonya, tombol navigasi menuju halaman *home*, halaman *training data*, bagian *user input* di halaman *home* dan tombol logout. Pada saat tombol navigasi halaman *training data* di klik pada halaman *home*, pengguna akan diarahkan ke halaman *login admin*. *Proses login admin* dan *training data* hanya dapat dilakukan oleh admin.



Gambar 4. 3 Implementasi Tampilan Login Admin

Sumber : Dokumentasi Pribadi

Setelah melakukan *login*, halaman akan menampilkan instruksi dalam melakukan proses *training data*. Terdapat *form* berupa *input file* beserta tombol “*Train Data*” untuk melakukan *training data*. Jika admin sudah menginput *file* sesuai dengan keterangan yang ditampilkan pada instruksi, proses *training data* dapat dilakukan dengan mengklik tombol “*Train Data*”



Gambar 4. 4 Implementasi Tampilan Training Data

Sumber : Dokumentasi Pribadi

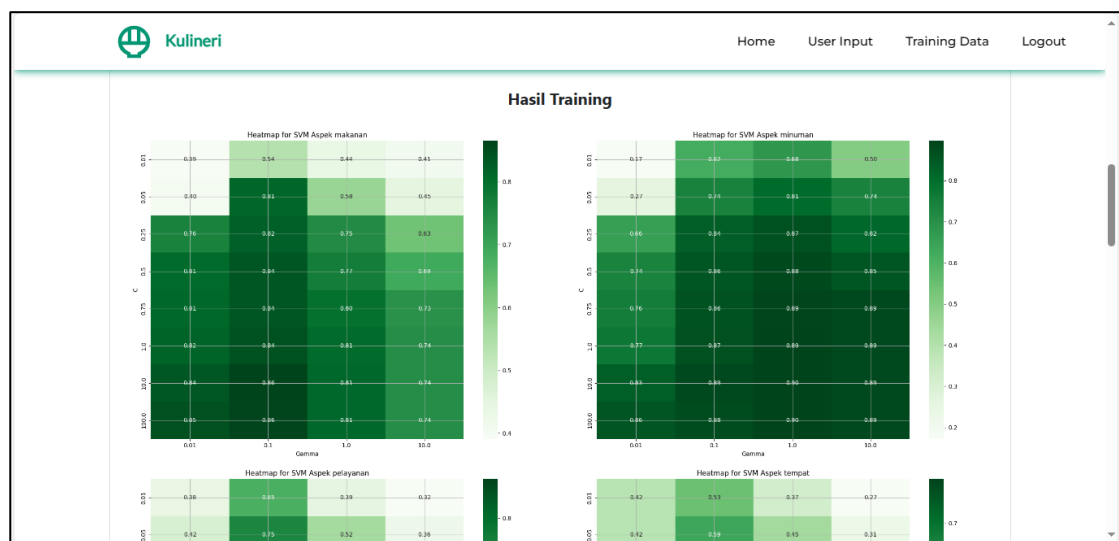
Setelah model diperoleh dari hasil proses training data berdasarkan input data, ditampilkan serangkaian proses *preprocessing* yang dilakukan terhadap salah satu kalimat pada data *text* yang digunakan dalam proses training.

Tahap Pre-Processing	
Pada bagian ini ditampilkan tahapan pre-processing beserta contoh penerapannya pada salah satu kalimat dari data training.	
Input	Penasaran dengan Mie Bangladeshnya, untuk rasa masih okelah, teh susunya enak, hari biasa tetap ramai, pelayanan cepat, toilet bersih
Punctuation Removal	Penasaran dengan Mie Bangladeshnya untuk rasa masih okelah teh susunya enak hari biasa tetap ramai pelayanan cepat toilet bersih
Case Folding	penasaran dengan mie bangladeshnya untuk rasa masih okelah teh susunya enak hari biasa tetap ramai pelayanan cepat toilet bersih
Tokenisasi	penasaran dengan mie bangladeshnya untuk rasa masih okelah teh susunya enak hari biasa tetap ramai pelayanan cepat toilet bersih
Stopword Removal	penasaran mie bangladeshnya okelah teh susunya enak ramai pelayanan cepat toilet bersih
Stemming	penasaran mie bangladeshnya oketeh susunya enak ramai pelayanan cepat toilet bersih
Convert Slangwords	penasaran mie bangladeshnya oketeh susunya enak ramai pelayanan cepat toilet bersih

Gambar 4. 5 Lanjutan Implementasi Tampilan Training Data

Sumber : Dokumentasi Pribadi

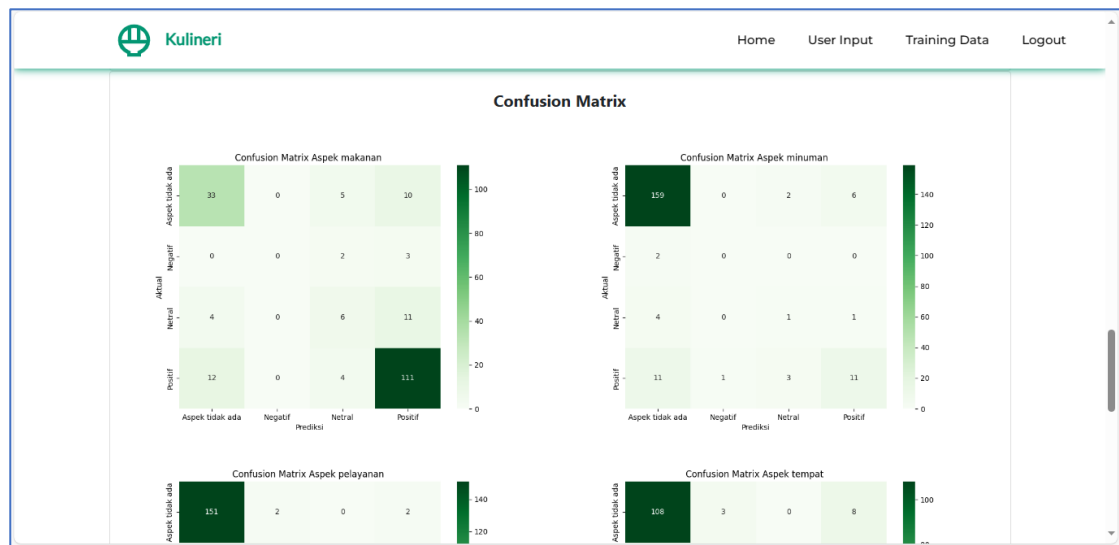
Halaman ini juga menampilkan beberapa diagram yang berisi keterangan hasil training dari masing-masing aspek berupa nilai *precision*, *recall*, *f1-score*, dan akurasi.



Gambar 4. 6 Lanjutan Implementasi Tampilan Training Data

Sumber : Dokumentasi Pribadi

Serta ditampilkan plot confusion matrix dari setiap aspek yaitun aspek makanan, minuman, pelayanan, tempat, dan harga.



Gambar 4. 7 Lanjutan Implementasi Tampilan Training Data

Sumber : Dokumentasi Pribadi

3. Tampilan User Input/Testing

Implementasi pada tampilan *User Input* dibuat berdasarkan rancangan yang telah dijelaskan pada bab 3. Tampilan User Input/Testing akan ditampilkan setelah pengguna melakukan klik tombol navigasi "User Input". Pengguna akan diarahkan ke bagian *User Input* di halaman *Home*. Pada halaman *User Input* . Pada tampilan *User Input*, ditampilkan form berupa kolom *search* yang akan diisi oleh pengguna berupa nama rumah makan yang ingin dicari, beserta tombol *search*. Ditampilkan pula panduan tata cara penggunaan kolom pencarian nama rumah makan

Gambar 4. 8 Implementasi Tampilan User Input

Sumber : Dokumentasi Pribadi

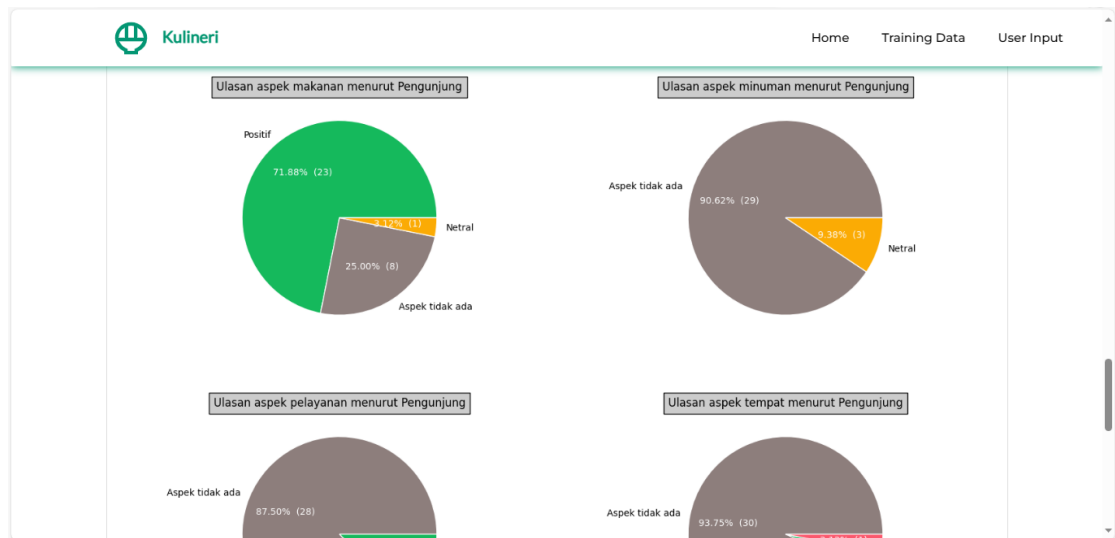
Setelah *input* berupa nama rumah makan dimasukkan dan tombol “Cari” di klik, proses *scraping* ulasan rumah akan dimulai. Setelah proses *scraping* selesai, dilanjutkan dengan proses sentimen ulasan yang telah dikumpulkan berdasarkan model yang sudah dilatih sebelumnya. Pengguna diminta untuk menunggu sampai serangkaian proses selesai.

Ulasan	Makanan	Minuman	Pelayanan	Tempat	Harga
Direkomendasikan	Aspek tidak ada	Aspek tidak ada	Aspek tidak ada	Aspek tidak ada	Aspek tidak ada
Mantab, lumayan enak dan cukup merakyat	Positif	Aspek tidak ada	Aspek tidak ada	Aspek tidak ada	Aspek tidak ada
Pegawai baik,	Positif	Aspek tidak ada	Aspek tidak ada	Aspek	Aspek

Gambar 4. 9 Output tabel pada Implementasi Tampilan User Input

Sumber : Dokumentasi Pribadi

Setelah semua proses selesai, ditampilkan *Output* berupa tabel ulasan yang sudah dikumpulkan, jumlah total ulasan yang dikumpulkan,



Gambar 4. 10 Output plot perbandingan jumlah sentimen pada Implementasi Tampilan User Input

Sumber : Dokumentasi Pribadi

Serta ditampilkan juga plot jumlah perbandingan sentimen positif, netral, dan negatif, yang ada dari setiap aspek yaitu aspek makanan, minuman, pelayanan, tempat, dan harga.

4.2 Implementasi Model

4.2.1 Pelatihan Model

Seperti yang sudah di jelaskan sebelumnya di bab 3, proses *training* yang dilakukan menggunakan pendekatan “*One vs One*” atau “*All vs All*”. Dengan demikian, model dibuat sebanyak 5 model berdasarkan jumlah kelas yang ada pada dataset yaitu kelas makanan, minuman, pelayanan, tempat, dan harga. Proses pelatihan menggunakan beberapa nilai parameter *C* atau parameter *Cost* dengan nilai parameter *C* nya antara lain 0.01, 0.05, 0.25, 0.5, 0.75, 1, 10, 100. Serta digunakan nilai parameter Gamma dengan nilai gamma antara lain 0.001, 0.1, 1, 10. Nilai- nilai tersebut diperoleh dari beberapa penelitian sebelumnya yang telah dijelaskan pada bab3. Proses pelatihan dilakukan beberapa kali dengan mengubah jumlah parameter *C* dan gamma pada setiap perulangan proses pelatihan guna mendapatkan model terbaik. Proses pelatihan juga menggunakan teknik Proses ini juga dapat disebut dengan *hyperparameter tuning*. Model dapat dikatakan baik ketika metrik evaluasi, yang pada kali ini digunakan *f1-score* sebagai metrik akurasi nya, memiliki nilai yang cukup tinggi, dan nilai evaluasi

yang dihasilkan pada setiap aspek tidak memiliki kesenjangan atau kejomplangan yang berarti antara satu dengan yang lain.

Dikarenakan *data training* yang digunakan bersifat tidak seimbang, proses resampling seperti *Borderline SMOTE* maupun *Tomek Links* perlu dilakukan untuk menyeimbangkan data. Proses pelatihan akan dilakukan dalam beberapa skenario, antara lain:

1. Support Vector Machine tanpa resampling
2. Support Vector Machine dengan resampling *Borderline SMOTE*
3. Support Vector Machine dengan resampling Tomek Links
4. Support Vector Machine dengan resampling *Borderline SMOTE* dan Tomek Links
5. Support Vector Machine dengan resampling Tomek Links dan *Borderline SMOTE*

4.2.1.1 Hasil Pelatihan

Proses pelatihan dimulai setelah *data training* berupa file berekstensi .csv telah melalui proses *preprocessing*. Proses pelatihan menghasilkan *f1-score*, baik tanpa dilakukan *cross validation* maupun dengan dilakukan *cross validation*, dari model setiap aspek-aspek yang ada antara lain aspek makanan, minuman, pelayanan, tempat, dan harga. Adapun nilai yang dihasilkan dari proses pelatihan ditunjukkan pada tabel 4. 1

Tabel 4. 1 Hasil Pelatihan Sistem dengan beberapa skenario.

No	Resampling	F1 -Score Makanan (Cross Validation)	F1-Score Minuman (Cross Validation)	F1-Score Pelayanan (Cross Validation)	F1-Score Tempat (Cross Validation)	F1-Score Harga (Cross Validation)
1	-	0.60	0.48	0.68	0.66	0.71

No	Resampling	F1-Score Makanan (Cross Validation)	F1-Score Minuman (Cross Validation)	F1-Score Pelayanan (Cross Validation)	F1-Score Tempat (Cross Validation)	F1-Score Harga (Cross Validation)
2	BSMOTE	0.84	0.85	0.84	0.76	0.87
3	Tomek Links	0.62	0.48	0.68	0.65	0.69
4	BSMOTE + Tomek Links	0.84	0.85	0.85	0.75	0.88
5	Tomek Links + BSMOTE	0.86	0.89	0.87	0.77	0.89

Dengan rincian dari hasil *training* Tomek Links *Borderline* SMOTE di setiap fold di masing-masing aspek dapat dilihat di lihat di gambar 4.11

Fold	F1-Score (Cross Validation)				
	Makanan	Minuman	Pelayanan	Tempat	Harga
1	0.8794606127279394	0.9202640677295055	0.8593545751633987	0.8050531089628835	0.8672205930270448
2	0.8597863247863248	0.9207824087567601	0.9009923397331225	0.7816352201257862	0.8937435897435897
3	0.8581263136823142	0.9003172074129858	0.8327381591322661	0.76993996993997	0.8670148161335627
4	0.8513247863247863	0.8793345888261143	0.8481481198551369	0.7775069722973743	0.870228674040173
5	0.8725063358676803	0.8677286844060771	0.8698682831797188	0.7795495407145893	0.9199631584545386
6	0.8526487171787943	0.9002449594874694	0.8235127113175894	0.804056804056804	0.8600111389585073
7	0.9065885964034298	0.8856546456407921	0.8955014823195794	0.7939162346792742	0.8647321428571427
8	0.8445499181669395	0.8889455483908701	0.9208238301371563	0.7681637933931512	0.9094383600028912
9	0.8913438937148697	0.8881651787783994	0.8846971645835863	0.6924051504510088	0.9040956008169122
10	0.8318367346938776	0.909708138061867	0.8847308741896063	0.7755074781525338	0.9104498875132817
Average	0.8648172233546957	0.8861345426633961	0.8719967722371554	0.7747734272773374	0.8867038061547833

Gambar 4. 11 Gambar Tabel F1-Score

Sumber : Dokumentasi Pribadi

Berdasarkan nilai yang dihasilkan dari proses pelatihan yang ditampilkan pada tabel 4.1, menunjukkan teknik yang menghasilkan *f1-score* paling tinggi dihasilkan melalui skenario nomor 5, yaitu dengan menggunakan teknik Tomek Links terlebih dahulu, dilanjutkan dengan penggunaan teknik *Borderline* SMOTE. Skenario tersebut menghasilkan nilai *f1-score* pada aspek makanan sebesar 0.86 dengan *cross validation*. Pada aspek minuman, dihasilkan nilai *f1-score* sebesar 0.89 dengan *cross validation*. Pada aspek pelayanan, dihasilkan nilai *f1-score* sebesar 0.87 dengan *cross validation*. Pada aspek tempat, dihasilkan nilai *f1-score* sebesar 0.77 dengan *cross validation*. Dan pada aspek minuman, dihasilkan nilai *f1-score* sebesar 0.89 dengan *cross validation*.

Dilampirkan diagram *heatmap* untuk menunjukkan pelatihan menggunakan teknik Grid Search. Dengan visualisasi *heatmap* berwarna kuning menunjukkan nilai *f1-score* yang lebih tinggi, dan warna yang lebih gelap menuju biru menunjukkan nilai *f1-score* yang lebih rendah, yang dapat dilihat pada gambar 4.12, 4.13, 4.14, 4.15, dan 4.16.

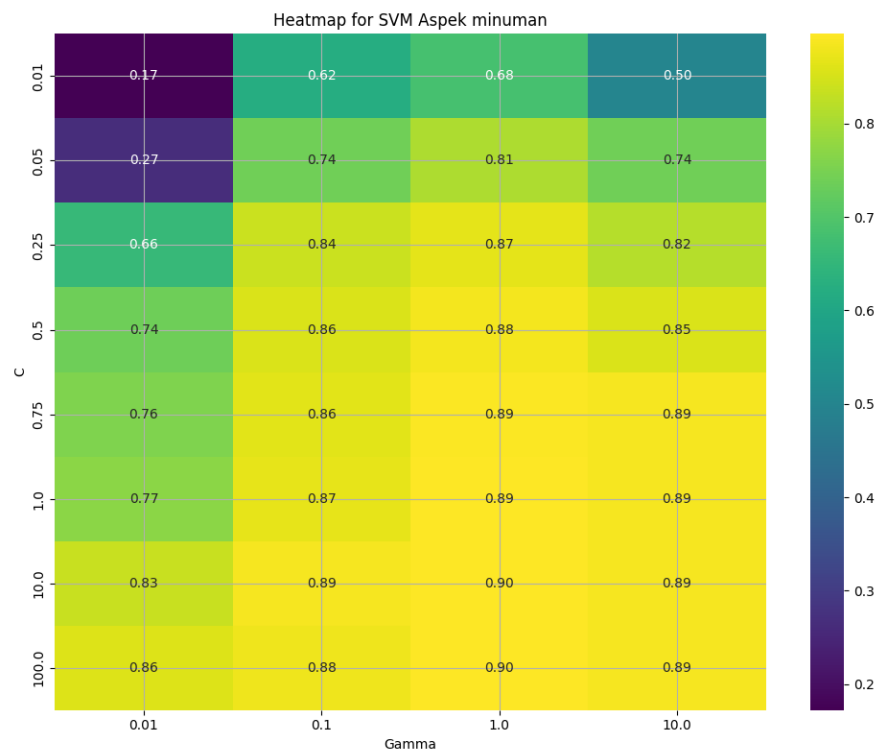
Setelah proses training selesai, dihasilkan model terbaik untuk kelas/aspek makanan dengan menggunakan nilai paramater C sebesar 10 dan gamma sebesar 0.1 dengan *f1-score* sebesar 0.8648. Visualisasi hasil pelatihannya dapat dilihat pada gambar 4.12.



Gambar 4. 12 Diagram kurva pelatihan model aspek makanan

Sumber : Dokumentasi Pribadi

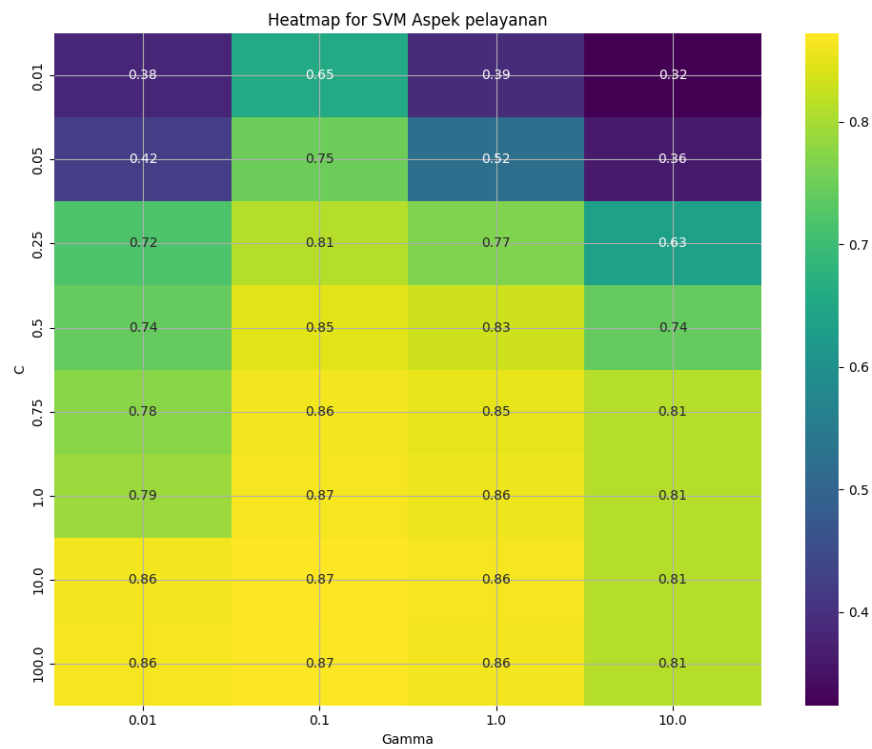
Dihasilkan model terbaik untuk kelas/aspek minuman dengan menggunakan nilai paramater C sebesar 10 dan gamma sebesar 0.1 dengan f1-score sebesar 0.8961. Visualisasi hasil pelatiahannya dapat dilihat pada gambar 4.13.



Gambar 4. 13 Diagram kurva pelatihan model aspek minuman

Sumber : Dokumentasi Pribadi

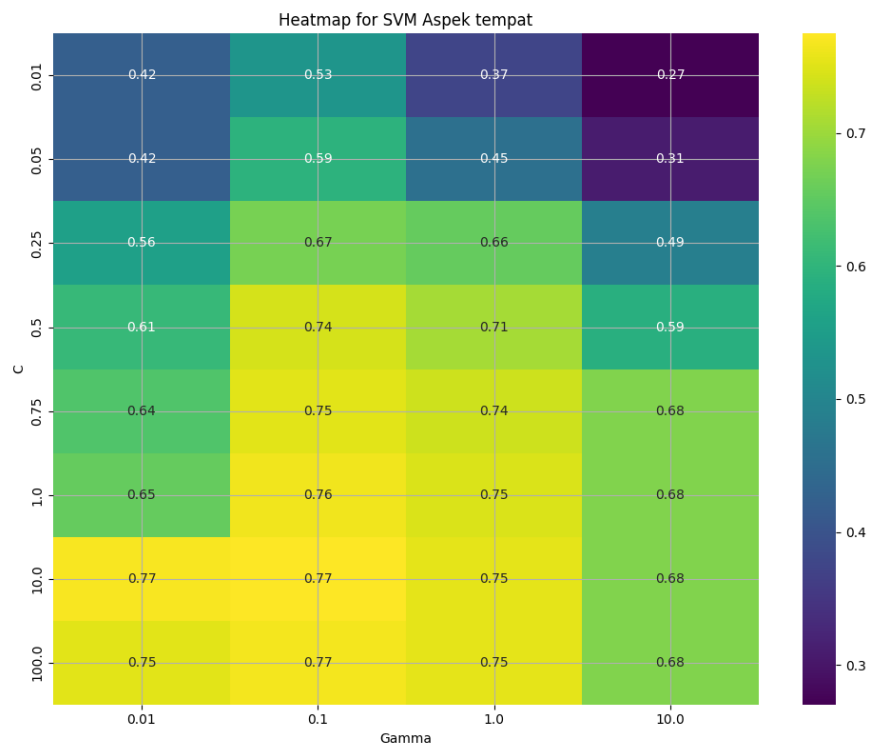
Dihasilkan model terbaik untuk kelas/aspek pelayanan dengan menggunakan nilai paramater C sebesar 10 dan gamma sebesar 0.1 dengan f1-score sebesar 0.8720. Visualisasi hasil pelatiahannya dapat dilihat pada gambar 4.14.



Gambar 4. 14 Diagram kurva pelatihan model aspek pelayanan

Sumber : Dokumentasi Pribadi

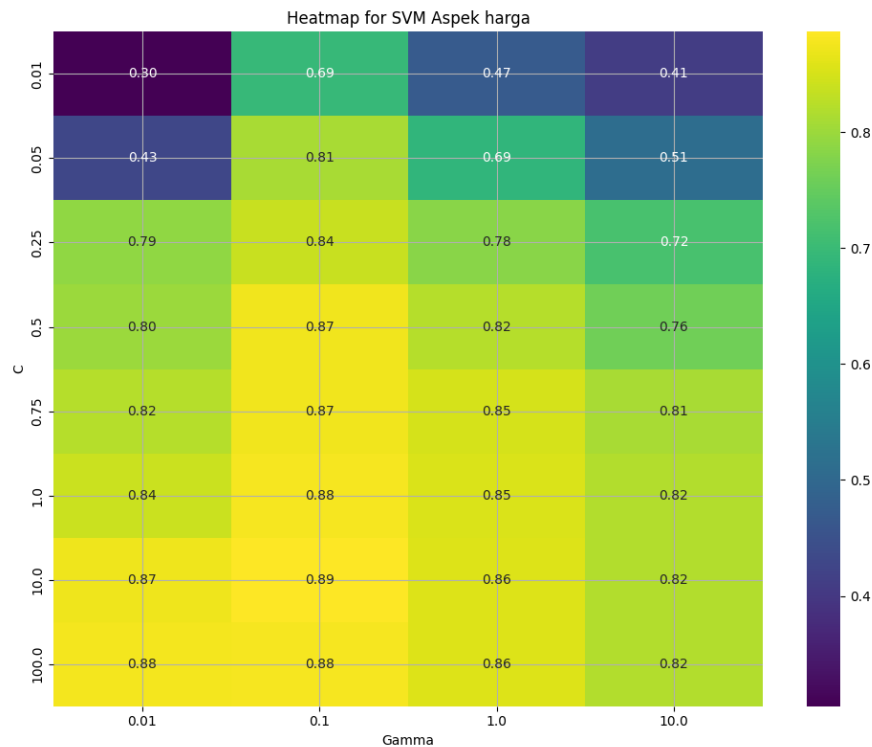
Dihasilkan model terbaik untuk kelas/aspek tempat dengan menggunakan nilai paramater C sebesar 10 dan gamma sebesar 1 dengan f1-score sebesar 0,7747. Visualisasi hasil pelatiahannya dapat dilihat pada gambar 4.15.



Gambar 4. 15 Diagram kurva pelatihan model aspek tempat

Sumber : Dokumentasi Pribadi

Dihasilkan model terbaik untuk kelas/aspek harga dengan menggunakan nilai paramater C sebesar 10 dan gamma sebesar 0.1 dengan f1-score 0.8866. Visualisasi hasil pelatihannya dapat dilihat pada gambar 4.16.



Gambar 4. 16 Diagram kurva pelatihan model aspek harga

Sumber : Dokumentasi Pribadi

4.2.2 Pengujian Model

Setelah proses training yang dilakukan model selesai, selanjutnya model akan diuji performanya terhadap data yang belum dikenal. Dataset yang digunakan pada tahap pengujian model adalah kumpulan ulasan rumah makan berdasarkan *input* pengguna yang dikumpulkan pada proses *scraping*. Pada proses pengujian kali ini, rumah makan Nasi Goreng Tiarbah yang beralamat di Jl. Ps. 1, Tj. Sari, Kec. Medan Selayang, Kota Medan dipilih sebagai sampel. Dikarenakan Nasi Goreng Tiarbah hanya memiliki 79 ulasan di Google Maps, jadi hanya sebanyak 79 ulasan yang untuk dijadikan data sampel pengujian. Jumlah tersebut berkurang menjadi 32 setelah proses *data cleaning* yang dilakukan pada tahap *preprocessing* sebelum data dimasukkan kedalam model untuk diprediksi. Data uji yang sudah di kumpulkan serta di lakukan tahap *preprocessing* berupa proses *cleaning* dapat dilihat pada Tabel 4.1 berikut ini :

Tabel 4. 2 Tabel Ulasan

No	Ulasan
1.	Direkomendasikan
2.	Mantab, lumayan enak dan cukup merakyat
3.	Awalnya pelanggan setia. Sampe ngasitau temen dan saudara wkwk terus ada moment dimana rasanya udah gak konsisten, mikir mungkin lagi apes kali ya jadi next day beli lagi. Tapi rasanya tetep gak seenak awal2 ternyata, kayak ada bahan yg diganti ya? terus nasinya kalo udah dingin keras bgt dan lemaknya gak bisa dikunyah alias kenyel bgt gak kemakan. Udah ngasi kritik dan saran ke admin. Tapi ternyata pas beli lagi eeh gada perubahan. Dari 2022 gapernah beli lagi dech
4.	Pegawai baik, namun pegawai laki-laki kurang sopan karena suka bicarain/gossip tentang pelanggan yang datang, namun makanan cukup enak, ada meong juga jadi terobati
5.	Nasi gorengnya enak. Harga mulai 23 rb an (yaa dengan porsi per cup) mnurut sy terlalu mahal jika melihat porsi yg diberikan. Kalau rasa boleh lah, untuk pilihan level pedas bisa by request. Bagi yg penyuka pedas, sy rekomen deh 👍. Lokasi di Jl. Setia Pasar 1 Medan. ...
6.	Lumayan lah
7.	Warungnya terletak di Jalan Pasar 1 Setiabudi, Medan. Dari Ringroad terletak di sebelah kanan jalan, dari Setiabudi sebelah kiri jalan. Menyajikan nasi goreng rempah dengan dendeng dan pilihan lainnya. Harga 1 porsi mulai dari 25K Rupiah. Tempatnya cukup bersih. Rasa menurut saya standar. Worth untuk dicoba.
8.	Biasa aja menurut aku, soalnya pas aku pesen kayak kebanyakan-an minyak gt trs asin bgtttt☐ ...

No	Ulasan
9.	Nasi goreng dendeng lemak nya lumayann enakkk. Cuman disayangkan untuk parkir mobil agak susah. Krnaaa gak ada lahan parkir nya.
10.	Bintang 3 dulu
11.	Nasi goreng dendeng nya enak, sesekali boleh dicoba buat yang suka makan nasi goreng.
12.	Enak kok, tp lebih baik lagi kalo kedepannya yg masak pakai masker
13.	Selama ini nyobain lemak amis, eh ternyata beda disini nikmat hahaha keren sih, sukses terus nasgor tiarbah 🤪 ...
14.	Terpesona dengan kelezatan nya.....yumiiii 😊😊😊 ...
15.	Racikan nasi gorengnya beda dari yang lain. Recomendeed ❤️
16.	nasi gorengnya enak, tapi klo beli langsung disana utk dibungkus, hrs siap2 ngantri hampir 40 menit (pdhl gak antri2 bgt, cuma antri sm 3 org driver gofood). terus jd kaget sendiri, harga seporsi di gofood dan di kedainya lgsg beda jauh, bukan hanya beda 2-3rb, yaa lebih lah
17.	Hai kak e, terima kasih sudah review nasi goreng kita yaah hihhi. perlu untuk diketahui, pihak ketiga (ojol) menetapkan pembagian 20% untuk setiap transaksi terhadap pihak merchant, jadinya harganya sedikit jauh beda dengan yang di warung 🙌 tapi tenang dengan harga segitu kami siap menjaga kualitas dari rasa kita hihhi Sekali lagi terima kasih yah kak e 🤪
18.	Nagih makan nasgornya
19.	Mantap rasa nya.. Baru X merasakan nasi goreng serasa makan di restoran..

No	Ulasan
20.	Tau dari medsos dan tertarik utk nyobain. Pilih yg ga pedas dan rasanya lumayan. Kayaknya lebih enak yang sedang pedesnya. Lemak sapinya terasa banget. Tapi harganya utk satu porsi sih lumayan.
21.	Tiarbah. Fast respon pelayanannya
22.	Kirain dendeng lemaknya bakalan ganggu, eh ternyata enggak Harus kali sih ini di coba, nasinya juga lembut
23.	nyobain nasgor dendeng lemak dan poinnya ★★★★★ lokasinya ga jauh dari belokan pasar 1, ada di sebelah kiri. tmptnya ga terlalu besar, kayaknya mostly orang pesan online atau take away.
24.	Heran, kok bisa enak ya ? Pantas tiap hari rame teros kutengok yg belik. Sampe ngantri lama. Worth it lah dengan harga segitu uda dapet cita rasa nasgor yg enak kali
25.	Delicious, flavourful spices. Enough said. Nasi Goreng Dendeng Lemak & Sate Lemak, Level Pedas Sedang.
26.	Ngga usah banyak bacod Langsung kesini, kalo pelayanan ga memuaskan cari aja aku. Biar ku bayar makanan klen! Oke lek kuu? Gasskan!
27.	Rasanya bikin nagihh, kuy dicoba gak akan nyesal.
28.	Recomended!! 🌟
29.	Dendeng lemaknya nyess dimulut ⚡
30.	Recomended enak

4.2.3 Hasil Pengujian Model

Data uji yang sudah dikumpulkan tersebut selanjutnya diprediksi oleh model yang sudah dilatih sebelumnya. Hasil prediksi yang dilakukan oleh model terhadap data uji ditampilkan pada Tabel 4.2 berikut :

Tabel 4. 3 Tabel Prediksi ulasan

Ulasan	Makanan	Minuman	Pelayanan	Tempat	Harga
Direkomendasikan	Negatif	Negatif	Negatif	Negatif	Negatif
Mantab, lumayan enak dan cukup merakyat	Positif	Netral	Netral	Netral	Netral
Pegawai baik, namun pegawai laki-laki kurang sopan karena suka bicarain/gossip tentang pelanggan yang datang, namun makanan cukup enak, ada meong juga jadi terobati	Positif	Netral	Netral	Netral	Netral
Awalnya pelanggan setia. Sampe ngasitau temen dan saudara wkwk terus ada moment dimana rasanya udah gak konsisten, mikir mungkin lagi apes kali ya jadi next day beli lagi. Tapi rasanya tetep gak	Positif	Netral	Netral	Netral	Netral

Ulasan	Makanan	Minuman	Pelayanan	Tempat	Harga
<p>seenak awal2 ternyata, kayak ada bahan yg diganti ya? terus nasinya kalo udah dingin keras bgt dan lemaknya gak bisa dikunyah alias kenyel bgt gak kemakan. Udah ngasi kritik dan saran ke admin. Tapi ternyata pas beli lagi eeh gada perubahan. Dari 2022 gapernah beli lagi deh</p>					
<p>Nasi gorengnya enak. Harga mulai 23 rb an (yaa dengan porsi per cup) mnurut sy terlalu mahal jika melihat porsi yg diberikan. Kalau rasa boleh lah, untuk pilihan level pedas bisa by request. Bagi yg penyuka pedas, sy rekomen deh 👍.</p>	Positif	Netral	Netral	Netral	Positif

Ulasan	Makanan	Minuman	Pelayanan	Tempat	Harga
Lokasi di Jl. Setia Pasar 1 Medan. ...					
Lumayan lah	Positif	Netral	Netral	Netral	Netral
Warungnya terletak di Jalan Pasar 1 Setiabudi, Medan. Dari Ringroad terletak di sebelah kanan jalan, dari Setiabudi sebelah kiri jalan. Menyajikan nasi goreng rempah dengan dendeng dan pilihan lainnya. Harga 1 porsi mulai dari 25K Rupiah. Tempatnya cukup bersih. Rasa menurut saya standar. Worth untuk dicoba.	Positif	Netral	Netral	Netral	Netral
Biasa aja menurut aku, soalnya pas aku pesen kayak kebanyak-an minyak gt trs asin bgtttt□ ...	Netral	Netral	Netral	Netral	Netral

Tabel tersebut di visualisasikan ke dalam tabel seperti yang ditampilkan pada gambar 4.16 berikut ini.



Ulasan sebanyak 29 ulasan, yang sudah diambil pada rumah makan **Nasi Goreng Tiarbah Ps.1** beserta aspeknya sebagai berikut :

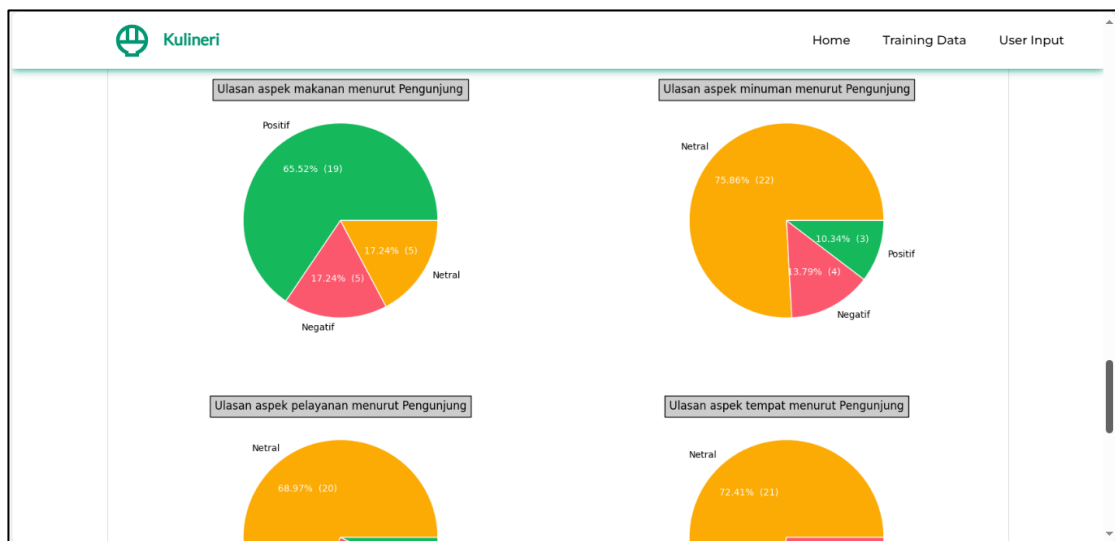
Show entries

Ulasan	Makanan	Minuman	Pelayanan	Tempat	Harga
Direkomendasikan	Negatif	Negatif	Negatif	Negatif	Negatif
Mantab, lumayan enak dan cukup merakyat	Positif	Netral	Netral	Netral	Netral
Pegawai baik, namun pegawai laki-laki kurang sopan karena suka bicarain/gossip tentang pelanggan yang datang, namun makanan cukup enak, ada meong	Positif	Netral	Netral	Netral	Netral

Gambar 4. 17 Tampilan *Testing*

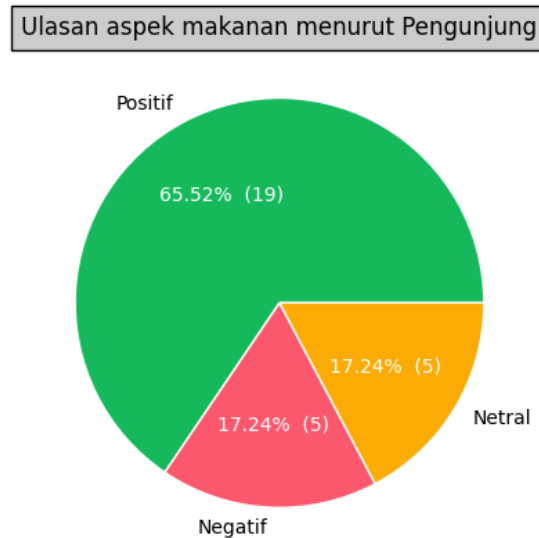
Sumber : Dokumentasi Pribadi

Semua ulasan yang telah diambil juga direkapitulasi dan disajikan kembali ke dalam bentuk *pie chart* berjumlah 5 buah sesuai dengan aspek-aspek yang telah ditentukan yaitu makanan, minuman, pelayanan, tempat, dan harga. Seperti yang ditampilkan pada gambar 4.17 berikut



Gambar 4. 18 Bagan *Pie Chart* Tampilan *Testing*

Sumber : Dokumentasi Pribadi

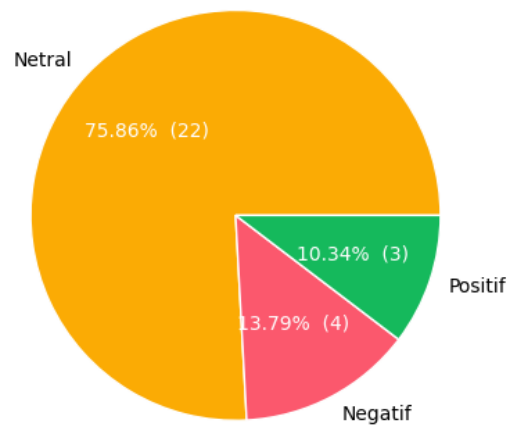


Gambar 4. 19 *Pie Chart* Aspek Makanan

Sumber : Dokumentasi Pribadi

Pada gambar 4.19 ditampilkan hasil rekapitulasi ulasan yang sudah diambil dan di prediksi, serta disajikan kembali dalam bentuk Pie chart untuk aspek makanan. Pie chart tersebut menunjukkan bahwa dari 29 ulasan yang telah diambil, 19 ulasan atau 65.52% menyatakan bahwa Restoran Nasi Goreng Tiarbah memiliki sentimen positif pada Aspek makanannya. Dilanjutkan dengan 5 ulasan atau 17.24% menyatakan sentimen netral, dan 5 ulasan atau 17.24% untuk sentimen negatif.

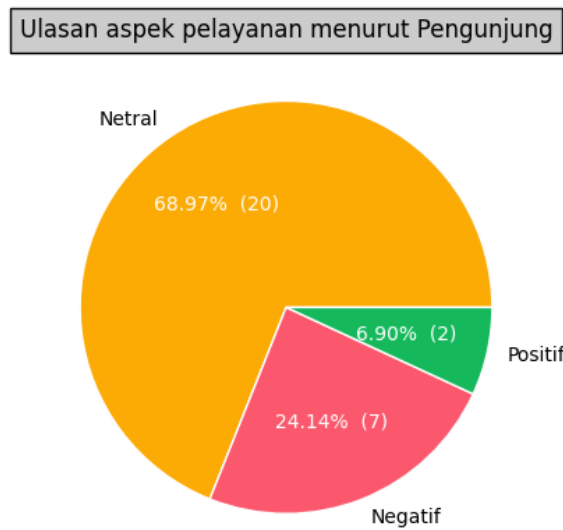
Ulasan aspek minuman menurut Pengunjung



Gambar 4. 20 *Pie Chart* Aspek Minuman

Sumber : Dokumentasi Pribadi

Pada gambar 4.20 ditampilkan hasil rekapitulasi ulasan dan disajikan kembali dalam bentuk Pie chart untuk aspek minuman. Pie chart tersebut menunjukkan bahwa dari 29 ulasan yang telah diambil, 22 ulasan atau 75.86% menyatakan bahwa Restoran Nasi Goreng Tiarbah memiliki sentimen netral pada Aspek minumannya. Dilanjutkan dengan 4 ulasan atau 13.79% menyatakan sentimen negatif, dan 5 ulasan atau 10.34% untuk sentimen positif.

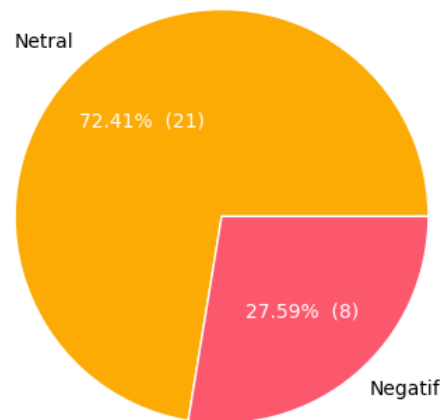


Gambar 4. 21 *Pie Chart* Aspek Pelayanan

Sumber : Dokumentasi Pribadi

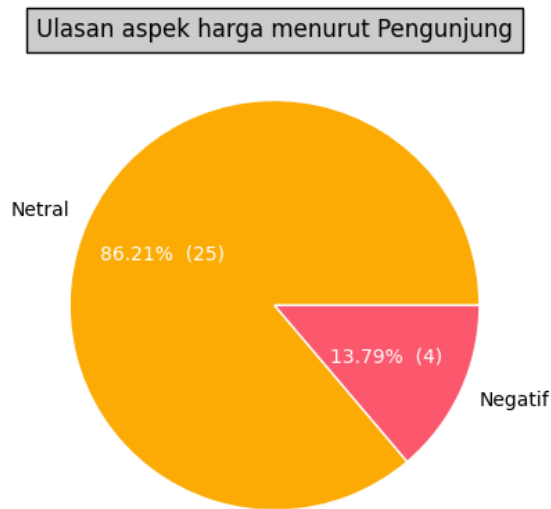
Pada gambar 4.21 ditampilkan hasil rekapitulasi ulasan dan disajikan kembali dalam bentuk Pie chart untuk aspek pelayanan. Pie chart tersebut menunjukkan bahwa dari 29 ulasan yang telah diambil, 20 ulasan atau 68.97% menyatakan bahwa Restoran Nasi Goreng Tiarbah memiliki sentimen netral pada Aspek minumannya. Dilanjutkan dengan 7 ulasan atau 23.14% menyatakan sentimen negatif, dan 2 ulasan atau 6.90% untuk sentimen positif.

Ulasan aspek tempat menurut Pengunjung

Gambar 4. 22 *Pie Chart* Aspek Tempat

Sumber : Dokumentasi Pribadi

Pada gambar 4.22 ditampilkan hasil rekapitulasi ulasan dan disajikan kembali dalam bentuk Pie chart untuk aspek tempat. Pie chart tersebut menunjukkan bahwa dari 29 ulasan yang telah diambil, 21 ulasan atau 72.41% menyatakan bahwa Restoran Nasi Goreng Tiarbah memiliki sentimen netral pada Aspek pelayanannya. Dilanjutkan dengan 8 ulasan atau 27.59% menyatakan sentimen negatif.



Gambar 4. 23 *Pie Chart* Aspek Harga

Sumber : Dokumentasi Pribadi

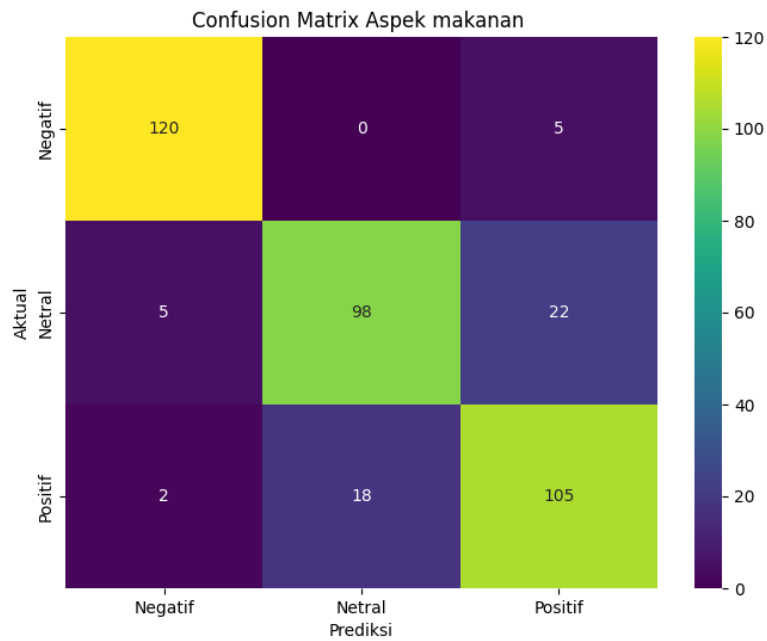
Pada gambar 4.23 ditampilkan hasil rekapitulasi ulasan dan disajikan kembali dalam bentuk Pie chart untuk aspek harga. Pie chart tersebut menunjukkan bahwa dari 29 ulasan yang telah diambil, 25 ulasan atau 86.21% menyatakan bahwa Restoran Nasi Goreng Tiarbah memiliki sentimen netral pada Aspek harga. Dilanjutkan dengan 4 ulasan atau 13.79% menyatakan sentimen negatif.

Secara keseluruhan, dapat disimpulkan bahwa Restoran Nasi Goreng Tiarbah yang terletak di Jl. Pasar 1 Setiabudi memiliki kualitas yang cukup baik, ditandai dengan kecenderungan sentimen positif untuk aspek makanan nya, dan kecenderungan sentimen netral untuk aspek minuman, pelayanan, tempat, dan harga, Walaupun masih terdapat beberapa sentimen negatif pada setiap aspek nya.

4.2.4 Evaluasi Model

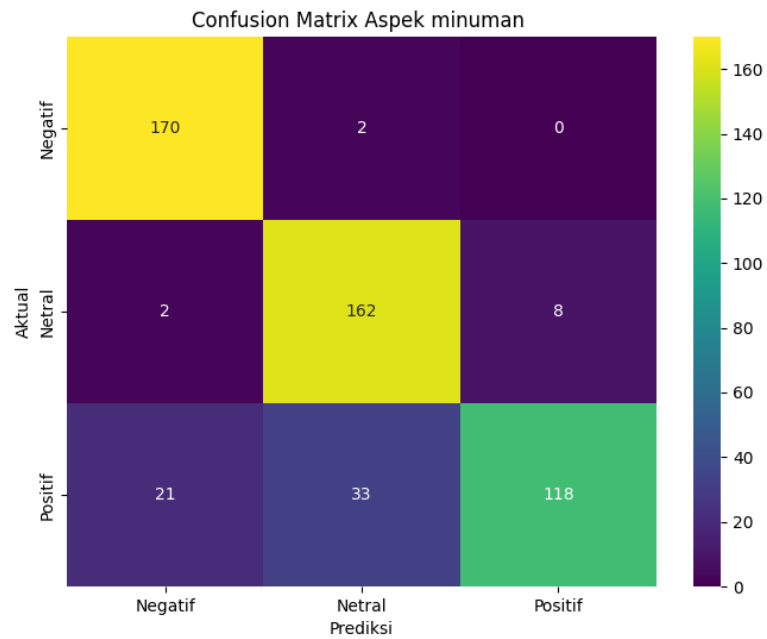
Tahap evaluasi model dilakukan untuk mengetahui performa serta menjadi tolak ukur apakah suatu model dapat berkerja dengan optimal dalam proses pengklasifikasian analisis sentimen berbasis aspek. Tahap evaluasi dilakukan dengan menghitung serta memperhatikan nilai akurasi, *precision*, *recall*, dan *f1-score* yang diperoleh melalui metode *confusion matrix*. Metode ini menghasilkan sebuah *output* visualisasi berupa diagram *heatmap* yang menampilkan pola berdasarkan *True Positive* (TP), (TP), *False*

Positive (FP), dan *False Negative* (FN) untuk aspek makanan, aspek minuman, aspek pelayanan, aspek tempat dan aspek harga. Dengan visualisasi *heatmap* berwarna kuning menunjukkan nilai yang lebih banyak, dan warna yang lebih gelap menuju biru menunjukkan nilai yang lebih sedikit.



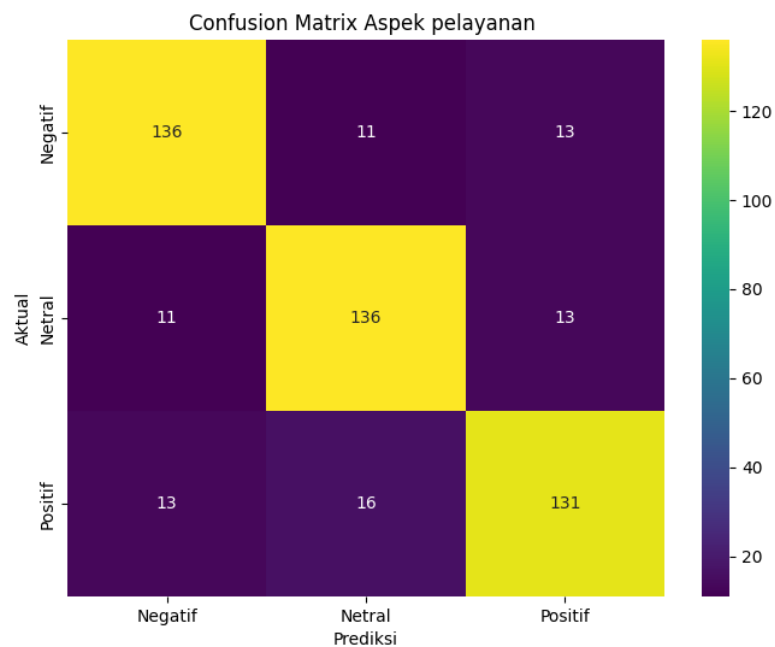
Gambar 4. 24 *Confusion Matrix* model aspek makanan

Sumber : Dokumentasi Pribadi



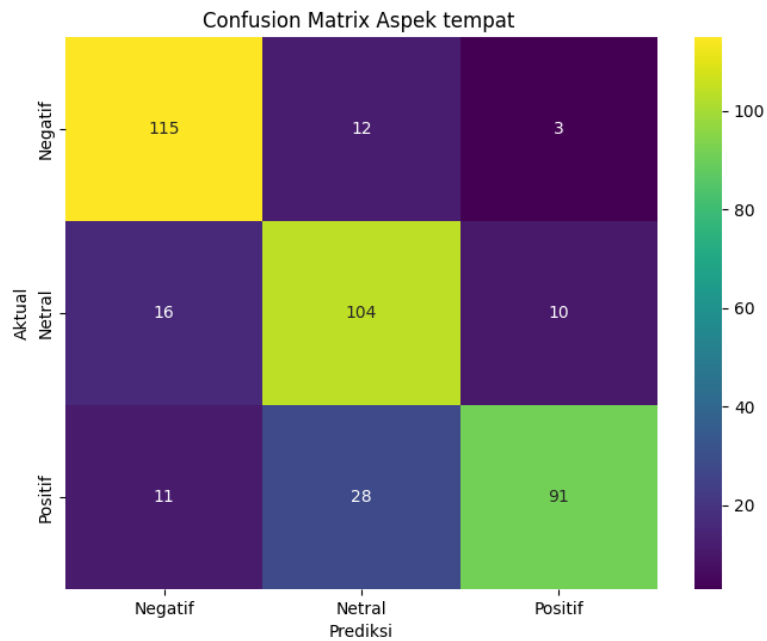
Gambar 4. 25 *Confusion Matrix* model aspek minuman

Sumber : Dokumentasi Pribadi



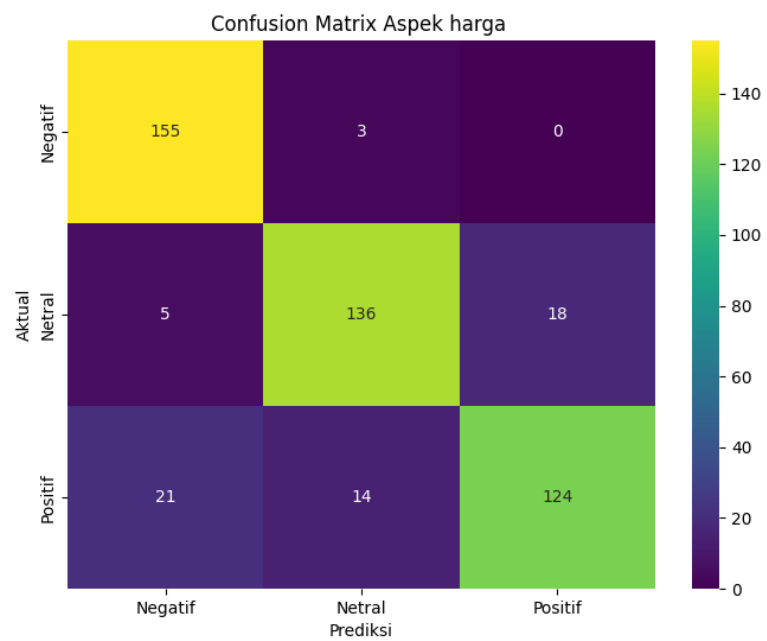
Gambar 4. 26 *Confusion Matrix* model aspek pelayanan

Sumber : Dokumentasi Pribadi



Gambar 4. 27 *Confusion Matrix* model aspek tempat

Sumber : Dokumentasi Pribadi



Gambar 4. 28 *Confusion Matrix* model aspek harga

Sumber : Dokumentasi Pribadi

Berdasarkan visualisasi *heatmap* dari aspek makanan, minuman, pelayanan, tempat, dan harga yang ditampilkan pada gambar 4.24 , 4.25 , 4.26, 4.27 , dan 4.28, diperoleh keterangan klasifikasi sentimen untuk aspek makanan yang dijabarkan pada beberapa tabel yaitu tabel 4.5, tabel 4.6, tabel 4.7 berikut ini :

Tabel 4. 4 Keterangan Sentimen Positif Aspek Makanan *Confussion Matrix*

No	Sentimen Positif	Total
1.	TPPosMak (<i>True Positive</i> Positif Makanan)	105
2.	TNPosMak (<i>True Negative</i> Positif Makanan)	223
3.	FPPosMak (<i>False Positive</i> Positif Makanan)	27
4.	FNPosMak (<i>False Negative</i> Positif Makanan)	20

Tabel 4. 5 Keterangan Sentimen Netral Aspek Makanan *Confussion Matrix*

No	Sentimen Netral	Total
1.	TPNetMak (<i>True Positive</i> Netral Makanan)	98
2.	TNNetMak (<i>True Negative</i> Netral Makanan)	232
3.	FPNetMak (<i>False Positive</i> Netral Makanan)	18
4.	FNNNetMak (<i>False Negative</i> Netral Makanan)	27

Tabel 4. 6 Keterangan Sentimen Negatif Aspek Makanan *Confussion Matrix*

No	Sentimen Negatif	Total
1.	TPNegMak (<i>True Positive</i> Negatif Makanan)	120
2.	TNNegMak (<i>True Negative</i> Negatif Makanan)	237
3.	FPNegMak (<i>False Positive</i> Negatif Makanan)	7
4.	FNNegMak (<i>False Negative</i> Negatif Makanan)	5

Lalu, ditampilkan keterangan klasifikasi sentimen untuk aspek makanan yang dijabarkan pada beberapa tabel yaitu tabel 4.8, tabel 4.9, tabel 4.10 berikut ini :

Tabel 4. 7 Keterangan Sentimen Positif Aspek Minuman *Confussion Matrix*

No	Sentimen Positif	Total
5.	TPPosMin(<i>True Positive</i> Positif Minuman)	118
6.	TNPosMin (<i>True Negative</i> Positif Minuman)	336
7.	FPPosMin (<i>False Positive</i> Positif Minuman)	8
8.	FNPosMin (<i>False Negative</i> Positif Minuman)	54

Tabel 4. 8 Keterangan Sentimen Netral Aspek Minuman *Confussion Matrix*

No	Sentimen Netral	Total
5.	TPNetMin (<i>True Positive</i> Netral Minuman)	162
6.	TNNetMin (<i>True Negative</i> Netral Minuman)	309
7.	FPNetMin (<i>False Positive</i> Netral Minuman)	35
8.	FNNetMin (<i>False Negative</i> Netral Minuman)	10

Tabel 4. 9Keterangan Sentimen Negatif Aspek Minuman *Confussion Matrix*

No	Sentimen Negative	Total
1.	TPNegMin (<i>True Positive</i> Negatif Minuman)	170
2.	TNNegMin (<i>True Negative</i> Negatif Minuman)	321
3.	FPNegMin (<i>False Positive</i> Negatif Minuman)	23
4.	FNNegMin (<i>False Negative</i> Negatif Minuman)	2

Ditampilkan keterangan klasifikasi sentimen untuk aspek pelayanan yang dijabarkan pada beberapa tabel yaitu tabel 4.11, tabel 4.12, tabel 4.13 berikut ini :

Tabel 4. 10 Keterangan Sentimen Positif Aspek Pelayanan *Confussion Matrix*

No	Sentimen Positif	Total
9.	TPPosPly (<i>True Positive</i> Positif Pelayanan)	131

10	TNPosPly (<i>True Negative</i> Positif Pelayanan)	294
11	FPPosPly (<i>False Positive</i> Positif Pelayanan)	26
12	FNPosPly (<i>False Negative</i> Positif Pelayanan)	29

Tabel 4. 11 Keterangan Sentimen Netral Aspek Pelayanan *Confussion Matrix*

No	Sentimen Netral	Total
9.	TPNetPly (<i>True Positive</i> Netral Pelayanan)	136
10	TNNetPly (<i>True Negative</i> Netral Pelayanan)	293
11	FPNetPly (<i>False Positive</i> Netral Pelayanan)	27
12	FNNetPly (<i>False Negative</i> Netral Pelayanan)	24

Tabel 4. 12 Keterangan Sentimen Netral Aspek Pelayanan *Confussion Matrix*

No	Sentimen Negative	Total
1.	TPNegPly (<i>True Positive</i> Negatif Pelayanan)	136
2.	TNNegPly (<i>True Negative</i> Negatif Pelayanan)	296
3.	FPNegPly (<i>False Positive</i> Negatif Pelayanan)	24
4.	FNNegPly (<i>False Negative</i> Negatif Pelayanan)	24

Ditampilkan keterangan klasifikasi sentimen untuk aspek tempat yang dijabarkan pada beberapa tabel yaitu tabel 4.14, tabel 4.15, tabel 4.16 berikut ini :

Tabel 4. 13 Keterangan Sentimen Positif Aspek Tempat *Confussion Matrix*

No	Sentimen Positif	Total
13	TPPosTpt (<i>True Positive</i> Positif Tempat)	91
14	TNPosTpt (<i>True Negative</i> Positif Tempat)	247
15	FPPosTpt (<i>False Positive</i> Positif Tempat)	13
16	FNPosTpt (<i>False Negative</i> Positif Tempat)	36

Tabel 4. 14 Keterangan Sentimen Netral Aspek Tempat *Confussion Matrix*

No	Sentimen Netral	Total
13	TPNetTpt (<i>True Positive</i> Netral Tempat)	104
14	TNNetTpt (<i>True Negative</i> Netral Tempat)	220
15	FPNetTpt (<i>False Positive</i> Netral Tempat)	40
16	FNNetTpt (<i>False Negative</i> Netral Tempat)	26

Tabel 4. 5 Keterangan Sentimen Negatif Aspek Tempat *Confussion Matrix*

No	Sentimen Negative	Total
5.	TPNegTpt (<i>True Positive</i> Negatif Tempat)	115
6.	TNNegTpt (<i>True Negative</i> Positif Tempat)	233
7.	FPNegTpt (<i>False Positive</i> Positif Tempat)	27
8.	FNNegTpt (<i>False Negative</i> Positif Tempat)	15

Ditampilkan keterangan klasifikasi sentimen untuk aspek tempat yang dijabarkan pada beberapa tabel yaitu tabel 4.16, tabel 4.17, tabel 4.18 berikut ini :

Tabel 4. 16 Keterangan Sentimen Positif Aspek Harga *Confussion Matrix*

No	Sentimen Positif	Total
17	TPPosTpt (<i>True Positive</i> Positif Harga)	124
18	TNPosTpt (<i>True Negative</i> Positif Harga)	299
19	FPPosTpt (<i>False Positive</i> Positif Harga)	18
20	FNPosTpt (<i>False Negative</i> Positif Harga)	34

Tabel 4. 17 Keterangan Sentimen Netral Aspek Harga *Confussion Matrix*

No	Sentimen Netral	Total
17	TPNetTpt (<i>True Positive</i> Netral Harga)	136

18	TNNetTpt (<i>True Negative</i> Netral Harga)	294
19	FPNetTpt (<i>False Positive</i> Netral Harga)	17
20	FNNetTpt (<i>False Negative</i> Netral Harga)	23

Tabel 4. 18 Keterangan Sentimen Negatif Aspek Harga *Confussion Matrix*

No	Sentimen Negative	Total
9.	TPNegTpt (<i>True Positive</i> Negatif Harga)	155
10	TNNegTpt (<i>True Negative</i> Negatif Harga)	292
11	FPNegTpt (<i>False Positive</i> Negatif Harga)	26
12	FNNegTpt (<i>False Negative</i> Negatif Harga)	3

Berdasarkan hasil analisis sentimen yang ditampilkan dalam Tabel 4.6 hingga Tabel 4.18, langkah evaluasi dapat dijalankan dengan menghitung nilai akurasi, precision, recall, dan f-score. Evaluasi ini dilakukan pada pengujian terbaik dengan melibatkan 1005 data, dan metrik-metrik tersebut dihitung menggunakan persamaan yang telah dijelaskan pada bab sebelumnya yaitu bab 3.

- Perhitungan Precision aspek makanan :

$$\text{Precision aspek makanan sentimen positif (+)} = \frac{TP}{TP+FP} \times 100 = \frac{105}{105+27} \times 100 = 0.79$$

$$\text{Precision aspek makanan sentimen netral (0)} = \frac{TP}{TP+FP} \times 100 = \frac{98}{98+18} \times 100 = 0.84$$

$$\text{Precision aspek makanan sentimen negatif (-)} = \frac{TP}{TP+FP} \times 100 = \frac{120}{120+7} \times 100 = 0.94$$

- Perhitungan Precision aspek minuman :

$$\text{Precision aspek minuman sentimen positif (+)} = \frac{TP}{TP+FP} \times 100 = \frac{118}{118+8} \times 100 = 0.93$$

$$\text{Precision aspek minuman sentimen netral (0)} = \frac{TP}{TP+FP} \times 100 = \frac{162}{162+35} \times 100 = 0.82$$

$$\text{Precision aspek minuman sentimen negatif (-)} = \frac{TP}{TP+FP} \times 100 = \frac{170}{170+23} \times 100 = 0.88$$

- Perhitungan Precision aspek pelayanan :

$$\text{Precision aspek pelayanan sentimen positif (+)} = \frac{TP}{TP+FP} \times 100 = \frac{131}{131+26} \times 100 = 0.83$$

$$\text{Precision aspek pelayanan sentimen netral (0)} = \frac{TP}{TP+FP} \times 100 = \frac{136}{136+27} \times 100 = 0.83$$

$$\text{Precision aspek pelayanan sentimen negatif (-)} = \frac{TP}{TP+FP} \times 100 = \frac{136}{136+24} \times 100 = 0.85$$

- Perhitungan Precision aspek tempat :

$$\text{Precision aspek tempat sentimen positif (+)} = \frac{TP}{TP+FP} \times 100 = \frac{91}{91+13} \times 100 = 0.87$$

$$\text{Precision aspek tempat sentimen netral (0)} = \frac{TP}{TP+FP} \times 100 = \frac{104}{104+40} \times 100 = 0.72$$

$$\text{Precision aspek tempat sentimen negatif (-)} = \frac{TP}{TP+FP} \times 100 = \frac{115}{115+27} \times 100 = 0.80$$

- Perhitungan Precision aspek harga :

$$\text{Precision aspek harga sentimen positif (+)} = \frac{TP}{TP+FP} \times 100 = \frac{124}{124+18} \times 100 = 0.87$$

$$\text{Precision aspek harga sentimen netral (0)} = \frac{TP}{TP+FP} \times 100 = \frac{136}{136+17} \times 100 = 0.88$$

$$\text{Precision aspek harga sentimen negatif (-)} = \frac{TP}{TP+FP} \times 100 = \frac{155}{155+26} \times 100 = 0.85$$

- Perhitungan Recall aspek makanan :

$$\text{Recall aspek makanan sentimen positif (+)} = \frac{TP}{TP+FN} \times 100 = \frac{105}{105+20} \times 100 = 0.84$$

$$\text{Recall aspek makanan sentimen netral (0)} = \frac{TP}{TP+FN} \times 100 = \frac{98}{98+27} \times 100 = 0.77$$

$$\text{Recall aspek makanan sentimen negatif (-)} = \frac{TP}{TP+FN} \times 100 = \frac{120}{120+5} \times 100 = 0.96$$

- Perhitungan Recall aspek minuman :

$$\text{Recall aspek minuman sentimen positif (+)} = \frac{TP}{TP+FN} \times 100 = \frac{118}{118+54} \times 100 = 0.68$$

$$\text{Recall aspek minuman sentimen netral (0)} = \frac{TP}{TP+FN} \times 100 = \frac{162}{162+10} \times 100 = 0.94$$

$$\text{Recall aspek minuman sentimen negatif (-)} = \frac{TP}{TP+FN} \times 100 = \frac{170}{170+2} \times 100 = 0.98$$

- Perhitungan Recall aspek pelayanan :

$$\text{Recall aspek pelayanan sentimen positif (+)} = \frac{TP}{TP+FN} \times 100 = \frac{131}{131+29} \times 100 = 0.81$$

$$\text{Recall aspek pelayanan sentimen netral (0)} = \frac{TP}{TP+FN} \times 100 = \frac{136}{136+27} \times 100 = 0.83$$

$$\text{Recall aspek pelayanan sentimen negatif (-)} = \frac{TP}{TP+FN} \times 100 = \frac{136}{136+24} \times 100 = 0.85$$

- Perhitungan Recall aspek tempat :

$$\text{Recall aspek tempat sentimen positif (+)} = \frac{TP}{TP+FN} \times 100 = \frac{91}{91+36} \times 100 = 0.71$$

$$\text{Recall aspek tempat sentimen netral (0)} = \frac{TP}{TP+FN} \times 100 = \frac{104}{104+26} \times 100 = 0.80$$

$$\text{Recall aspek tempat sentimen negatif (-)} = \frac{TP}{TP+FN} \times 100 = \frac{115}{115+15} \times 100 = 0.88$$

- Perhitungan Precision aspek harga :

$$\text{Precision aspek harga sentimen positif (+)} = \frac{TP}{TP+FN} \times 100 = \frac{124}{124+34} \times 100 = 0.78$$

$$\text{Precision aspek harga sentimen netral (0)} = \frac{TP}{TP+FN} \times 100 = \frac{136}{136+23} \times 100 = 0.85$$

$$\text{Precision aspek harga sentimen negatif (-)} = \frac{TP}{TP+FN} \times 100 = \frac{155}{155+3} \times 100 = 0.98$$

- Perhitungan F1-score aspek makanan :

$$\text{F1-score aspek makanan sentimen positif (+)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{210}{210+20+27} \times 100 = 0.81$$

$$\text{F1-score aspek makanan sentimen netral (0)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{196}{196+27+18} \times 100 = 0.81$$

$$\text{F1-score aspek makanan sentimen negatif (-)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{240}{240+5+7} \times 100 = 0.95$$

- Perhitungan F1-score aspek minuman :

$$\text{F1-score aspek minuman sentimen positif (+)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{236}{236+54+8} \times 100 = 0.79$$

$$\text{F1-score aspek minuman sentimen netral (0)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{196}{196+27+18} \times 100 = 0.81$$

$$\text{F1-score aspek minuman sentimen negatif (-)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{240}{240+5+7} \times 100 = 0.95$$

- Perhitungan F1-score aspek pelayanan :

$$\text{F1-score aspek pelayanan sentimen positif (+)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{262}{262+29+26} \times 100 = 0.82$$

$$\text{F1-score aspek pelayanan sentimen netral (0)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{272}{272+24+27} \times 100 = 0.84$$

$$\text{F1-score aspek pelayanan sentimen negatif (-)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{272}{272+24+24} \times 100 = 0.85$$

- Perhitungan F1-score aspek tempat :

$$\text{F1-score aspek tempat sentimen positif (+)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{182}{182+36+13} \times 100 = 0.82$$

$$\text{F1-score aspek tempat sentimen netral (0)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{208}{208+26+40} \times 100 = 0.75$$

$$\text{F1-score aspek tempat sentimen negatif (-)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{230}{230+15+27} \times 100 = 0.84$$

- Perhitungan F1-score aspek harga :

$$\text{F1-score aspek harga sentimen positif (+)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{248}{248+34+18} \times 100 = 0.82$$

$$\text{F1-score aspek harga sentimen netral (0)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{272}{272+23+17} \times 100 = 0.87$$

$$\text{F1-score aspek harga sentimen negatif (-)} = \frac{2TP}{2TP+FN+FP} \times 100 = \frac{310}{310+3+26} \times 100 = 0.91$$

- Perhitungan Akurasi :

$$\text{Akurasi Makanan} = \frac{TP+TN}{TP+TN++FP+FN} = \frac{105+223}{105+223+20+28} = \frac{328}{378} = 0.86$$

$$\text{Akurasi Minuman} = \frac{TP+TN}{TP+TN++FP+FN} = \frac{118+336}{118+336+54+8} = \frac{454}{516} = 0.87$$

$$\text{Akurasi Pelayanan} = \frac{TP+TN}{TP+TN++FP+FN} = \frac{131+294}{131+294+29+26} = \frac{425}{480} = 0.88$$

$$\text{Akurasi Tempat} = \frac{TP+TN}{TP+TN++FP+FN} = \frac{91+247}{91+247++36+13} = \frac{338}{387} = 0.87$$

$$\text{Akurasi Harga} = \frac{TP+TN}{TP+TN++FP+FN} = \frac{124+299}{124+299++34+18} = \frac{423}{475} = 0.89$$

Sehingga, total rata-rata nilai accuracy berdasarkan kelima aspek:

$$\text{Akurasi rata-rata} = \frac{0.86+0.87+0.88+0.87+0.89}{5} \times 100 = \frac{4.37}{5} \times 100 = 87.4$$

F1-Score rata-rata (diperoleh dari proses *training* model) :

$$\frac{0.86 + 0.89 + 0.87 + 0.77 + 0.89}{5} \times 100 = \frac{4.22}{5} \times 100 = 85.6$$

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, beberapa kesimpulan yang diperoleh terkait penelitian aspect-based sentiment analysis pada ulasan makanan menggunakan Support Vector Machine dengan penerapan teknik resampling *Borderline* SMOTE dan Tomek Links adalah sebagai berikut:

1. Algoritma *Support Vector Machine* dengan teknik resampling *Borderline* SMOTE dan Tomek Links menunjukkan kinerja yang cukup baik dalam melakukan analisis sentimen pada ulasan makanan berdasarkan lima aspek. Evaluasi dilakukan dengan menggunakan metode *confusion matrix*, dan hasil rata-rata akurasi mencapai 87,4%. Selain itu, nilai rata-rata f1-score dari proses pelatihan adalah sebesar 85,6%. Walaupun masih terdapat misklasifikasi dalam melakukan analisis ulasan ke dalam sentimen.
2. Penggunaan *Borderline* SMOTE dan Tomek Links berdampak cukup signifikan dalam mengatasi ketidakseimbangan data pada penelitian ini. F1-Score yang dihasilkan sebelum penggunaan *Borderline* SMOTE dan Tomek Links sebesar 60% untuk aspek makanan, 48% untuk aspek minuman, 68% untuk aspek pelayanan, 66% untuk aspek tempat, dan 71% untuk aspek harga. Sedangkan setelah penggunaan *Borderline* SMOTE dan Tomek Links sebesar 86% untuk aspek makanan, 89% untuk aspek minuman, 87% untuk aspek pelayanan, 77% untuk aspek tempat, dan 89% untuk aspek harga. Peningkatan yang terjadi pada f1-score berkisar 26% untuk aspek makanan, 41% untuk aspek minuman, 19% untuk aspek pelayanan, 11% untuk aspek tempat, dan 18% untuk aspek harga. Hal ini berdampak pada peningkatan kinerja model yang digunakan dalam melakukan *aspect based sentiment analysis*.

5.2 Saran

Adapun saran yang dapat diajukan penulis untuk penelitian kedepannya antara lain:

1. Pada penelitian selanjutnya, diharapkan data yang digunakan dalam proses *training* lebih beragam, lebih banyak, serta lebih berkualitas.
2. Pada penelitian berikutnya, mempertimbangkan untuk menggunakan algoritma yang lebih mendalam dan kompleks, untuk memperoleh kinerja sistem yang lebih baik.
3. Pada penelitian berikutnya, diharapkan dapat mengatasi kendala dalam proses input nama rumah makan, di mana nama rumah makan yang dimasukkan harus spesifik dan sesuai dengan informasi yang terdapat pada platform Google Maps. Kesalahan dapat terjadi jika tidak sesuai, dan hal ini perlu diatasi.

DAFTAR PUSTAKA

- Al Farobi, O. (2021). Implementasi Metode Support Vector Machine (SVM) untuk mengetahui respon Masyarakat Indonesia terhadap pemberian vaksin sinovac (Bachelor's thesis, Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta).
- Al-Mejibli, I. S., Alwan, J. K., & Abd Dhafar, H. (2020). The effect of gamma value on support vector machine performance with different kernels. *International Journal of Electrical and Computer Engineering*, 10(5), 5497.
- Amalia, P. R. Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3).
- Anandan, V. (2020). Resample an extremely imbalanced datasets. Retrieved from <https://vijay-anandan.medium.com/how-to-resample-an-imbalanced-datasets-8e413dabbc21>
- Chopra, D., Joshi, N., & Mathur, I. (2016). *Mastering natural language processing with python*. Packt Publishing.
- De, S., & Sirisuriya, S. (2015). *A Comparative Study on Web Scraping*.
- DURAHİM, A. O. (2016). Comparison of sampling techniques for imbalanced learning. *Yönetim Bilişim Sistemleri Dergisi*, 2(2), 181-191.
- Gagić, S., Tešanović, D., & Jovičić, A. (2013). The vital components of restaurant quality that affect guest satisfaction. *Turizam*, 17(4), 166-176.
- Glez-Peña, D., Lourenço, A., López-Fernández, H., Reboiro-Jato, M., & Fdez-Riverola, F. (2013). Web scraping technologies in an API world. *Briefings in Bioinformatics*, 15(5), 788–797. <https://doi.org/10.1093/bib/bbt026>
- Gotama, J. W. (2020). Pengenalan Konsep Pembelajaran Mesin dan Deep learning. *Department of Computer Science, Tokyo Institute of Technology*.

Govindarajan, M. (2014). *Sentiment Analysis Of Restaurant Reviews Using Hybrid Classification Method Proceedings of 2 nd IRF International Conference*.

Gu, Q., Wang, X. M., Wu, Z., Ning, B., & Xin, C. S. (2016). An improved SMOTE algorithm based on genetic algorithm for imbalanced data classification. *Journal of Digital Information Management*, 14(2), 92-103.

Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link. *JOIV: International Journal on Informatics Visualization*, 7(1), 258-264.

Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing* (pp. 878-887). Berlin, Heidelberg: Springer Berlin Heidelberg.

Juwiantho, H., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2020). Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2Vec Menggunakan Deep Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(1), 181-188.

Kumar, Ajitesh . (2023). K-fold cross validation - python example. Retrieved from <https://vitalflux.com/k-fold-cross-validation-python-example/>

Liu, Z., Lin, Y., & Sun, M. (2020). *Representation learning for natural language processing* (p. 334). Springer Nature.

Marsum, W. A. (2005). Restoran dan segala permasalahannya.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mullen, T., & Collier, N. (2004). *Sentiment Analysis using Support Vector Machines with Diverse Information Sources. Sentiment analysis using support vector machines with diverse information sources*.
<https://www.researchgate.net/publication/221012677>

Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). *Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1*. <http://asnugroho.net>

Nurifan, F., Sarno, R., & Sungkono, K. R. (2019). Aspect based sentiment analysis for restaurant reviews using hybrid elmo-wikipedia and hybrid expanded opinion lexicon-senticircle. *International Journal of Intelligent Engineering and Systems*, 12(6), 47-58.

Pamungkas, T. J., & Romadhony, A. (2021). Analisis Sentimen Berbasis Aspek Terhadap Ulasan Restoran Berbahasa Indonesia menggunakan Support Vector Machines. *eProceedings of Engineering*, 8(4).

Pilehvar, M. T., & Camacho-Collados, J. (2020). Embeddings in natural language processing: theory and advances in vector representations of meaning. *Synthesis Lectures on Human Language Technologies*, 13(4), 1-175.

Pulungan, A. F., & Selvida, D. (2022). Kombinasi Metode Sampling pada Pengklasifikasian Data Tidak Seimbang Menggunakan Algoritma SVM. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(2), 277-282.

Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges. *Proceedings of the 5th International Engineering Conference, IEC 2019*, 200–204. <https://doi.org/10.1109/IEC47844.2019.8950616>

Sampaio, Cássia. Understanding SVM Hyperparameters. <https://stackabuse.com/understanding-svm-hyperparameters/>

Sasada, T., Liu, Z., Baba, T., Hatano, K., & Kimura, Y. (2020). A resampling method for imbalanced datasets considering noise and overlap. *Procedia Computer Science*, 176, 420-429.

Sari, F. V., Sari, Fransiska Vina & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 681-686.

Seo, S., Kim, Y., Han, H. J., Son, W. C., Hong, Z. Y., Sohn, I., ... & Hwang, C. (2021). Predicting successes and failures of clinical trials with outer product-based convolutional neural network. *Frontiers in Pharmacology*, 12, 670670.

Trianto, T., Muliawati, A., & Irmanda, H. N. (2022, August). Penerapan Borderline-SMOTE dan Grid Search pada Bagging-SVM untuk Klasifikasi Penyakit Diabetes.

In Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (Vol. 3, No. 2, pp. 610-619).

Yang, M., Xu, J., Luo, K., & Zhang, Y. (2021). Sentiment analysis of Chinese text based on Elmo-RNN model. In *Journal of Physics: Conference Series* (Vol. 1748, No. 2, p. 022033). IOP Publishing

Yustihan, S., Adikara, P., & Indriati, I. (2021). Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(3), 1017-1023.

Zhang, Min., Geng, Guohua., Chen, Jing. (2020). Semi-Supervised Bidirectional Long Short-Term Memory and Conditional Random Fields Model for Named-Entity Recognition Using Embeddings from Language Models Representations, 22. <https://doi.org/10.3390/e22020252>

Zhao, B. (2017). Web Scraping. In *Encyclopedia of Big Data* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_483-1