

**ANALISIS SENTIMEN ULASAN PENGUNJUNG SIMPANG LIMA
GUMUL KEDIRI MENGGUNAKAN METODE BM25 DAN
NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

Inosensius Karelo Hesay
NIM: 175150200111059



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA**

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BRAWIJAYA

MALANG

2021



PENGESAHAN

ANALISIS SENTIMEN ULASAN PENGUNJUNG SIMPANG LIMA GUMUL KEDIRI
MENGUNAKAN METODE BM25 DAN *NEIGHBOR-WEIGHTED*
K-NEAREST NEIGHBOR

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Inosensius Karelo Hesay
NIM: 175150200111059

Skripsi ini telah diuji dan dinyatakan lulus pada
18 Juni 2021

Telah diperiksa dan disetujui oleh:

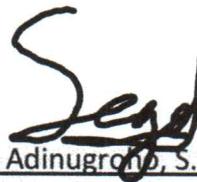
Dosen Pembimbing I



Indriati, S.T., M.Kom.

NIP: 19831013 201504 2 002

Dosen Pembimbing 2



Sigit Adinugroho, S.Kom., M.Sc.

NIK: 201607 880701 1 000

Mengetahui

Ketua Jurusan Teknik Informatika



Achmad Basuki, S.T. M.MG., Ph.D.

NIP: 19741118 200312 1 002

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar referensi.

Apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 18 Juni 2021



Inosensius Karelo Hesay

NIM: 175150200111059

ABSTRAK

Inosensius Karelo Hesay, Analisis Sentimen Ulasan Pengunjung Simpang Lima Gumul Kediri Menggunakan Metode BM25 Dan *Neighbor-Weighted K-Nearest Neighbor*.

Pembimbing: Indriati, S.T., M.Kom. dan Sigit Adinugroho, S.Kom., M.Sc.

Simpang Lima Gumul (SLG) merupakan monumen yang menjadi bangunan ikon serta salah satu destinasi wisata dari daerah Kediri. Para pengunjung yang datang dapat memberikan ulasan pada *Google Review* SLG untuk membantu pihak pengelola mengetahui kekurangan dan kelebihan dari sarana prasarana di SLG. Namun demikian, pihak pengelola belum memiliki sistem yang dapat mengelompokkan ulasan positif dan negatif secara otomatis. Permasalahan tersebut dapat diselesaikan menggunakan sistem analisis sentimen untuk melakukan klasifikasi pada setiap ulasan secara otomatis. Sistem analisis sentimen pada penelitian ini menggunakan kombinasi metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) dan BM25. Tahapan dari sistem analisis sentimen ini meliputi proses *preprocessing* untuk setiap data latih dan data uji, pembobotan TF-IDF, pemeringkatan tiap ulasan menggunakan metode BM25, dan terakhir proses klasifikasi menggunakan metode NWKNN dimana tiap ulasan yang diujikan akan mendapatkan hasil klasifikasinya. Total data yang digunakan pada penelitian ini sejumlah 1000 data, dengan pembagian 800 untuk data latih dan 200 untuk data uji. Penelitian ini melakukan pengujian dengan menggunakan bantuan *5-fold cross validation* untuk menguji parameter nilai k dan nilai eksponen pada metode NWKNN serta nilai k_1 dan b pada metode BM25. Berdasarkan pengujian yang dilakukan pada tiap parameter yang diujikan, didapatkan hasil bahwa nilai terbaik untuk parameter nilai $k_1 = 1,2$, $b = 0,5$, $k = 20$, dan eksponen = 2. Kombinasi dari nilai parameter tersebut menghasilkan nilai rata – rata nilai *precision* sebesar 0,9509, *recall* sebesar 0,9589, *accuracy* sebesar 0,93, dan *f-measure* sebesar 0,9548.

Kata kunci: analisis sentimen, neighbor-weighted k-nearest neighbor, BM25, simpang lima gumul, kediri

ABSTRACT

Inosensius Karelo Hesay, Sentiment Analysis of Visitors Review at Simpang Lima Gumul Kediri using the BM25 and Neighbor-Weighted K-Nearest Neighbor Methods.

Pembimbing: Indriati, S.T., M.Kom. and Sigit Adinugroho, S.Kom., M.Sc.

Simpang Lima Gumul (SLG) is a monument that has become an iconic building as well as one of the tourist destinations of Kediri. Visitors who come can provide reviews on Google Review SLG to help managers know the advantages and disadvantages of facilities at SLG. However, the manager does not yet have a system that can automatically classify positive and negative reviews. This problem can be solved using a sentiment analysis system to automatically classify each review. The sentiment analysis system in this study uses a combination of Neighbor-Weighted K-Nearest Neighbor (NWKNN) and BM25 methods. The stages of this sentiment analysis system include the preprocessing process for each training data and test data, weighting TF-IDF, ranking each review using the BM25 method, and finally the classification process using the NWKNN method where each review tested will get its classification results. The total data used in this study were 1000 data, with the division of 800 for training data and 200 for test data. This study conducted a test using 5-fold cross validation to test the k value and exponential value parameters in the NWKNN method also the k1 and b values in the BM25 method. Based on the tests carried out on each tested parameter, it was found that the best value for the parameter value $k_1 = 1,2$, $b = 0,5$, $k = 20$, and the exponent = 2. The combination of these parameter values produces an average value precision of 0,9509, recall of 0,9589, accuracy of 0,93, and f-measure of 0,9548.

Keywords: sentiment analysis, neighbor-weighted k-nearest neighbor, BM25, simpang lima gumul, kediri

DAFTAR ISI

PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
PRAKATA	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Batasan Masalah	4
1.6 Sistematika Pembahasan	4
BAB 2 LANDASAN KEPUSTAKAAN	6
2.1 Kajian Pustaka	6
2.2 Simpang Lima Gumul	8
2.3 <i>Google Review</i>	9
2.4 Analisis Sentimen	9
2.5 Spesifikasi Sentimen Simpang Lima Gumul	9
2.6 <i>Text Mining</i>	10
2.7 <i>Preprocessing</i>	10
2.7.1 <i>Cleaning</i>	10
2.7.2 <i>Case Folding</i>	10
2.7.3 <i>Tokenizing</i>	11
2.7.4 <i>Filtering</i>	11
2.7.5 <i>Stemming</i>	12
2.8 Metode BM25	13



2.9 Metode <i>Neighbor-Weighted K-Nearest Neighbor</i> (NWKNN)	14
2.10 Evaluasi	15
2.10.1 <i>K-Fold Cross Validation</i>	15
2.10.2 <i>Confusion Matrix</i>	16
BAB 3 METODOLOGI PENELITIAN	18
3.1 Tipe Penelitian	18
3.2 Strategi Penelitian	18
3.3 Partisipan Penelitian	18
3.4 Lokasi Penelitian	18
3.5 Teknik Pengumpulan Data	19
3.6 Data Penelitian	19
3.7 Teknik Analisis Data	19
3.8 Peralatan Pendukung	19
3.9 Implementasi Algoritma	20
3.10 Penarikan Kesimpulan	20
BAB 4 PERANCANGAN	21
4.1 Deskripsi Umum Sistem	21
4.2 Perancangan Proses <i>Training</i>	22
4.2.1 <i>Cleaning</i>	22
4.2.2 <i>Case Folding</i>	23
4.2.3 <i>Tokenizing</i>	24
4.2.4 <i>Filtering</i>	25
4.2.5 <i>Stemming</i>	26
4.2.6 Perhitungan <i>Term Frequency</i>	28
4.3 Perancangan Proses <i>Testing</i>	28
4.3.1 Perhitungan <i>Document Frequency</i>	30
4.3.2 Perhitungan <i>Inverse Document Frequency</i>	31
4.3.3 <i>Scoring</i> BM25	32
4.3.4 Pemingkatan Ulasan	34
4.3.5 Ambil Data Sejumlah Nilai K	36
4.3.6 Pembobotan Kelas	36
4.3.7 <i>Scoring</i> NWKNN	37

4.4	Perhitungan Manual.....	39
4.4.1	Perhitungan Manual <i>Preprocessing</i>	41
4.4.1.1	Cleaning	41
4.4.1.2	Case Folding.....	42
4.4.1.3	Tokenizing.....	42
4.4.1.4	Filtering.....	43
4.4.1.5	Stemming.....	44
4.4.2	Perhitungan Manual Metode BM25.....	45
4.4.2.1	Perhitungan Manual Term Frequency.....	45
4.4.2.2	Perhitungan Manual <i>Document Frequency</i>	48
4.4.2.3	Perhitungan Manual <i>Inverse Document Frequency</i>	49
4.4.2.4	<i>Perhitungan Manual Scoring</i> BM25.....	50
4.4.2.5	Pemeringkatan Ulasan.....	52
4.4.3	Perhitungan Manual Neighbor-Weighted K-Nearest Neighbor	52
4.4.3.1	Mengambil Ulasan Sejumlah Nilai K.....	52
4.4.3.2	Pembobotan Kelas.....	53
4.4.3.3	<i>Scoring</i> NWKNN	53
4.4.4	Perhitungan Manual Nilai <i>Precision, Recall, F-measure, dan Accuracy</i>	54
4.5	Perancangan Skenario Pengujian.....	55
4.5.1	Pengujian Nilai K pada NWKNN	55
4.5.2	Pengujian Nilai Eksponen pada NWKNN.....	56
4.5.3	Pengujian Nilai K1 pada BM25	56
4.5.4	Pengujian Nilai B pada BM25	57
4.5.5	Pengujian Faktor Kesalahan Hasil Klasifikasi.....	58
BAB 5	IMPLEMENTASI	59
5.1	Implementasi <i>Main Program</i>	59
5.2	Implementasi <i>Preprocessing</i>	62
5.2.1	Implementasi <i>Cleaning</i>	62
5.2.2	Implementasi <i>Case Folding</i>	62
5.2.3	Implementasi <i>Tokenizing</i>	63
5.2.4	Implementasi <i>Filtering</i>	63
5.2.5	Implementasi <i>Stemming</i>	64



5.3 Implementasi BM25	65
5.3.1 Implementasi Perhitungan <i>Term Frequency</i>	65
5.3.2 Implementasi Perhitungan <i>Document Frequency</i>	66
5.3.3 Implementasi Perhitungan <i>Inverse Document Frequency</i>	67
5.3.4 Implementasi Perhitungan <i>Scoring</i> BM25	67
5.3.5 Implementasi Pemeringkatan Ulasan	68
5.4 Implementasi NWKNN	69
5.4.1 Implementasi Mengambil Data Sejumlah Nilai K	69
5.4.2 Implementasi Pembobotan Tiap Kelas	70
5.4.3 Implementasi <i>Scoring</i> NWKNN	71
BAB 6 PENGUJIAN DAN ANALISIS	73
6.1 Pengujian Nilai K dengan <i>5-Fold Cross Validation</i>	73
6.1.1 Pengujian <i>Fold</i> ke-1	73
6.1.2 Pengujian <i>Fold</i> ke-2	74
6.1.3 Pengujian <i>Fold</i> ke-3	74
6.1.4 Pengujian <i>Fold</i> ke-4	75
6.1.5 Pengujian <i>Fold</i> ke-5	75
6.1.6 Rata Rata Hasil Pengujian Nilai K	76
6.2 Pengujian Nilai Eksponen dengan <i>5-Fold Cross Validation</i>	77
6.2.1 Pengujian <i>Fold</i> ke-1	77
6.2.2 Pengujian <i>Fold</i> ke-2	77
6.2.3 Pengujian <i>Fold</i> ke-3	78
6.2.4 Pengujian <i>Fold</i> ke-4	79
6.2.5 Pengujian <i>Fold</i> ke-5	79
6.2.6 Rata Rata Hasil Pengujian Nilai Eksponen	80
6.3 Pengujian Nilai K1 dengan <i>5-Fold Cross Validation</i>	80
6.3.1 Pengujian <i>Fold</i> ke-1	80
6.3.2 Pengujian <i>Fold</i> ke-2	81
6.3.3 Pengujian <i>Fold</i> ke-3	81
6.3.4 Pengujian <i>Fold</i> ke-4	82
6.3.5 Pengujian <i>Fold</i> ke-5	82
6.3.6 Rata Rata Hasil Pengujian Nilai K1	83

6.4 Pengujian Nilai B dengan 5-Fold Cross Validation	84
6.4.1 Pengujian <i>Fold</i> ke-1	84
6.4.2 Pengujian <i>Fold</i> ke-2	84
6.4.3 Pengujian <i>Fold</i> ke-3	85
6.4.4 Pengujian <i>Fold</i> ke-4	85
6.4.5 Pengujian <i>Fold</i> ke-5	85
6.4.6 Rata Rata Hasil Pengujian Nilai B	86
6.5 Pengujian Faktor Kesalahan Hasil Klasifikasi	86
6.6 Analisis Pengujian	88
6.6.1 Analisis Pengujian Nilai K	88
6.6.2 Analisis Pengujian Nilai Eksponen	89
6.6.3 Analisis Pengujian Nilai K1	90
6.6.4 Analisis Pengujian Nilai B	91
6.6.5 Analisis Pengujian Faktor Kesalahan Hasil Klasifikasi	92
BAB 7 PENUTUP	93
7.1 Kesimpulan	93
7.2 Saran	93
DAFTAR REFERENSI	94
LAMPIRAN A DATA	96
LAMPIRAN B HASIL KLASIFIKASI SISTEM	99

DAFTAR TABEL

Tabel 2.1 Contoh Tahapan <i>Cleaning</i>	10
Tabel 2.2 Contoh Tahapan <i>Case Folding</i>	11
Tabel 2.3 Contoh Tahapan <i>Tokenizing</i>	11
Tabel 2.4 Contoh Tahapan <i>Filtering</i>	11
Tabel 2.5 Contoh Tahapan <i>Stemming</i>	12
Tabel 2.6 <i>Confusion Matrix</i>	16
Tabel 3.1 Spesifikasi <i>Hardware</i>	19
Tabel 3.2 Spesifikasi <i>Software</i>	19
Tabel 4.1 10 Data Latih	40
Tabel 4.2 1 Data Uji	40
Tabel 4.3 Hasil <i>Cleaning</i> 10 Data Latih	41
Tabel 4.4 Hasil <i>Cleaning</i> 1 Data Uji	41
Tabel 4.5 Hasil <i>Case Folding</i> 10 Data Latih	42
Tabel 4.6 Hasil <i>Case Folding</i> 1 Data Uji	42
Tabel 4.7 Hasil <i>Tokenizing</i> 10 Data Latih	43
Tabel 4.8 Hasil <i>Tokenizing</i> 1 Data Uji	43
Tabel 4.9 Hasil <i>Filtering</i> 10 Data Latih	44
Tabel 4.10 Hasil <i>Filtering</i> 1 Data Uji	44
Tabel 4.11 Hasil <i>Stemming</i> 10 Data Latih	44
Tabel 4.12 Hasil <i>Stemming</i> 1 Data Uji	45
Tabel 4.13 Hasil Perhitungan TF Data Latih	45
Tabel 4.14 Hasil Perhitungan TF Data Uji	48
Tabel 4.15 Hasil Perhitungan DF	49
Tabel 4.16 Hasil Perhitungan IDF	50
Tabel 4.17 Hasil Perhitungan Panjang dan Rata – Rata Ulasan	50
Tabel 4.18 Hasil <i>Scoring</i> BM25 (Ulasan ke-1 – ke-5)	51
Tabel 4.19 Hasil <i>Scoring</i> BM25 (Ulasan ke-6 – ke-10)	51
Tabel 4.20 Hasil Pemeringkatan Ulasan	52
Tabel 4.21 Ulasan Sejumlah Nilai K	52
Tabel 4.22 Jumlah Kelas Positif dan Negatif	53
Tabel 4.23 Hasil Pembobotan Kelas NWKNN	53

Tabel 4.24 Hasil Perhitungan <i>Scoring</i> NWKNN	54
Tabel 4.25 Contoh Hasil Klasifikasi	54
Tabel 4.26 <i>Confusion Matrix</i>	55
Tabel 4.27 Skenario Pengujian Nilai K.....	55
Tabel 4.28 Skenario Pengujian Nilai Eksponen	56
Tabel 4.29 Skenario Pengujian Nilai K_1	57
Tabel 4.30 Skenario Pengujian Nilai B.....	57
Tabel 4.31 Skenario Pengujian Faktor Kesalahan Hasil Klasifikasi.....	58
Tabel 6.1 Hasil Pengujian <i>Fold</i> ke-1 pada Nilai K	73
Tabel 6.2 Hasil Pengujian <i>Fold</i> ke-2 pada Nilai K	74
Tabel 6.3 Hasil Pengujian <i>Fold</i> ke-3 pada Nilai K	74
Tabel 6.4 Hasil Pengujian <i>Fold</i> ke-4 pada Nilai K	75
Tabel 6.5 Hasil Pengujian <i>Fold</i> ke-5 pada Nilai K	75
Tabel 6.6 Rata – Rata Hasil Pengujian Nilai K.....	76
Tabel 6.7 Hasil Pengujian <i>Fold</i> ke-1 pada Nilai Eksponen.....	77
Tabel 6.8 Hasil Pengujian <i>Fold</i> ke-2 pada Nilai Eksponen.....	78
Tabel 6.9 Hasil Pengujian <i>Fold</i> ke-3 pada Nilai Eksponen.....	78
Tabel 6.10 Hasil Pengujian <i>Fold</i> ke-4 pada Nilai Eksponen.....	79
Tabel 6.11 Hasil Pengujian <i>Fold</i> ke-5 pada Nilai Eksponen.....	79
Tabel 6.12 Rata – Rata Hasil Pengujian Nilai Eksponen.....	80
Tabel 6.13 Hasil Pengujian <i>Fold</i> ke-1 pada Nilai K_1	81
Tabel 6.14 Hasil Pengujian <i>Fold</i> ke-2 pada Nilai K_1	81
Tabel 6.15 Hasil Pengujian <i>Fold</i> ke-3 pada Nilai K_1	82
Tabel 6.16 Hasil Pengujian <i>Fold</i> ke-4 pada Nilai K_1	82
Tabel 6.17 Hasil Pengujian <i>Fold</i> ke-5 pada Nilai K_1	83
Tabel 6.18 Rata – Rata Hasil Pengujian Nilai K_1	83
Tabel 6.19 Hasil Pengujian <i>Fold</i> ke-1 pada Nilai B	84
Tabel 6.20 Hasil Pengujian <i>Fold</i> ke-2 pada Nilai B	84
Tabel 6.21 Hasil Pengujian <i>Fold</i> ke-3 pada Nilai B	85
Tabel 6.22 Hasil Pengujian <i>Fold</i> ke-4 pada Nilai B	85
Tabel 6.23 Hasil Pengujian <i>Fold</i> ke-5 pada Nilai B	86
Tabel 6.24 Rata – Rata Hasil Pengujian Nilai B.....	86

Tabel 6.25 Ulasan yang Mendapatkan Hasil Klasifikasi Salah..... 87

Tabel 6.26 Hasil Pengujian Faktor Kesalahan Hasil Klasifikasi..... 87



DAFTAR GAMBAR

Gambar 2.1 Alur <i>Preprocessing</i>	10
Gambar 2.2 Ilustrasi <i>5-fold Cross Validation</i>	16
Gambar 4.1 Diagram Alir Umum Sistem, (a) Proses <i>Training</i> , (b) Proses <i>Testing</i>	21
Gambar 4.2 Diagram Alir Proses <i>Training</i>	22
Gambar 4.3 Diagram Alir Proses <i>Cleaning</i>	23
Gambar 4.4 Diagram Alir Proses <i>Case Folding</i>	23
Gambar 4.5 Diagram Alir Proses <i>Tokenizing</i>	24
Gambar 4.6 Diagram Alir Proses <i>Filtering</i>	25
Gambar 4.7 Diagram Alir Proses <i>Stemming</i>	26
Gambar 4.8 Diagram Alir Proses <i>Term Frequency</i>	28
Gambar 4.9 Diagram Alir Proses <i>Testing</i>	29
Gambar 4.10 Diagram Alir Proses <i>Document Frequency</i>	30
Gambar 4.11 Diagram Alir Proses <i>Inverse Document Frequency</i>	31
Gambar 4.12 Diagram Alir Proses <i>Scoring BM25</i>	33
Gambar 4.13 Diagram Alir Proses <i>Pemeringkatan Ulasan</i>	34
Gambar 4.14 Diagram Alir Proses <i>Pengambilan Data Sejumlah Nilai K</i>	36
Gambar 4.15 Diagram Alir Proses <i>Pembobotan Kelas</i>	36
Gambar 4.16 Diagram Alir Proses <i>Scoring NWKNN</i>	38
Gambar 6.1 Grafik Rata – Rata Pengujian pada Nilai K.....	88
Gambar 6.2 Grafik Rata – Rata Pengujian pada Nilai Eksponen.....	89
Gambar 6.3 Grafik Rata – Rata Pengujian pada Nilai K1.....	90
Gambar 6.4 Grafik Rata – Rata Pengujian pada Nilai B.....	91

DAFTAR LAMPIRAN

LAMPIRAN A DATA.....	96
LAMPIRAN B HASIL KLASIFIKASI SISTEM.....	99



BAB 1 PENDAHULUAN

Bab ini akan menguraikan bagian – bagian yang menjadi dasar penulis dalam melakukan penelitian. Terdiri dari latar belakang permasalahan, rumusan masalah, tujuan serta manfaat dari penelitian, lingkup batasan permasalahan, serta sistematika pembahasan.

1.1 Latar Belakang

Setiap daerah memiliki ciri khas masing – masing yang menunjukkan keistimewaan daerah tersebut. Ciri khas tersebut bisa dalam bentuk bangunan, kuliner, ataupun peninggalan masa lalu. Salah satu contoh bangunan yang menjadi ciri khas / ikon dari suatu tempat adalah Simpang Lima Gumul (SLG), yang merupakan bangunan ikon dari Kabupaten Kediri. Pada mulanya SLG hanyalah bangunan yang dibangun sebagai penanda jalan yang menghubungkan beberapa daerah di Kabupaten Kediri. Namun seiring berjalannya waktu, pemerintah Kabupaten Kediri memperluas dan menambahkan beberapa fasilitas di sekitar area SLG sehingga menjadi ikon serta destinasi wisata ketika berkunjung ke Kediri (Kusuma, 2019).

Perubahan fungsional SLG yang awalnya hanya berupa penanda jalan menjadi destinasi wisata mengharuskan pemerintah Kabupaten Kediri untuk memperhatikan sarana dan prasarana yang ada di sekitar SLG. Hal ini bertujuan untuk memberikan kenyamanan serta menarik pengunjung. Salah satu cara mengetahui kelebihan serta kekurangan suatu tempat adalah dengan meminta pendapat dari pengunjung. Pendapat dari pengunjung dapat mempermudah pengelola dalam menentukan apa yang sudah baik ataupun yang perlu ditingkatkan dari tempat tersebut. Pendapat yang diberikan pengunjung terkait sarana prasarana SLG dapat diketahui melalui ulasan pada *Google Review* SLG. Namun *Google Review* SLG tidak memiliki *filter* untuk membedakan mana ulasan yang bersifat positif ataupun negatif. Semua ulasan bercampur menjadi satu, sehingga cara untuk mengetahui mana ulasan yang merupakan saran dan bersifat positif / negatif adalah dengan membaca secara manual satu persatu untuk menemukan ulasan tersebut. Terkait masalah tersebut, maka diperlukan proses yang dapat mengelompokkan ulasan secara otomatis berdasarkan sifatnya (positif dan negatif) untuk membantu pemerintah Kabupaten Kediri mengetahui mana saran yang berisikan kepuasan ataupun ketidakpuasan terhadap sarana prasarana di SLG dengan lebih cepat menggunakan analisis sentimen.

Pada ulasan *Google Review* selain ulasan terdapat pula *star rating* untuk mewakili tingkat kepuasan penulis, dengan asumsi bahwa semakin banyak *star rating* yang diberikan maka penulis semakin puas. Namun *star rating* pada *Google Review* tidak bisa secara langsung dijadikan acuan bahwa ulasan tersebut merupakan ulasan positif ataupun ulasan negatif. Hal ini disebabkan karena penulis dapat memberikan jumlah *star rating* serta isi ulasan secara bebas, sehingga bisa saja terjadi ketidaksesuaian antara kedua hal tersebut. Oleh karena itu penerapan analisis sentimen pada ulasan yang diambil dari *Google Review*

diperlukan untuk memastikan apakah ulasan tersebut merupakan ulasan positif atau negatif. *Google Review* SLG memiliki *star rating* 4/5 yang diasumsikan bahwa ulasan tersebut tidak seimbang karena ulasan positif lebih banyak dibandingkan ulasan negatif, dimana ulasan positif merepresentasikan kepuasan pengunjung sementara ulasan negatif merepresentasikan ketidakpuasan pengunjung terkait sarana prasarana di SLG.

Pada sistem analisis sentimen ketika membedakan serta memisahkan ulasan berdasarkan kategorinya (positif dan negatif) dapat menggunakan beberapa metode klasifikasi seperti *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-Nearest Neighbor* (KNN), dan sebagainya (Sahara & Wahyudi, 2015). Penelitian oleh Desta & Gurmessa (2019) berisikan tentang perbandingan metode KNN, SVM, dan *Naive Bayes* dalam klasifikasi teks pada data *email*. Penelitian tersebut mendapatkan hasil bahwa metode KNN memiliki akurasi tertinggi dibandingkan metode SVM dan *Naive Bayes*, yaitu sebesar 90,76%. Metode KNN melakukan proses klasifikasi kelas dari data baru dengan mengambil data *training* sejumlah nilai *k* yang memiliki nilai kemiripan paling dekat, kemudian memilih kelas yang paling banyak muncul dari data *training* yang diambil sebagai kelas dari data baru (Pristiyanti, et al., 2018). Meskipun demikian, metode KNN memiliki kelemahan yaitu tidak dapat menangani data tidak seimbang atau tidak terdistribusi secara merata. Permasalahan tersebut memunculkan metode baru yang merupakan pengembangan dari metode KNN, yaitu metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) (Tan, 2005). Secara garis besar metode NWKNN memiliki cara kerja yang mirip dengan KNN, namun dalam metode NWKNN terjadi proses pembobotan yang membuat tiap kelas yang digunakan memiliki nilai bobotnya sendiri. Proses pembobotan pada tiap kelas tersebut merupakan alasan metode NWKNN dapat menangani permasalahan data tidak seimbang (Martha, et al., 2018).

Metode NWKNN dalam awal pengembangannya menggunakan proses pembobotan dan pemeringkatan TF-IDF. Meskipun demikian penelitian oleh Harjanto, et al., (2012) menunjukkan bahwa hasil dari TF-IDF tidak selalu memberikan hasil performa yang baik karena data yang terambil tidak 100% merupakan data yang relevan dengan *query*. Terkait permasalahan tersebut, maka diperlukan metode pembobotan dan pemeringkatan lain yang memiliki kinerja lebih baik dari TF-IDF. Penelitian oleh Yang, et al., (2012) tentang deteksi duplikasi laporan *bug* melakukan pengujian terhadap kinerja TF-IDF dan BM25, pengujian tersebut mendapatkan hasil bahwa metode BM25 lebih unggul dibandingkan TF-IDF dengan perbedaan kinerja tertinggi sebesar 11%. Selain itu penelitian oleh Pardede, et al., (2018) menunjukkan bahwa metode BM25 efektif dalam pemeringkatan dokumen berdasarkan nilai kedekatan dokumen yang diujikan dengan dokumen latih. Hal tersebut terbukti dari nilai *recall* sebesar 100% yang diartikan bahwa semua data yang relevan dengan *query* berhasil terambil.

Berdasarkan permasalahan yang dikemukakan serta penelitian sebelumnya, maka penelitian ini akan berfokus dalam proses pembuatan sistem analisis sentimen terhadap ulasan pengunjung SLG Kediri. Sistem analisis sentimen akan

dibuat menggunakan metode NWKNN sebagai proses klasifikasi karena metode tersebut memiliki kinerja yang lebih baik dibandingkan metode lainnya terutama dalam permasalahan data tidak seimbang serta menggunakan metode BM25 untuk proses pembobotan dan pemeringkatan menggantikan metode TF-IDF karena memiliki kinerja yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dipaparkan pada latar belakang, maka didapatkan rumusan masalah sebagai berikut:

1. Bagaimana hasil pengujian yang diperoleh dari metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri?
2. Bagaimana pengaruh nilai k_1 dan nilai b pada metode BM25 terhadap nilai akurasi dan f -measure pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri?
3. Bagaimana pengaruh nilai k dan nilai eksponen pada metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) terhadap nilai akurasi dan f -measure pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri?

1.3 Tujuan

Tujuan dari diadakannya penelitian ini yaitu:

1. Mengetahui hasil pengujian dari metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri.
2. Menjelaskan pengaruh nilai k_1 dan nilai b pada metode BM25 terhadap nilai akurasi dan f -measure pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri
3. Menjelaskan pengaruh nilai k dan nilai eksponen pada metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) terhadap nilai akurasi dan f -measure pada analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri?

1.4 Manfaat

Adapun manfaat yang didapatkan dari penelitian ini dijabarkan sebagai berikut:

1. Membantu pemerintah Kabupaten Kediri terutama bagian pengelola Simpang Lima Gumul (SLG) untuk mengetahui sarana prasarana apa saja yang sudah layak dan belum layak (perlu ditingkatkan) di sekitar area Simpang Lima Gumul (SLG) dari sudut pandang pengunjung.

2. Ikut berkontribusi dengan memberikan referensi proses analisis sentimen pada ulasan pengunjung Simpang Lima Gumul Kediri menggunakan metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN).
3. Memberikan referensi terkait penggunaan kombinasi metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) dalam proses analisis sentimen kepada peneliti selanjutnya.

1.5 Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yaitu:

1. Sumber data hanya berasal dari *Google Review* Simpang Lima Gumul (SLG) Kediri.
2. Data yang digunakan sejumlah 1000 ulasan tidak seimbang.
3. Data diklasifikasikan ke dalam dua kategori yaitu positif dan negatif.
4. Ulasan yang digunakan terbatas hanya untuk ulasan berbahasa Indonesia.
5. Apabila ditemukan emoji dalam ulasan *Google Review*, maka emoji tersebut tidak digunakan dalam analisis sentimen.
6. Data diambil dalam rentang waktu Januari 2019 – Maret 2021.

1.6 Sistematika Pembahasan

Sistematika pembahasan pada penelitian ini terbagi menjadi tujuh bab dengan beberapa sub bab. Berikut adalah sistematika pembahasan yang ada dalam penelitian ini yaitu:

BAB 1 PENDAHULUAN

Bab pendahuluan menjabarkan tentang latar belakang permasalahan yang diangkat, rumusan masalah, tujuan serta manfaat dari penelitian yang dilakukan, lingkup batasan masalah, dan sistematika pembahasan dari penelitian analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri.

BAB 2 LANDASAN KEPUSTAKAAN

Bab landasan kepastakaan menjabarkan tentang kajian pustaka dan dasar teori berdasarkan penelitian yang telah dilakukan sebelumnya. Tujuan dari adanya berbagai referensi dari penelitian sebelumnya adalah sebagai landasan dalam pembuatan penelitian ini.

BAB 3 METODOLOGI PENELITIAN

Bab metodologi penelitian berisikan uraian tentang bagaimana metodologi yang akan diterapkan dalam perancangan, implementasi, pengujian, dan analisis hasil dari penelitian. Pada bab ini juga terdapat gambaran singkat dari keseluruhan rangkaian tahapan yang dilakukan selama penelitian.

BAB 4 PERANCANGAN

Bab perancangan menjabarkan tentang perancangan dari metode BM25 dan NWKNN pada sistem, contoh perhitungan manual, dan pengujian sistem.

BAB 5 IMPLEMENTASI

Bab implementasi berisikan penjabaran dari metode yang digunakan serta merancang pembuatan kode program berdasarkan perancangan sistem pada bab perancangan.

BAB 6 PENGUJIAN DAN ANALISIS

Bab pengujian dan analisis menjabarkan tentang proses pengujian validasi sistem beserta analisis hasil pengujiannya.

BAB 7 PENUTUP

Bab penutup merupakan bagian akhir yang berisikan hasil kesimpulan dari penelitian yang dikerjakan beserta saran dari penulis untuk penelitian selanjutnya.



BAB 2 LANDASAN KEPUSTAKAAN

Bab ini berisikan konsep dasar serta referensi dari penelitian sebelumnya yang menjadi landasan dari topik penelitian. Konsep dasar tersebut meliputi penggunaan serta implementasi metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) yang nantinya diimplementasikan ke dalam sistem analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri.

2.1 Kajian Pustaka

Penelitian yang dilakukan oleh Desta dan Gurmessa (2019) membahas tentang proses klasifikasi email. Tujuan dari penelitian ini adalah menganalisis serta membandingkan kemampuan metode klasifikasi dalam kasus penyortiran email yang masuk sesuai dengan kelasnya, yaitu email spam dan email ham. Penelitian ini menggunakan tiga jenis metode klasifikasi yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Naive Bayes* (NB) yang kemudian dibandingkan hasil akhirnya untuk melihat metode mana yang mendapatkan hasil terbaik. *Dataset* pada penelitian ini berasal dari repositori *online* dengan total data sejumlah 4061, dimana keseluruhan data tersebut terbagi atas 2 kelas yaitu 1813 data pada kelas email spam dan 2248 data pada kelas email ham. Setelah dilakukan pengujian terhadap ketiga metode yang digunakan, didapatkan hasil akurasi pada metode KNN sebesar 90,76%, SVM sebesar 90,43%, dan NB sebesar 79,28%. Hasil akhir dari akurasi yang didapatkan menunjukkan bahwa metode KNN memiliki kinerja terbaik dalam proses klasifikasi email spam dan email ham dibandingkan metode SVM dan NB.

Penelitian selanjutnya dilakukan oleh Martha, et al., (2018) untuk membandingkan kinerja dari metode klasifikasi *K-Nearest Neighbor* (KNN) dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) pada analisis sentimen data *microblog*. Data yang digunakan diambil menggunakan metode *web crawling* dari beberapa *platform*, yaitu Twitter, Facebook, dan Youtube. Data yang telah didapatkan terbagi atas 5 kelas secara tidak seimbang, yaitu: sangat negatif, negatif, netral, positif, dan sangat positif. Penelitian ini menggunakan 550 data sebagai data latih dengan pembagian 224 data kelas negatif, 169 data kelas positif, 71 data kelas netral, 62 data kelas sangat negatif, dan 24 data kelas sangat positif. Proses pengujian dilakukan dengan menggunakan 50 data uji yang terdiri atas 10 data untuk tiap kelas sesungguhnya. Hasil pengujian didapatkan bahwa pada saat nilai $K=9$ metode NWKNN mendapatkan nilai akurasi tertinggi sebesar 86% karena berhasil mengklasifikasi secara benar 43 dari 50 data uji, sedangkan pada saat $K=3$ metode KNN mendapatkan nilai akurasi tertinggi sebesar 82% karena berhasil mengklasifikasi secara benar 41 dari 50 data uji. Hasil pengujian berdasarkan nilai k mulai dari 1 hingga 15 mendapatkan hasil bahwa metode KNN memiliki hasil akurasi lebih tinggi dibandingkan metode NWKNN sebanyak 3 kali pada nilai $k=3$, $k=4$, dan $k=5$ sementara metode NWKNN lebih tinggi dibandingkan KNN sebanyak 12 kali. Perbandingan hasil dari penelitian tersebut menunjukkan bahwa dengan

data yang tidak seimbang metode NWKNN cenderung memiliki hasil akurasi yang lebih tinggi dibandingkan metode KNN.

Penelitian lainnya mengenai metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) dilakukan oleh Indriati dan Ridok (2016). Penelitian tersebut melakukan analisis terhadap *review* pada aplikasi *mobile* dengan menggunakan data seimbang dan data tidak seimbang. Melihat perbandingan kinerja dari metode *K-Nearest Neighbor* (KNN) serta metode pengembangannya, yaitu *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) merupakan tujuan dari penelitian ini. Data seimbang yang digunakan dalam penelitian ini sejumlah 200 *review* dengan pembagian 100 *review* untuk kelas positif dan 100 *review* untuk kelas negatif, sementara data tidak seimbang sejumlah 550 *review* dengan pembagian 50 *review* untuk kelas positif dan 500 *review* untuk kelas negatif. Perbandingan antara data latih dan data uji yang digunakan pada pengujian adalah 80%-20%. Hasil pengujian menunjukkan bahwa pada data seimbang kinerja dari metode NWKNN lebih baik dari pada metode KNN namun perbedaannya tidak signifikan, hasil pengujian menunjukkan keunggulan metode NWKNN pada nilai *f-measure* dengan metode KNN tidak lebih dari 0,03 dengan nilai tertinggi ketika $k=20$ sebesar 0,9. Sementara pada pengujian data tidak seimbang, terdapat perbedaan kinerja yang signifikan lebih unggul dari metode NWKNN dibandingkan KNN. Metode NWKNN mendapatkan nilai *f-measure* tertinggi pada nilai $k=45$ sebesar 0,797 sementara metode KNN mendapatkan nilai tertinggi sebesar 0,476. Hasil penelitian menunjukkan bahwa metode NWKNN memiliki kinerja yang lebih baik dibandingkan metode KNN terutama ketika data yang digunakan merupakan data yang tidak seimbang.

Penelitian selanjutnya oleh Septrinas, et al., (2019) menggunakan metode BM25 dan *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan berita olahraga. *Dataset* penelitian ini sejumlah 280 data, yang terdiri atas 240 data latih dan 40 data uji. Hasil klasifikasi dari penelitian ini terbagi dalam lima kelas, dimana tiap kelas mewakili bidang olahraga, yaitu sepak bola, bulu tangkis, basket, balap, dan bela diri. Pengujian pada penelitian ini dilakukan menggunakan metode *7-fold cross validation* agar pembagian data untuk tiap pengujian dapat merata. Hasil dari pengujian pada penelitian ini mendapatkan nilai tertinggi ketika nilai $k = 20$, dimana *confusion matrix* yang menghasilkan memiliki nilai *precision* = 0,92157, *recall* = 0,91428, dan nilai *f-measure* = 0,91791 sedangkan nilai terendah didapatkan ketika nilai $k = 200$, yang menghasilkan nilai *precision* = 0,732889, nilai *recall* = 0,714286, dan nilai *f-measure* = 0,699871. Pada proses pengujian ketika nilai $k = 2$ sampai $k = 20$ mengalami perubahan nilai dalam rentang yang rendah, sementara ketika nilai k lebih dari 20 hasil dari pengujian menunjukkan bahwa terjadi penurunan nilai dalam rentang perubahan yang tinggi dan menandakan bahwa nilai k yang digunakan dalam penelitian ini mempengaruhi hasil dari proses klasifikasi. Selain itu dari penelitian ini juga menunjukkan bahwa metode BM25 dapat digunakan bersamaan dengan metode KNN.

Penelitian lain terkait penggunaan serta kinerja metode BM25 dilakukan oleh Pardede, et al., (2018) yang berfokus pada perbandingan metode BM25 dengan metode *Probabilistic Latent Semantic Analysis* (PLSA) dalam pencarian dokumen yang sesuai dengan *query* yang diinginkan. Data yang digunakan pada penelitian ini sejumlah 100 dokumen. Pada penelitian ini proses pengujian dilakukan sebanyak 10 kali dengan pengujian pertama menggunakan 10 dokumen dan terus bertambah sebanyak 10 dokumen untuk tiap pengujian selanjutnya dengan pengujian terakhir menggunakan 100 dokumen. Hasil dari pengujian didapatkan bahwa metode BM25 mendapatkan nilai *precision* yang lebih tinggi dibandingkan metode PLSA. Hal tersebut terjadi karena pencarian dokumen menggunakan BM25 hanya memberikan bobot pada dokumen yang sesuai dengan *query*, sementara PLSA memberikan bobot pada keseluruhan dokumen yang ada. Nilai *recall* pada kedua metode tersebut adalah 100%, yang menandakan bahwa dokumen yang sesuai dapat ditemukan secara menyeluruh. Pada nilai *confusion matrix* yang lain, seperti *f-measure* didapatkan bahwa metode BM25 juga lebih unggul dibandingkan PLSA yaitu sebesar 61,649 dibanding 56,887. Secara garis besar bisa disimpulkan bahwa metode BM25 lebih efektif dalam perangkingan dokumen karena mengutamakan nilai tiap dokumen dan memiliki waktu pemrosesan yang lebih cepat dibandingkan metode PLSA.

Berdasarkan pemaparan dari penelitian sebelumnya dapat dilihat bahwa metode BM25 memiliki kemampuan pembobotan dan pemeringkatan yang baik, serta dapat digunakan untuk membantu proses klasifikasi dalam melakukan pemeringkatan pada data yang digunakan. Sementara metode NWKNN memiliki kinerja yang lebih baik dari pada metode KNN, terbukti dari kemampuannya untuk mendapatkan hasil akurasi yang lebih tinggi ketika data yang digunakan tidak seimbang maupun seimbang. Berdasarkan penelitian sebelumnya oleh Septrinas, et al., (2019) menggunakan kombinasi metode BM25 dan KNN untuk klasifikasi berita olahraga, maka penelitian ini akan menggunakan kombinasi metode BM25 dan metode pengembangan dari KNN, yaitu NWKNN untuk melakukan analisis sentimen terhadap ulasan pengunjung Simpang Lima Gumul Kediri.

2.2 Simpang Lima Gumul

Simpang Lima Gumul (SLG) merupakan bangunan ikon dari Kabupaten Kediri yang diresmikan pada tahun 2008. Pada mulanya SLG hanya digunakan sebagai bangunan penanda jalan transportasi yang menghubungkan 5 kawasan di daerah Kabupaten Kediri. Namun seiring berkembangnya zaman, bangunan ini beralih fungsi menjadi salah satu destinasi wisata di daerah Kediri (Kusuma, 2019). Keseluruhan kawasan SLG adalah sekitar 37 hektar yang terdiri atas beberapa bangunan dan secara khusus SLG memiliki luas 804 meter persegi dengan tinggi bangunan 25 meter. Bentuk dari SLG menyerupai Arc de Triomphe yang ada di Perancis. SLG dibangun karena terinspirasi dari kisah kepahlawanan Prabu Joyoboyo pada masa Kerajaan Kediri (sekitar abad ke-12) yang bertekad untuk menyatukan 5 kawasan di sekitar Kediri. Selain itu relief yang ada pada dinding SLG Kediri juga menceritakan tentang sejarah, kesenian, dan kebudayaan Kediri (Averiani, 2019).

2.3 Google Review

Google Maps memiliki sebuah fitur yang bernama *Google Review*, dimana fitur tersebut dibuat dengan tujuan agar pengguna dapat memberikan ulasan secara bebas pada suatu tempat yang muncul di *Google Maps* (Haq & Rachmat, 2020). Suatu tempat harus mendaftar terlebih dahulu pada *Google My Business* agar tempat tersebut muncul di *Google Maps*. Berdasarkan riset dari Trustklik disebutkan bahwa 70% orang Indonesia lebih mempercayai *review online* dibandingkan keterangan yang tertulis secara resmi (Mubarok, 2018). Siapapun dapat memberikan ulasan pada *Google Review* secara bebas namun ulasan tersebut bersifat publik, sehingga siapapun dapat melihat apa yang tertulis pada ulasan tempat tersebut. Terlebih lagi dalam *Google Review* selain isi ulasan terdapat pula *star rating* yang digunakan sebagai parameter untuk menilai kepuasan penulis. Aspek *star rating* tersebut menjadi pembeda dengan media lain yang digunakan sebagai sarana mengungkapkan pendapat terhadap suatu hal. Penelitian oleh Haq dan Rachmat (2020) menjelaskan bahwa *Google Review* dapat digunakan sebagai evaluasi terkait sarana prasarana dalam sektor pariwisata, karena melalui *Google Review* pihak pengelola dapat mengetahui apa yang sudah baik maupun masih kurang dari tempat pariwisata tersebut berdasarkan sudut pandang pengunjung.

2.4 Analisis Sentimen

Analisis sentimen merupakan salah satu cabang ilmu pengetahuan dari *data mining* yang berfokus untuk menganalisis serta mengolah informasi tekstual yang berbentuk opini publik pada suatu objek tertentu (Novantiran, et al., 2015). Analisis sentimen secara umum kerap digunakan untuk menilai kesukaan atau ketidaksukaan publik terhadap suatu hal tertentu. Secara garis besar analisis sentimen berfokus pada opini yang mengekspresikan sentimen positif dan negatif. Nilai ekspresi berupa positif atau negatif yang didapat pada analisis sentimen dapat digunakan oleh masyarakat sebagai acuan untuk mengambil suatu keputusan kedepannya (Indriati & Ridok, 2016).

2.5 Spesifikasi Sentimen Simping Lima Gumul

Sentimen SLG yang digunakan berasal dari *Google Review* SLG, dimana sentimen tersebut tidak seimbang (ulasan positif lebih banyak dari pada negatif) serta hanya fokus pada sentimen yang berisikan saran terhadap sarana prasarana SLG. Ulasan *Google Review* tidak memiliki aturan tertulis dalam penulisan ulasan, sehingga setiap orang dapat menulis ulasan terkait apapun dan dengan format penulisan apapun sesuai keinginan hati penulis. Selain itu jumlah *star rating* pada tiap ulasan tidak bisa secara langsung menjadi acuan bahwa sentimen SLG tersebut bersifat positif atau negatif, serta secara spesifik tidak bisa menunjukkan bahwa ulasan tersebut berisikan saran terkait sarana prasarana SLG. Hal tersebut terjadi karena penulis dapat memberikan nilai *star rating* dan isi ulasan secara bebas, sehingga bisa saja terjadi ketidaksesuaian antara kedua hal tersebut.

2.6 Text Mining

Text mining merupakan bagian dalam proses penambangan data (*data mining*), dimana sumber data berasal dari sekumpulan teks yang tidak terstruktur. Tujuan proses *text mining* adalah untuk mendapatkan kata – kata yang mampu mewakili keseluruhan dokumen (Putri, et al., 2020). Secara umum proses dari *text mining* terdiri atas beberapa tahapan, yaitu: pengumpulan data, pembersihan teks (*cleaning*), pemrosesan teks, analisis pola, dan terakhir adalah mengekstrak informasi dari teks (Talib, et al., 2016).

2.7 Preprocessing

Pada tahapan *preprocessing* setiap dokumen yang didapatkan akan diolah untuk mendapatkan *term* atau token yang dianggap penting dan mewakili dokumen tersebut. Secara garis besar tujuan dari *preprocessing* adalah untuk mempermudah proses komputasi dengan cara melakukan pemecahan suatu dokumen untuk diolah lebih lanjut (Septrinas, et al., 2019). *Preprocessing* terdiri atas beberapa tahapan, yaitu: *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Alur proses *preprocessing* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Alur Preprocessing

2.7.1 Cleaning

Tahapan *cleaning* merupakan proses yang bertujuan untuk membersihkan kalimat dari karakter yang dirasa tidak perlu untuk ikut serta dalam tahapan berikutnya seperti tanda baca, angka, simbol, dan *emoticon* (Pradhana, et al., 2020). *Google Review* bukanlah sebuah forum formal, sehingga tahapan ini diperlukan untuk menghilangkan karakter yang dianggap tidak diperlukan karena tidak memiliki nilai dalam mewakili isi dari suatu kalimat. Contoh penerapan dari tahapan *cleaning* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Contoh Tahapan *Cleaning*

Sebelum proses <i>cleaning</i>	Setelah proses <i>cleaning</i>
Tolong bapak pemkab Kediri, ada 3 tempat parkir, dan 2 kamar mandi masih kotor. Tolong dibersihkan!	Tolong bapak pemkab Kediri ada tempat parkir dan kamar mandi masih kotor Tolong dibersihkan

2.7.2 Case Folding

Tahapan *case folding* bertujuan untuk merubah huruf kapital yang ada dalam kalimat menjadi huruf kecil (*lowercase*). Hal ini diperlukan karena tidak keseluruhan data teks konsisten dalam penggunaan huruf (Pristiyanti, et al., 2018). Sebagai contoh apabila tidak dilakukan proses *case folding* dan

menemukan dua kata yang sama namun secara struktur karakter berbeda (kata pertama huruf kecil semua, sedangkan kata kedua terdapat 1 karakter huruf kapital) akan dianggap berbeda. Contoh penerapan dari tahapan *case folding* dapat dilihat pada Tabel 2.2.

Tabel 2.2 Contoh Tahapan Case Folding

Sebelum proses <i>case folding</i>	Setelah proses <i>case folding</i>
Tolong bapak pemkab Kediri ada tempat parkir dan kamar mandi masih kotor Tolong dibersihkan	tolong bapak pemkab kediri ada tempat parkir dan kamar mandi masih kotor tolong dibersihkan

2.7.3 Tokenizing

Tahapan *tokenizing* merupakan proses dimana terjadi pemisahan masing – masing kata yang ada pada kalimat untuk menjadi sebuah token. Pemisah yang digunakan untuk proses ini adalah spasi. Secara garis besar proses ini akan memotong kalimat berdasarkan kata penyusunnya (Septrinas, et al., 2019). Contoh penerapan tahapan *tokenizing* dapat dilihat pada Tabel 2.3.

Tabel 2.3 Contoh Tahapan Tokenizing

Sebelum proses <i>tokenizing</i>	Setelah proses <i>tokenizing</i>
tolong bapak pemkab kediri ada tempat parkir dan kamar mandi masih kotor tolong dibersihkan	'tolong', 'bapak', 'pembkab', 'kediri', 'ada', 'tempat', 'parkir', 'dan', 'kamar', 'mandi', 'masih', 'kotor', 'tolong', 'dibersihkan'

2.7.4 Filtering

Tahapan *filtering* juga bisa disebut *stopword removal*. Tujuan dari proses ini adalah menghilangkan kata atau token hasil dari proses *tokenizing* yang dianggap tidak penting. Penghilangan kata atau token pada proses *filtering* menggunakan bantuan dari *stoplist*. *Stoplist* adalah sekumpulan kata yang dianggap tidak penting dan tidak memiliki arti (Indriati & Ridok, 2016). Cara penghilangan kata adalah dengan cara membandingkan setiap token hasil proses *tokenizing* dengan kata yang ada pada *stoplist*. Apabila token tersebut terdaftar pada *stoplist*, maka token tersebut akan dihilangkan dari kumpulan token. Penelitian ini menggunakan bantuan *stoplist* Tala sebagai acuan untuk mengetahui mana kata yang dianggap tidak penting. Contoh penerapan tahapan *filtering* dapat dilihat pada Tabel 2.4.

Tabel 2.4 Contoh Tahapan Filtering

Sebelum proses <i>filtering</i>	Setelah proses <i>filtering</i>
tolong	tolong
bapak	bapak
pembkab	pembkab

Tabel 2.4 Contoh Tahapan *Filtering* (Lanjutan)

Sebelum proses <i>filtering</i>	Setelah proses <i>filtering</i>
kediri	kediri
ada	-
tempat	tempat
parkir	parkir
dan	-
kamar	kamar
mandi	mandi
masih	-
kotor	kotor
tolong	tolong
dibersihkan	dibersihkan

2.7.5 Stemming

Tahapan *stemming* merupakan proses akhir dari proses *preprocessing*, tujuan dari *stemming* adalah untuk mendapatkan bentuk dasar dari sebuah kata (Pradhana, et al., 2020). Proses *stemming* dilakukan dengan cara melakukan pengecekan untuk tiap token, apabila token merupakan sebuah kata yang mempunyai imbuhan (baik pada awal kata, akhir kata, atau campuran) maka proses *stemming* dilakukan pada token tersebut untuk mendapatkan kata dasarnya. Melalui proses *stemming* setiap kata yang sebenarnya serupa namun memiliki imbuhan berbeda bisa terhitung menjadi 1 jenis token saja, sehingga dapat mengurangi jumlah token yang ada. Contoh penerapan tahapan *stemming* dapat dilihat pada Tabel 2.5.

Tabel 2.5 Contoh Tahapan *Stemming*

Sebelum proses <i>stemming</i>	Setelah proses <i>stemming</i>
tolong	tolong
bapak	bapak
pemkab	pemkab
kediri	kediri
ada	-
tempat	tempat
parkir	parkir
dan	-

Tabel 2.5 Contoh Tahapan *Stemming* (Lanjutan)

Sebelum proses <i>stemming</i>	Setelah proses <i>stemming</i>
kamar	kamar
mandi	mandi
masih	masih
kotor	kotor
tolong	tolong
dibersihkan	bersih

2.8 Metode BM25

Metode BM25 merupakan formula terbaik dalam kelas *best match* karena menghasilkan keefektifan dan ketepatan yang tinggi ketika menghitung kesamaan antara *query* dan dokumen (Pardede, et al., 2018). Penelitian oleh Pardede, et al., (2018) menjelaskan bahwa metode BM25 lebih baik dari PLSA karena memiliki kinerja yang lebih baik dalam mendapatkan data sesuai dengan *query* yang diminta. Sementara penelitian oleh Yang, et al., (2012) juga menjelaskan bahwa metode BM25 memiliki kinerja yang lebih baik dari TF-IDF dalam menemukan data yang relevan dari sekumpulan data yang ada. Langkah pertama dalam metode BM25 adalah melakukan perhitungan bobot *term*, proses ini didapat dari perhitungan nilai TF (*Term Frequency*), IDF (*Inverse Document Frequency of term*), dan panjang dokumen. Setelah itu dilakukan pemeringkatan dari hasil pembobotan tersebut dimulai dari yang memiliki nilai terbesar hingga yang terkecil. Berdasarkan penelitian oleh Septrinas, et al., (2019), rumus perhitungan BM25 dapat dilihat pada Persamaan 2.1.

$$BM25 = \sum_{i=1}^n idf(q_i) \frac{tf(q_i, D)(k_1+1)}{tf(q_i, D)+k_1(1-b+b \frac{|D|}{avgdl})} \quad (2.1)$$

Keterangan:

$tf(q_i, D)$ = Banyak *term frekuensi* q_i pada dokumen

k_1 = Bernilai antara 1,2 – 2,0

b = Bernilai antara 0,5 – 0,8

$|D|$ = Panjang dokumen

$avgdl$ = Rata – rata panjang dari keseluruhan dokumen

Pada metode BM25 terdapat pula rumus IDF yang dijabarkan pada Persamaan 2.2.

$$idf(q_i) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} \quad (2.2)$$

Keterangan:

$df(q_i)$ = Document frequency dari kata q_i

N = Total data latih yang digunakan

2.9 Metode Neighbor-Weighted K-Nearest Neighbor (NWKNN)

Distribusi data latih yang digunakan pada suatu penelitian untuk setiap kategori yang ada memiliki kecenderungan untuk tidak terbagi secara rata. Hal ini dapat menyebabkan kesalahan pada hasil klasifikasi. Untuk mengatasi permasalahan tersebut dapat menggunakan metode *Neighbor-Weighted K-Nearest Neighbor* (NWKNN). Metode NWKNN tercipta dari pengembangan metode KNN yang memiliki kekurangan dalam mengatasi data tidak seimbang dimana metode NWKNN bekerja dengan menerapkan prinsip pembobotan (Indriati & Ridok, 2016). Proses pembobotan tersebut menjadi pembeda antara KNN dan NWKNN. Proses tersebut bertujuan untuk memberikan nilai antar ketetangaan data yang digunakan, dimana nilainya didapatkan berdasarkan nilai bobot kelas data tersebut. Penelitian yang dilakukan oleh Martha, et al., (2018) juga mendukung bahwa metode NWKNN memiliki kinerja yang lebih baik dibandingkan KNN serta dapat mengatasi permasalahan pembagian data yang tidak seimbang karena menggunakan proses pembobotan untuk kelas yang digunakan, sehingga data dengan kelas mayoritas tidak mendominasi hasil klasifikasi. Kelas yang mempunyai jumlah data lebih banyak akan memiliki nilai bobot yang kecil dan kelas yang jumlah datanya lebih sedikit akan memiliki nilai bobot yang lebih besar. Berdasarkan penelitian oleh Indriati & Ridok (2016), dapat diuraikan langkah – langkah perhitungan metode NWKNN adalah sebagai berikut:

1. Menghitung nilai bobot untuk mendapatkan nilai kemiripan kemudian diurutkan berdasarkan yang paling besar ke yang paling kecil. Pada penelitian ini untuk menghitung nilai bobot dan proses pemeringkatan menggunakan metode BM25.
2. Menentukan bobot tiap kategori kelas, penjabaran rumus dapat dilihat pada Persamaan 2.3.

$$Weight_i = \frac{1}{\left(\frac{Num(C_i^d)}{\min\{Num(C_i^d) | j = 1, \dots, K^*\}} \right)^{\frac{1}{exp}}} \quad (2.3)$$

Keterangan:

$Weight_i$ = Bobot kelas ke – i

$Num(C_i^d)$ = Total data latih pada kelas ke – i

$Num(C_i^d)$ = Banyak data latih d pada kelas l , dimana l menyesuaikan berapa banyak kategori kelas yang ada

exp = Eksponen, dimana nilainya > 1

- Menghitung nilai skor untuk mendapatkan hasil klasifikasi, penjabaran rumus untuk menghitung skor dapat dilihat pada Persamaan 2.4.

$$score(d, c_i) = Weight_i \left(\sum_{d_j \in KNN(q)} Sim(q, d_j) \delta(d_j, c_i) \right) \quad (2.4)$$

Keterangan:

$score(d, c_i)$ = Hasil perhitungan skor dari data uji d pada kategori kelas c_i

$Weight_i$ = Bobot kelas ke – i

$d_j \in KNN(q)$ = Data latih d_j yang berada pada kumpulan tetangga terdekat dari data uji

$Sim(q, d_j)$ = Nilai similaritas antara data uji dengan data latih

$\delta(d_j, c_i)$ = Jika data latih termasuk dalam kelas c_i maka bernilai 1, jika bukan maka bernilai 0

c_i = Kelas / kategori ke – i

2.10 Evaluasi

Hasil klasifikasi dapat diuji menggunakan metode pengujian dimana akan diukur tingkat akurasi dari sistem yang dibuat (Pristiyanti, et al., 2018). Terdapat beberapa metode untuk menguji hasil klasifikasi dari sistem. Berikut adalah metode yang digunakan dalam penelitian ini yaitu: pengujian menggunakan *K-fold cross validation* dan pengujian menggunakan *confusion matrix* untuk menghitung nilai *precision*, *recall*, *f-measure*, dan *accuracy*.

2.10.1 K-Fold Cross Validation

K-fold cross validation merupakan sebuah metode yang digunakan untuk membandingkan dan mengevaluasi metode yang digunakan dengan cara membagi *dataset* menjadi dua bagian yaitu data uji dan data latih (Pradhana, et al., 2020). Cara kerja metode ini adalah dengan membagi *dataset* sejumlah *fold* yang digunakan dengan syarat 1 bagian adalah untuk data uji sementara bagian sisanya untuk data latih. Pada setiap pengujian *fold*, data uji yang digunakan tidak boleh sama antara *fold* yang satu dengan lainnya (Putri, et al., 2020). Pengujian pada penelitian ini menggunakan 5 *fold* (jumlah keseluruhan data dibagi 5 secara rata, kemudian 1 bagian untuk data uji dan sisanya untuk data latih) atau lebih umum dikenal dengan istilah *5-fold cross validation*. Ilustrasi dari *5-fold cross validation* dapat dilihat pada Gambar 2.2.



Gambar 2.2 Ilustrasi 5-fold Cross Validation

Sumber: (Mujtaba, 2020)

2.10.2 Confusion Matrix

Confusion matrix berisikan informasi hasil proses evaluasi dari kelas sebenarnya (*actual class*) dan kelas hasil prediksi (*predicted class*) yang dilakukan oleh metode klasifikasi (Pradhana, et al., 2020). *Confusion matrix* terbagi atas 4 bagian, yaitu: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Ilustrasi dari *confusion matrix* dapat dilihat pada Tabel 2.6.

Tabel 2.6 Confusion Matrix

Confusion Matrix		Predicted	
		True	False
Actual	True	TP	FN
	False	FP	TN

Keterangan:

TP = jumlah data yang memiliki kelas aktual positif yang mendapatkan hasil klasifikasi sebagai data positif.

FP = jumlah data yang memiliki kelas aktual negatif yang mendapatkan hasil klasifikasi sebagai data positif.

FN = jumlah data yang memiliki kelas aktual positif yang mendapatkan hasil klasifikasi sebagai data negatif.

TN = jumlah data yang memiliki kelas aktual negatif yang mendapatkan hasil klasifikasi sebagai data negatif.

Melalui *confusion matrix* dapat pula menghitung nilai *precision*, *recall*, *f-measure*, dan *accuracy* untuk mengevaluasi hasil klasifikasi serta melihat performa dari sistem yang dibuat. Berdasarkan penelitian dari Pristiyantri, et al., (2018) didapatkan penjelasan sebagai berikut:

Precision adalah tingkat ketepatan sistem untuk memberikan data yang sesuai berdasarkan apa yang diminta oleh pengguna. Rumus menghitung nilai *precision* dapat dilihat pada Persamaan 2.5.

$$Precision = \frac{TP}{TP+FP} \quad (2.5)$$

Recall adalah tingkat keberhasilan sistem dalam menemukan data yang relevan dari sekumpulan data yang ada. Rumus menghitung nilai *recall* dapat dilihat pada Persamaan 2.6.

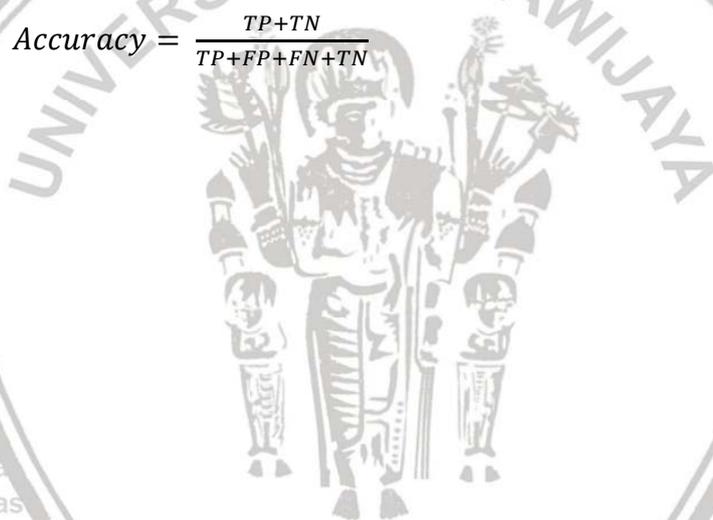
$$Recall = \frac{TP}{TP+FN} \quad (2.6)$$

F-measure adalah nilai rata – rata harmonik yang didapatkan dari nilai *precision* dan *recall*. Rumus menghitung nilai *f-measure* dapat dilihat pada Persamaan 2.7.

$$F - measure = \frac{2 \times precision \times recall}{precision+recall} \quad (2.7)$$

Accuracy adalah tingkat kedekatan antara nilai prediksi dengan nilai yang sebenarnya. Rumus menghitung nilai *accuracy* dapat dilihat pada Persamaan 2.8.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2.8)$$



BAB 3 METODOLOGI PENELITIAN

Bab ini berisikan uraian tentang bagaimana metodologi yang akan diterapkan dalam perancangan, implementasi, pengujian, dan analisis hasil dari penelitian. Isi dari bab ini meliputi tipe penelitian, strategi, partisipan penelitian, lokasi penelitian, teknik yang digunakan dalam pengumpulan data, sumber data yang digunakan, teknik yang digunakan untuk menganalisis data hingga penyusunan jadwal penelitian.

3.1 Tipe Penelitian

Jenis penelitian analisis sentimen pada ulasan pengunjung Simpang Lima Gumul Kediri merupakan tipe penelitian non-implementatif analitik. Penelitian non-implementatif analitik berarti penelitian ini bertujuan untuk menyelesaikan suatu permasalahan dimana penelitian ini juga menjelaskan hubungan antara objek yang diteliti dengan metode penyelesaiannya, dalam hal ini objek yang dimaksud adalah SLG Kediri dan metode penyelesaiannya yaitu analisis sentimen.

3.2 Strategi Penelitian

Strategi penelitian dari penelitian ini dengan cara studi kasus. Studi kasus dilakukan berdasarkan ulasan yang ditulis oleh pengunjung SLG pada *Google Review*. Strategi penelitian dilakukan dengan diawali proses pengumpulan data, memisahkan data sesuai kelasnya sebagai data latih, dilanjutkan proses *pre-processing*, melakukan pembobotan dan pemeringkatan pada data uji dan diklasifikasikan untuk mendapatkan kelas tertentu.

3.3 Partisipan Penelitian

Partisipan dalam penelitian ini adalah seluruh orang yang memiliki akun *Google* dan memberikan ulasan pada SLG di *Google Review*. Dimana diasumsikan bahwa orang yang memberikan ulasan sudah pernah mengunjungi SLG setidaknya minimal satu kali. Ulasan yang diambil berfokus pada ulasan yang menunjukkan kepuasan dan ketidakpuasan terhadap sarana prasarana di area SLG.

Proses pemberian label positif dan negatif pada keseluruhan data yang digunakan melibatkan lima orang penduduk Kediri dan pernah mengunjungi SLG Kediri sebagai pakar. Lima orang tersebut bernama Ade Nur Ashari, Annisa Fitri Rahmawati, Cindy Rezma Fanny, Dea Panca Wardana, dan Veronica Deandra Christabella. Penentuan label pada data yang digunakan dilakukan dengan pemberian label pada keseluruhan data oleh tiap pakar, dimana label terbanyak yang diberikan oleh pakar untuk tiap data akan menjadi label dari data tersebut.

3.4 Lokasi Penelitian

Lokasi penelitian bertempat di rumah penulis, yaitu Jl Letjend S Parman No.4 Kota Kediri.

3.5 Teknik Pengumpulan Data

Pada penelitian analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri data yang digunakan adalah data primer. Teknik pengumpulan data yang digunakan pada penelitian ini dilakukan secara manual dengan cara mengamati dan mengambil tiap ulasan pada *Google Review* Simpang Lima Gumul Kediri. Data yang digunakan berasal dari ulasan SLG Kediri sejak bulan Januari 2019 hingga Maret 2021.

3.6 Data Penelitian

Data penelitian berasal dari ulasan pada *Google Review* Simpang Lima Gumul Kediri. Data yang diambil hanya ulasan berbahasa Indonesia yang berisikan tentang kepuasan dan ketidakpuasan dari sarana prasarana di sekitar area Simpang Lima Gumul Kediri. Keseluruhan data yang diambil sejumlah 1000 data yang terdiri dari 800 data dengan sentimen positif dan 200 data sentimen negatif. Dimana data tersebut terbagi menjadi 800 data latih dan 200 data uji.

3.7 Teknik Analisis Data

Teknik analisis data dalam penelitian ini berdasarkan hasil yang diperoleh pada hasil perhitungan menggunakan metode BM25 dan metode klasifikasi *Neighbor-Weighted K-Nearest Neighbor* (NWKNN). Pengujian pada penelitian ini menggunakan metode *k-fold cross validation* dan *confusion matrix* dengan evaluasi nilai *precision*, *recall*, *f-measure*, dan *accuracy*.

3.8 Peralatan Pendukung

Dalam pelaksanaan penelitian ini, peralatan pendukung yang digunakan yaitu *hardware* dan *software* dengan spesifikasi sebagai berikut:

Tabel 3.1 Spesifikasi Hardware

Spesifikasi	Keterangan
Laptop	Acer F5-572G-5105
CPU	Intel Core i5-6200U
GPU	NVIDIA GeForce 920M
RAM	8GB DDR3
Storage	1000GB HDD

Tabel 3.2 Spesifikasi Software

Jenis	Keterangan
Sistem Operasi (OS)	Windows 10
Bahasa pemrograman	Python 3.7.6

Tabel 3.2 Spesifikasi Software (Lanjutan)

Jenis	Keterangan
Text editor	Visual Studio Code
Library	Sastrawi, Pandas, Counter, RE, Math, Sys

3.9 Implementasi Algoritma

Implementasi dari metode BM25 dan *Neighbor-Weighted K-Nearest Neighbor* (NWKNN) pada penelitian analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri secara garis besar dapat digambarkan sebagai berikut:

1. Memasukkan keseluruhan data berupa ulasan pengunjung SLG.
2. Melakukan *preprocessing* pada keseluruhan data.
3. Melakukan pembobotan dan pemeringkatan pada ulasan hasil *preprocessing* dengan menggunakan metode BM25.
4. Melakukan klasifikasi terhadap data uji menggunakan metode NWKNN.
5. Menghasilkan keluaran berupa klasifikasi terhadap data uji berupa kelas positif atau negatif.

3.10 Penarikan Kesimpulan

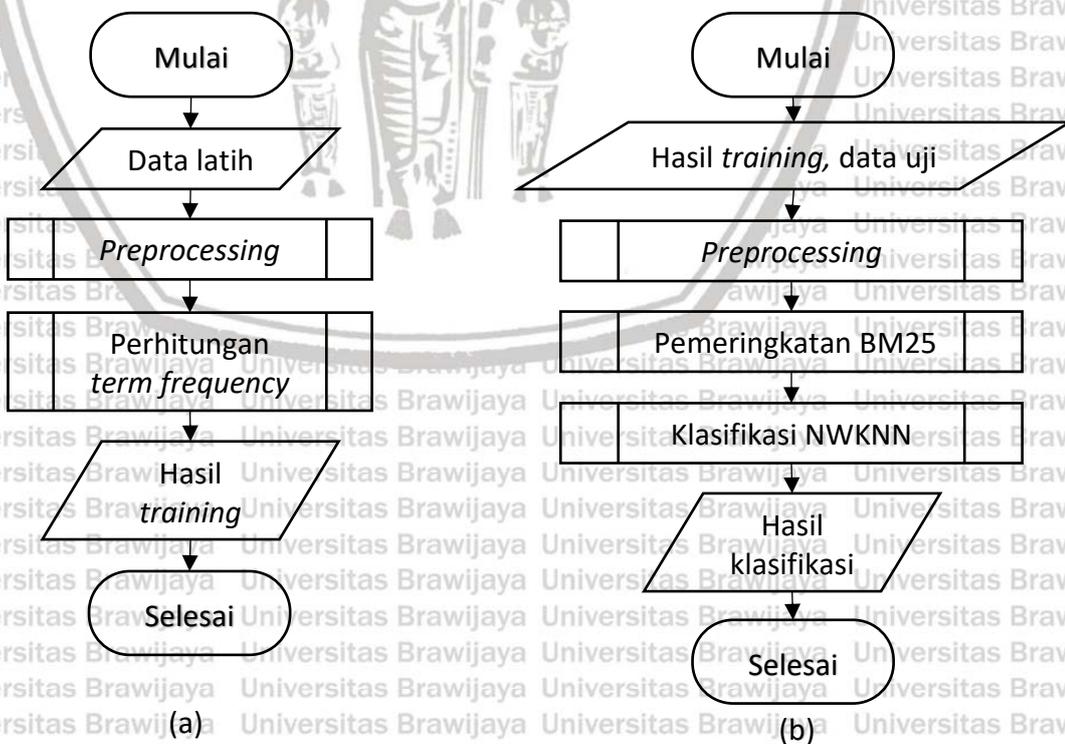
Pada tahap ini dilakukan penarikan kesimpulan terhadap hasil dari penelitian yang telah dilakukan. Proses ini akan memaparkan hasil dari sistem analisis sentimen dengan menggunakan metode BM25 dan NWKNN, pemaparan tersebut meliputi persentase akurasi hasil pengujian yang merepresentasikan kinerja dari sistem serta penjelasan terkait pengaruh parameter pada metode BM25 dan NWKNN terhadap sistem yang dibuat.

BAB 4 PERANCANGAN

Bab ini membahas terkait perancangan dari sistem dan proses manual dari metode yang digunakan dalam analisis sentimen terhadap ulasan pengunjung SLG menggunakan metode BM25 dan NWKNN.

4.1 Deskripsi Umum Sistem

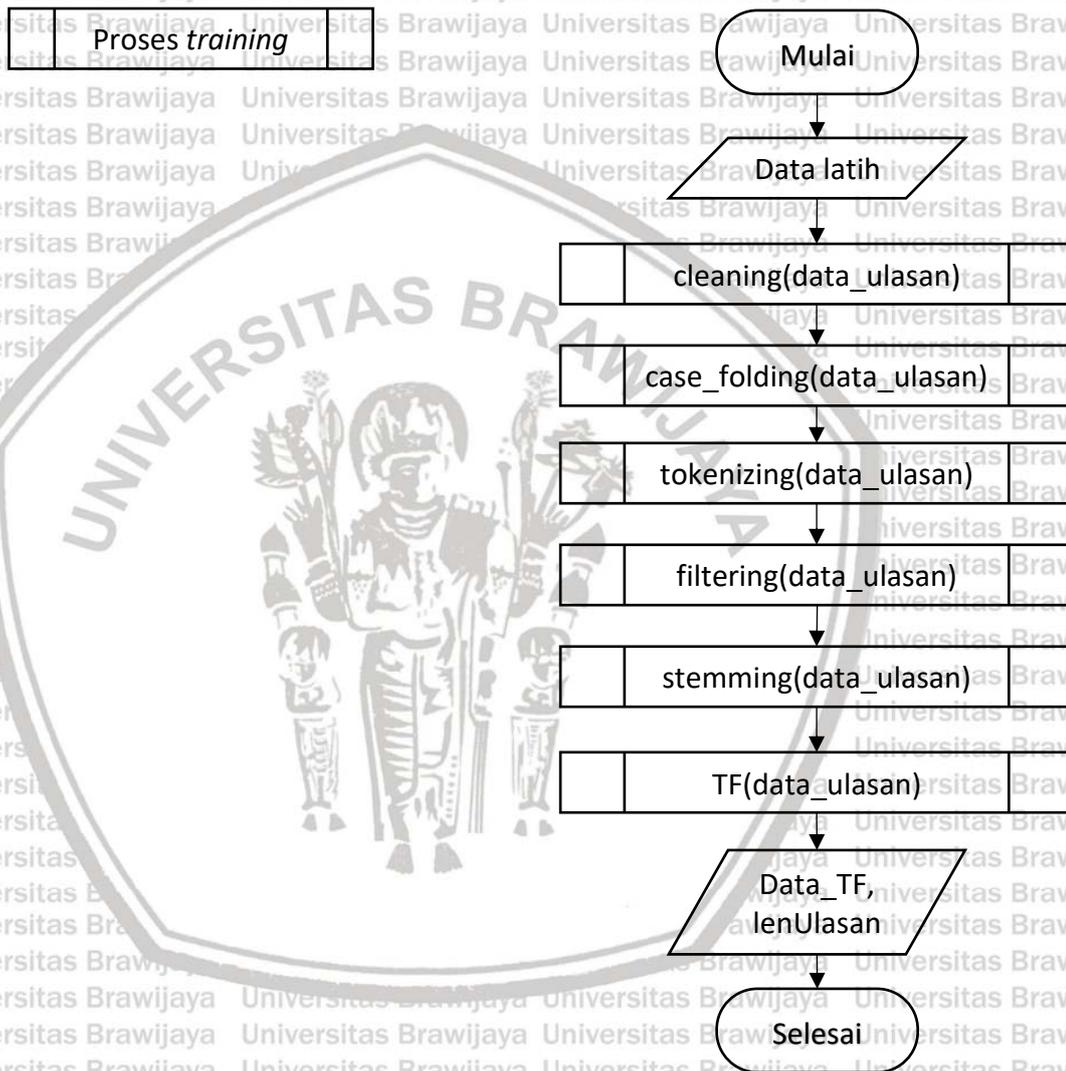
Deskripsi umum sistem berisikan penjabaran terkait tahapan dari sistem yang dibangun. Sistem yang dibangun yaitu sebuah sistem analisis sentimen terhadap ulasan pengunjung SLG Kediri yang didapatkan dari *Google Review* menggunakan metode BM25 untuk pembobotan dan pemeringkatan antar data serta metode NWKNN untuk proses klasifikasi. Klasifikasi pada sistem ini dibedakan menjadi 2, yaitu kelas positif dan negatif. Sistem ini juga terdiri atas 2 proses, yaitu proses *training* dan *testing*. Pada proses *training*, data latih yang digunakan akan terlebih dahulu dilakukan *preprocessing* untuk memilah kata yang nantinya digunakan sebagai *term*. Setelah itu dilanjutkan dengan perhitungan *term frequency* (TF) untuk mendapatkan jumlah *term* pada tiap data yang nantinya digunakan pada perhitungan metode BM25 dan NWKNN. Sementara pada proses *testing*, data uji yang digunakan juga melalui *preprocessing* terlebih dahulu. Setelah itu dilanjutkan perhitungan nilai *document frequency* (DF) dan *inverse document frequency* (IDF) yang merupakan bagian dari proses pembobotan dan pemeringkatan menggunakan metode BM25. Kemudian dilakukan proses klasifikasi menggunakan metode NWKNN untuk mengetahui hasil klasifikasi pada tiap data uji. Diagram alir dari perancangan umum sistem dapat dilihat pada Gambar 4.1.



Gambar 4.1 Diagram Alir Umum Sistem, (a) Proses *Training*, (b) Proses *Testing*

4.2 Perancangan Proses Training

Proses *training* pada sistem akan melakukan pengolahan pada data latih yang berupa ulasan pengunjung SLG Kediri menjadi kumpulan *term* dimana *term* tersebut akan digunakan sebagai perbandingan kemiripan dengan data uji. Tahapan pada proses *training* diawali dengan *preprocessing* yang terdiri atas *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming* serta melakukan perhitungan *term frequency* untuk setiap *term* hasil dari *preprocessing*. Diagram alir dari proses *testing* dapat dilihat pada Gambar 4.2.

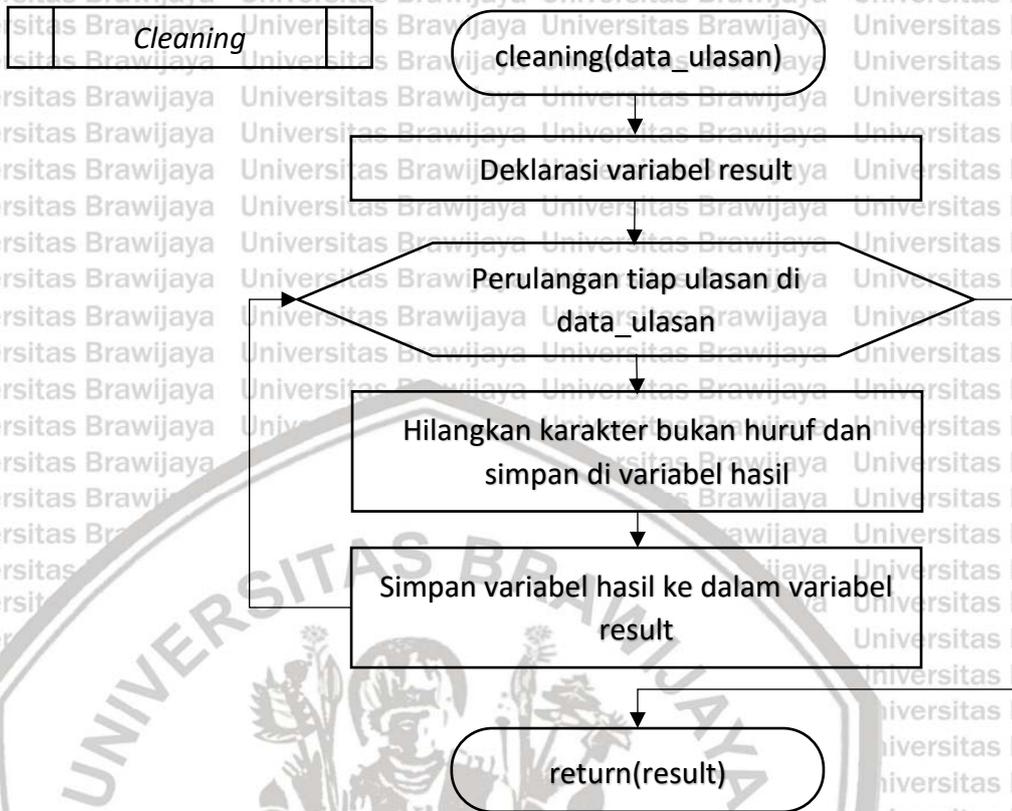


Gambar 4.2 Diagram Alir Proses Training

4.2.1 Cleaning

Proses *cleaning* merupakan langkah pertama dalam *preprocessing*. Pada proses ini akan dilakukan pengecekan untuk tiap ulasan yang digunakan untuk menghilangkan karakter atau simbol yang dirasa tidak perlu karena tidak memiliki bobot nilai yang menentukan apakah suatu ulasan bernilai positif atau negatif.

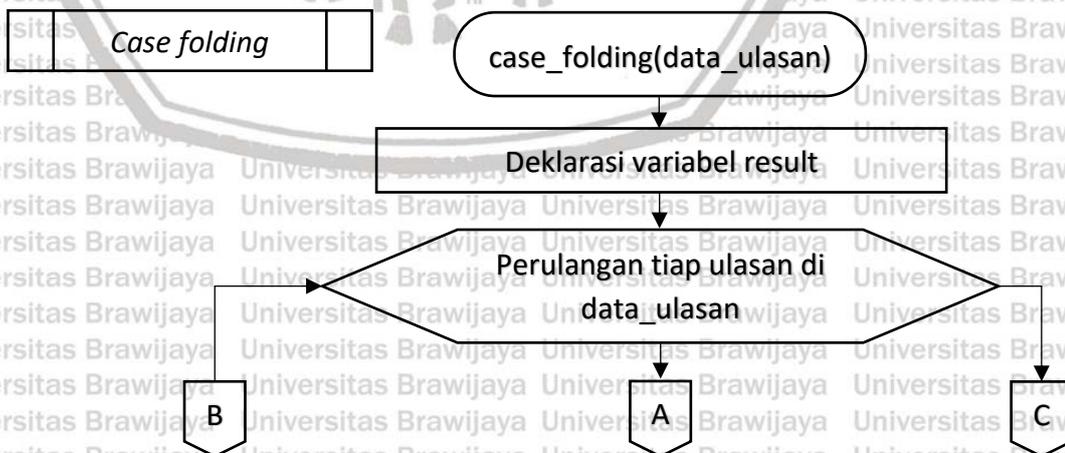
Karakter yang dimaksud berupa tanda baca, angka, dan simbol. Diagram alir dari proses *cleaning* dapat dilihat pada Gambar 4.3.



Gambar 4.3 Diagram Alir Proses *Cleaning*

4.2.2 Case Folding

Proses *case folding* bertujuan untuk melakukan perubahan pada setiap huruf kapital yang ada pada ulasan menjadi huruf kecil. Diagram alir dari proses *case folding* dapat dilihat pada Gambar 4.4.



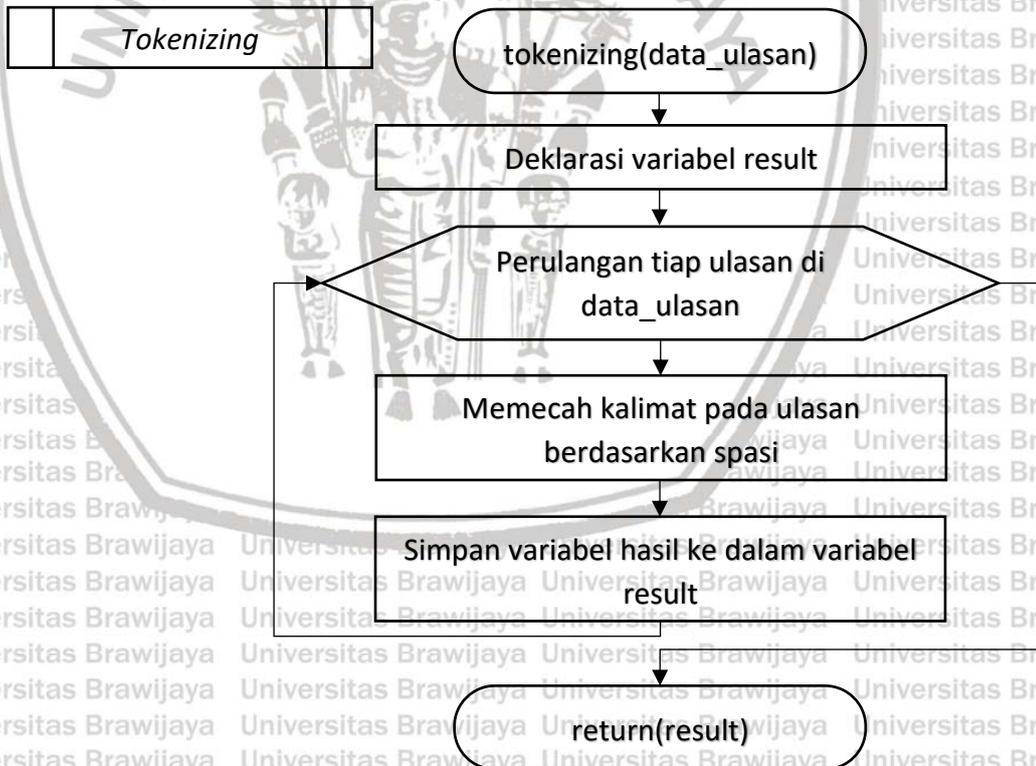
Gambar 4.4 Diagram Alir Proses *Case Folding*



Gambar 4.4 Diagram Alir Proses *Case Folding* (Lanjutan)

4.2.3 Tokenizing

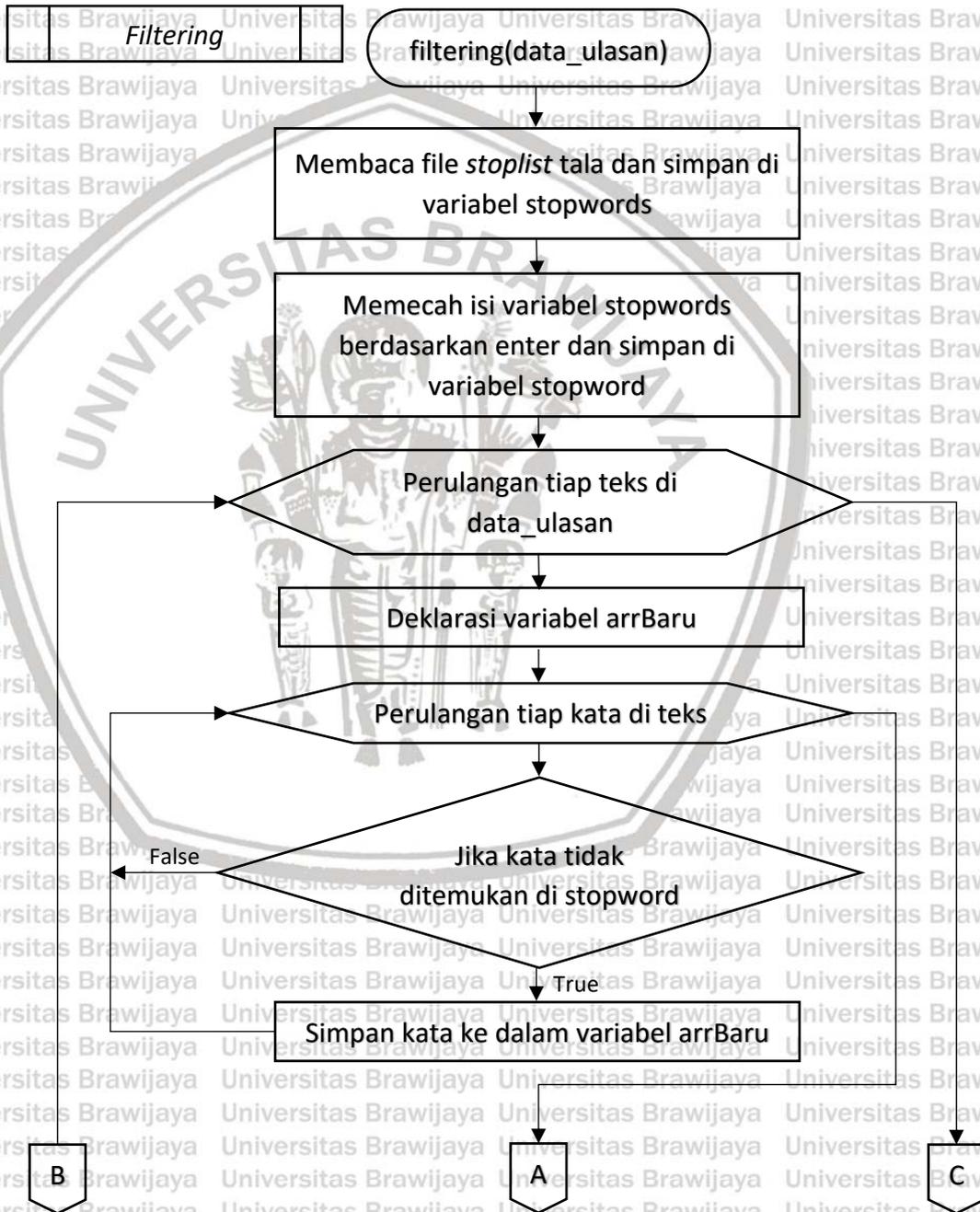
Proses *tokenizing* atau tokenisasi bertujuan untuk memisahkan tiap kata yang ada dalam setiap ulasan, dimana karakter yang digunakan sebagai faktor pemisah antar kata adalah karakter spasi. Hasil dari proses ini berupa kumpulan potongan kata untuk tiap ulasan. Diagram alir dari proses *tokenizing* dapat dilihat pada Gambar 4.5.



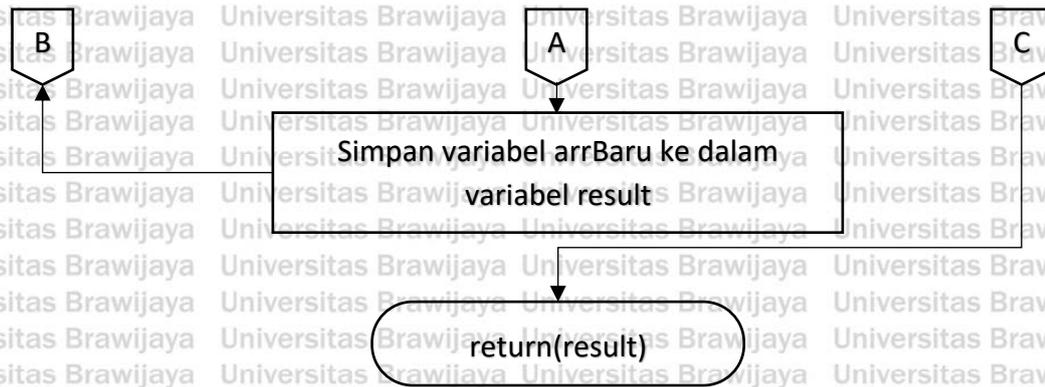
Gambar 4.5 Diagram Alir Proses *Tokenizing*

4.2.4 Filtering

Proses *filtering* bertujuan untuk menghilangkan kata hasil dari proses tokenizing yang dianggap tidak diperlukan pada proses selanjutnya karena memiliki makna yang umum dan tidak dapat merepresentasikan isi dari tiap ulasan. Kata yang dimaksud berupa kata ganti ataupun kata penghubung seperti: aku, kamu, akan, ini, ada, hanya, ke, lalu, mau, dan sebagainya. Proses *filtering* melakukan penghapusan kata berdasarkan kumpulan kata pada file *stoplist*, jika kata pada hasil proses *tokenizing* terdapat pada file *stoplist*, maka kata tersebut akan dihapus. Diagram alir dari proses *filtering* dapat dilihat pada Gambar 4.6.



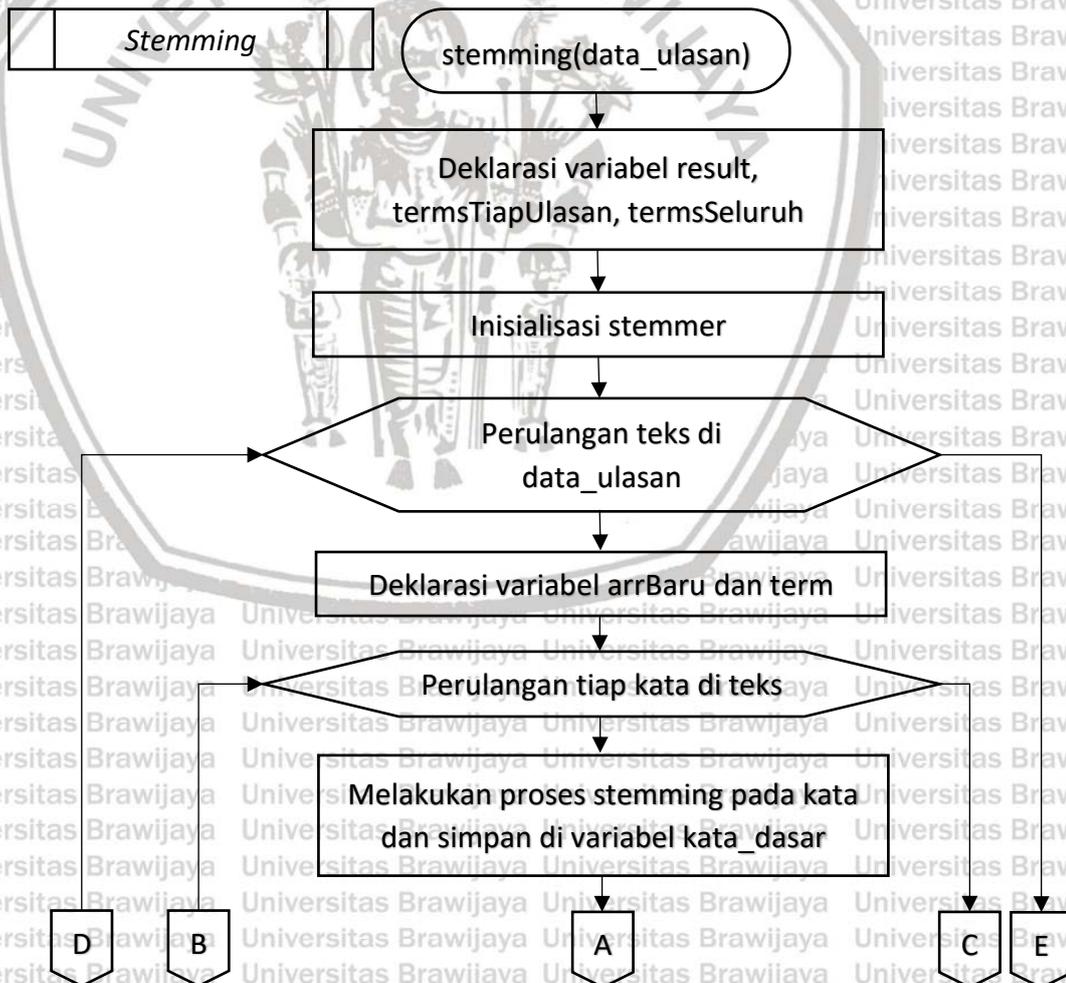
Gambar 4.6 Diagram Alir Proses *Filtering*



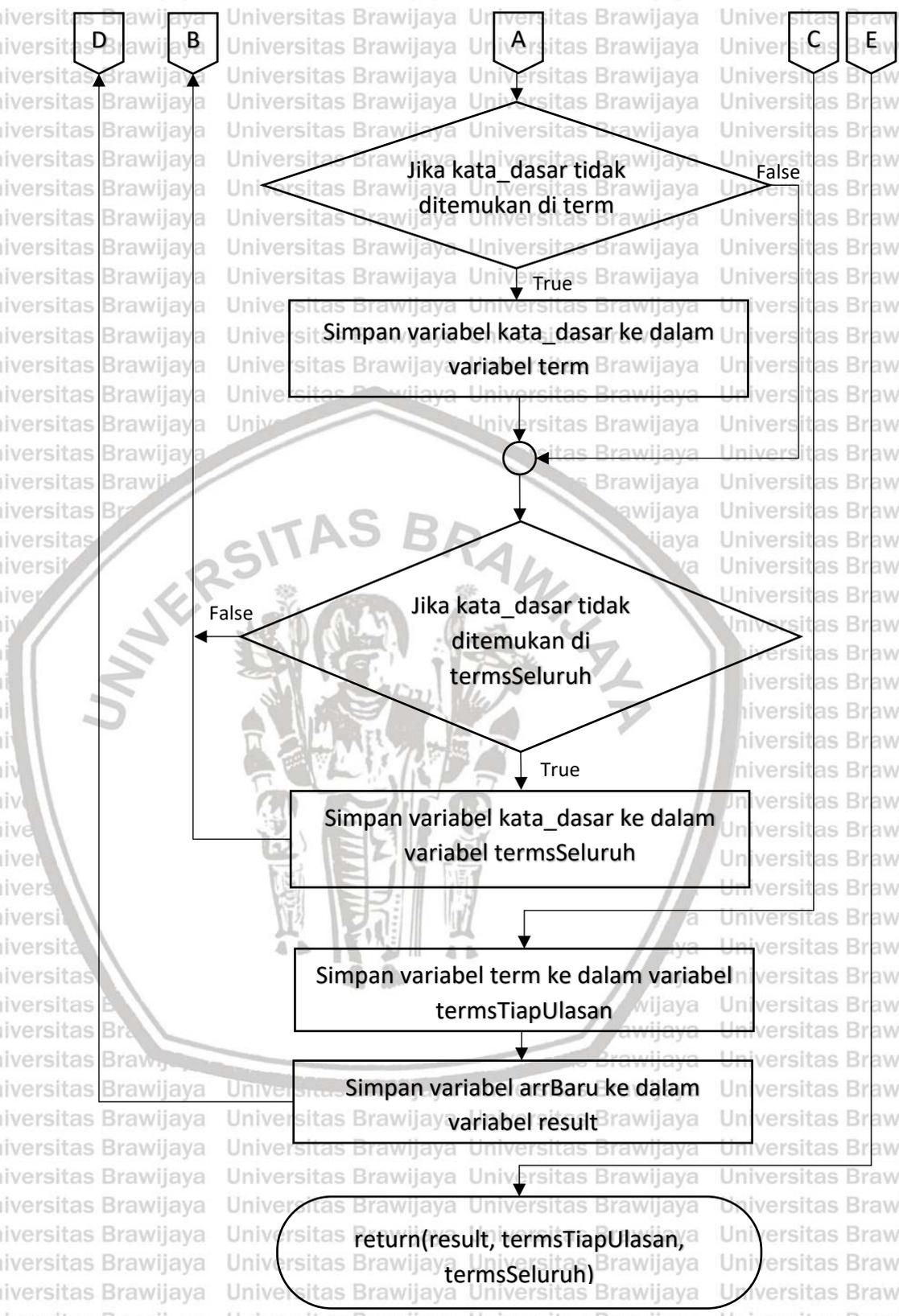
Gambar 4.6 Diagram Alir Proses *Filtering* (Lanjutan)

4.2.5 Stemming

Proses *stemming* bertujuan untuk mengembalikan kata yang memiliki imbuhan menjadi kata dasar. Proses *stemming* ini menggunakan bantuan dari *library* Sastrawi selama prosesnya. *Library* Sastrawi dapat secara otomatis melakukan pengecekan pada tiap kata dan mengembalikan ke kata dasarnya. Diagram alir dari proses *stemming* dapat dilihat pada Gambar 4.7.



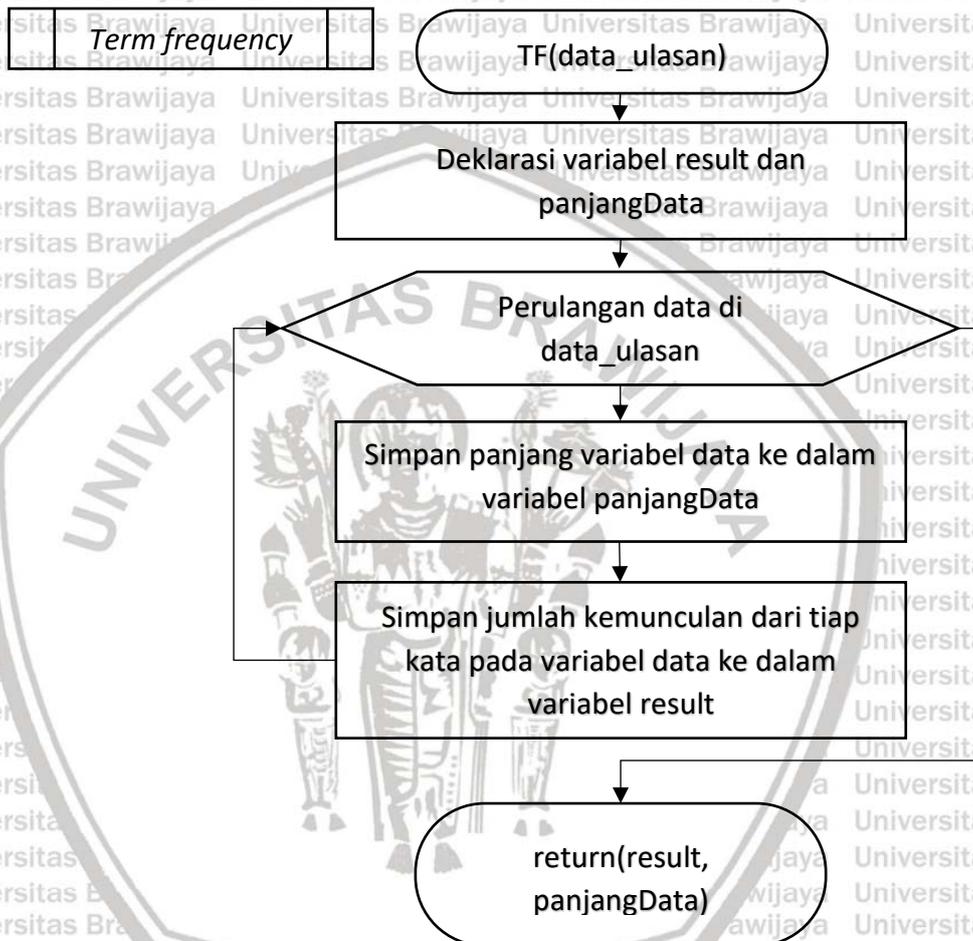
Gambar 4.7 Diagram Alir Proses *Stemming*



Gambar 4.7 Diagram Alir Proses Stemming (Lanjutan)

4.2.6 Perhitungan Term Frequency

Hasil dari proses *preprocessing* berupa *term* unik yang bisa digunakan untuk merepresentasikan isi dari tiap ulasan. Keseluruhan term yang didapat kemudian dilanjutkan dengan melakukan perhitungan nilai TF. Proses ini bertujuan untuk mengetahui seberapa banyak *term* muncul pada suatu ulasan. Proses perhitungan TF menggunakan bantuan dari *function Counter* yang terdapat pada *library collections* untuk menghitung berapa banyak *term* tersebut muncul. Diagram alir dari proses perhitungan TF dapat dilihat pada Gambar 4.8.



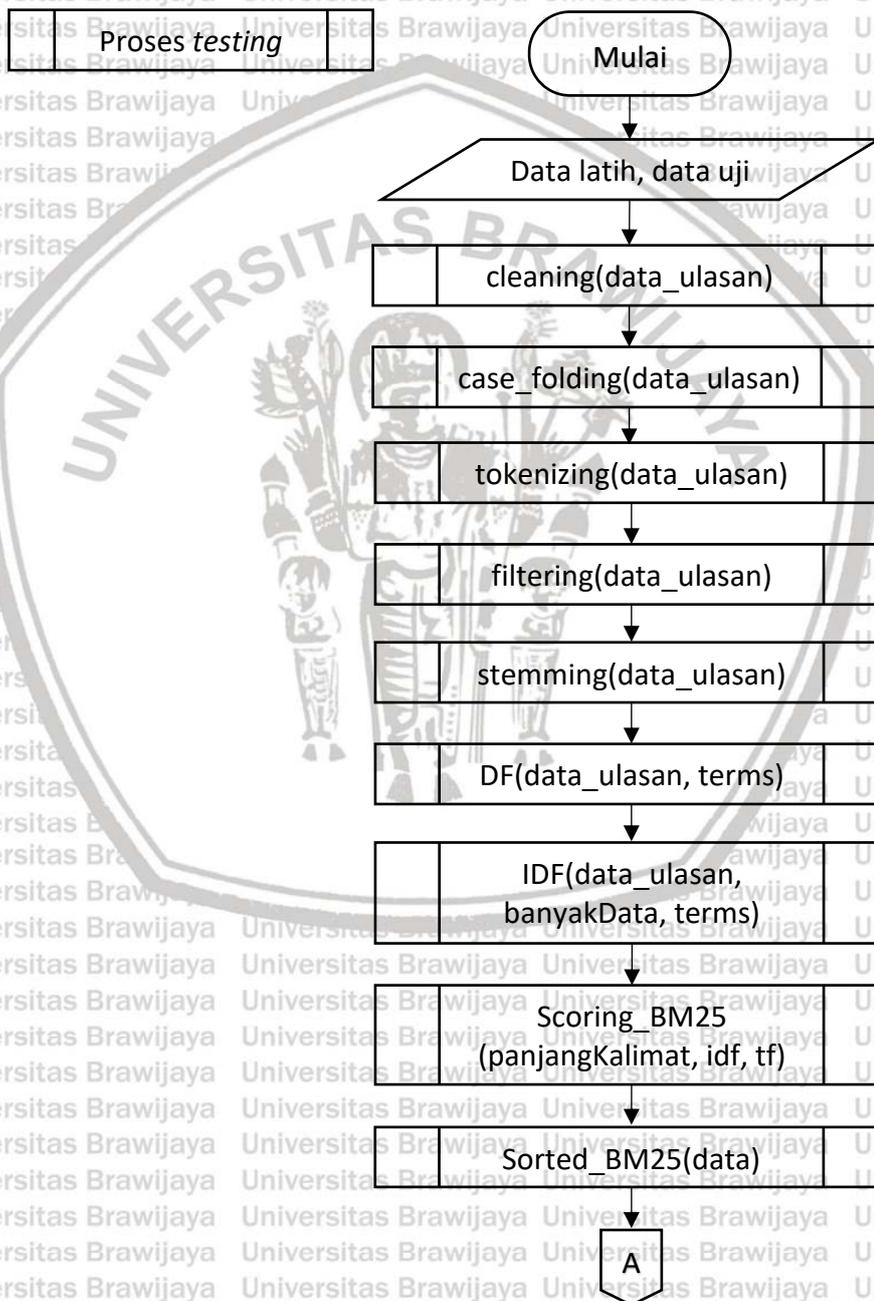
Gambar 4.8 Diagram Alir Proses Term Frequency

4.3 Perancangan Proses Testing

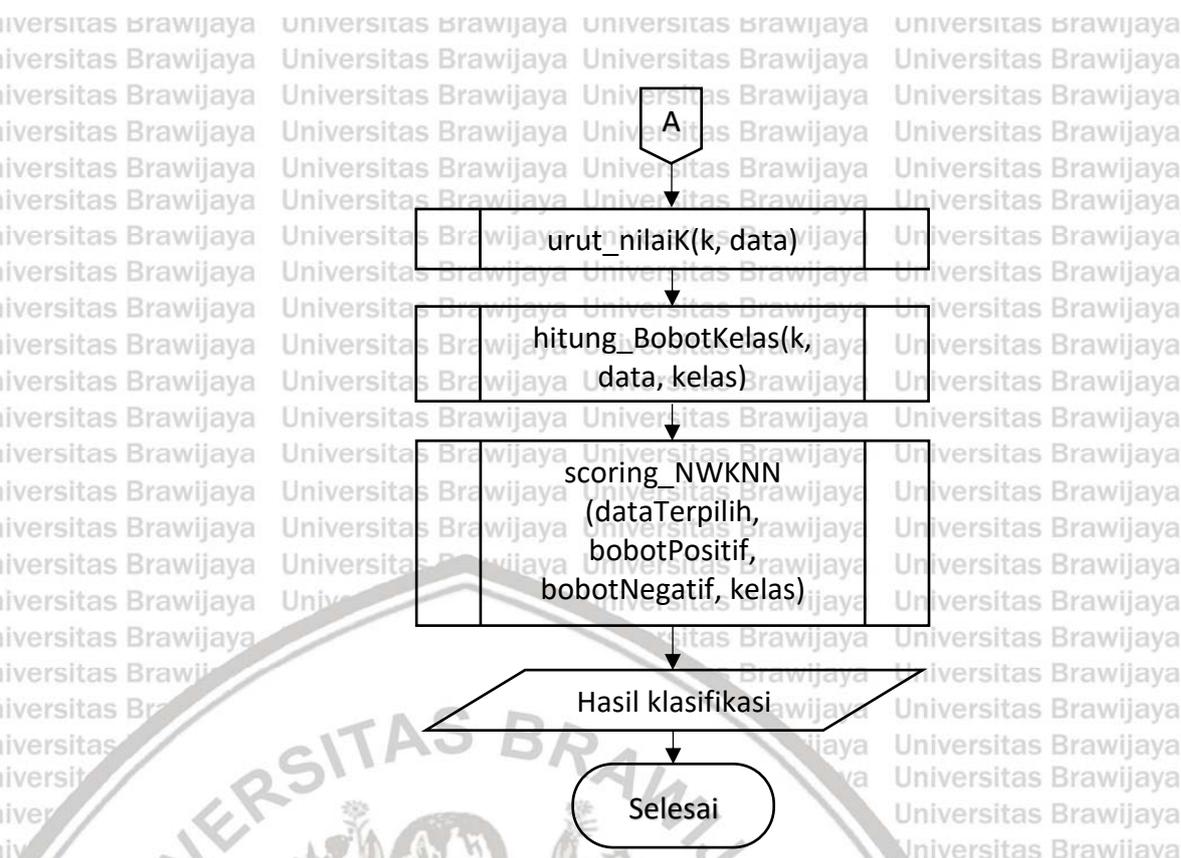
Proses *testing* akan melakukan pengolahan pada data uji yang berupa ulasan pengunjung SLG Kediri untuk mendapatkan hasil berupa klasifikasi kelas untuk tiap ulasan pada data uji. Tahapan pada proses *testing* meliputi *preprocessing* yang terdiri atas *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Kemudian dilanjutkan proses pembobotan dan pemeringkatan menggunakan metode BM25. Proses tersebut akan menghasilkan keluaran berupa ulasan beserta bobotnya yang sudah diperingkatkan dari nilai terbesar ke nilai terkecil sesuai dengan tingkat kemiripan antara ulasan data latih dengan data uji. Proses pembobotan dan

pemeringkatan menggunakan metode BM25 terbagi atas beberapa tahapan, yaitu: perhitungan *document frequency* (DF), perhitungan *inverse document frequency* (IDF), perhitungan nilai bobot BM25, dan mengurutkan data berdasarkan bobot.

Setelah melakukan proses pembobotan dan pemeringkatan maka langkah selanjutnya adalah melakukan klasifikasi pada tiap ulasan menggunakan metode NWKNN. Proses klasifikasi menggunakan metode NWKNN terdiri dari beberapa tahapan, yaitu: mengambil ulasan sejumlah nilai k, pembobotan tiap kelas, dan terakhir *scoring* NWKNN untuk mendapatkan hasil klasifikasi. Diagram alir dari proses *testing* dapat dilihat pada Gambar 4.9.



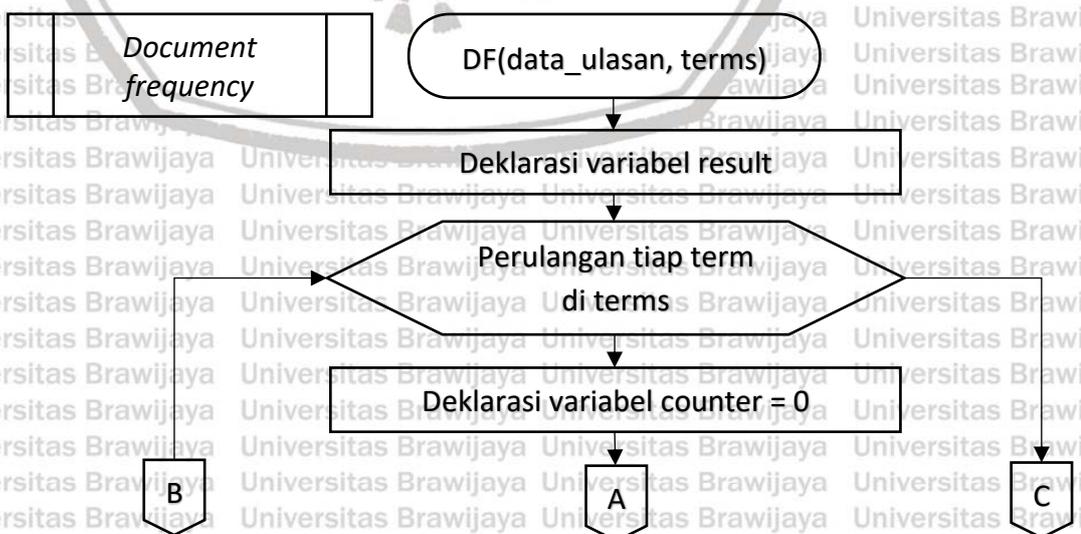
Gambar 4.9 Diagram Alir Proses *Testing*



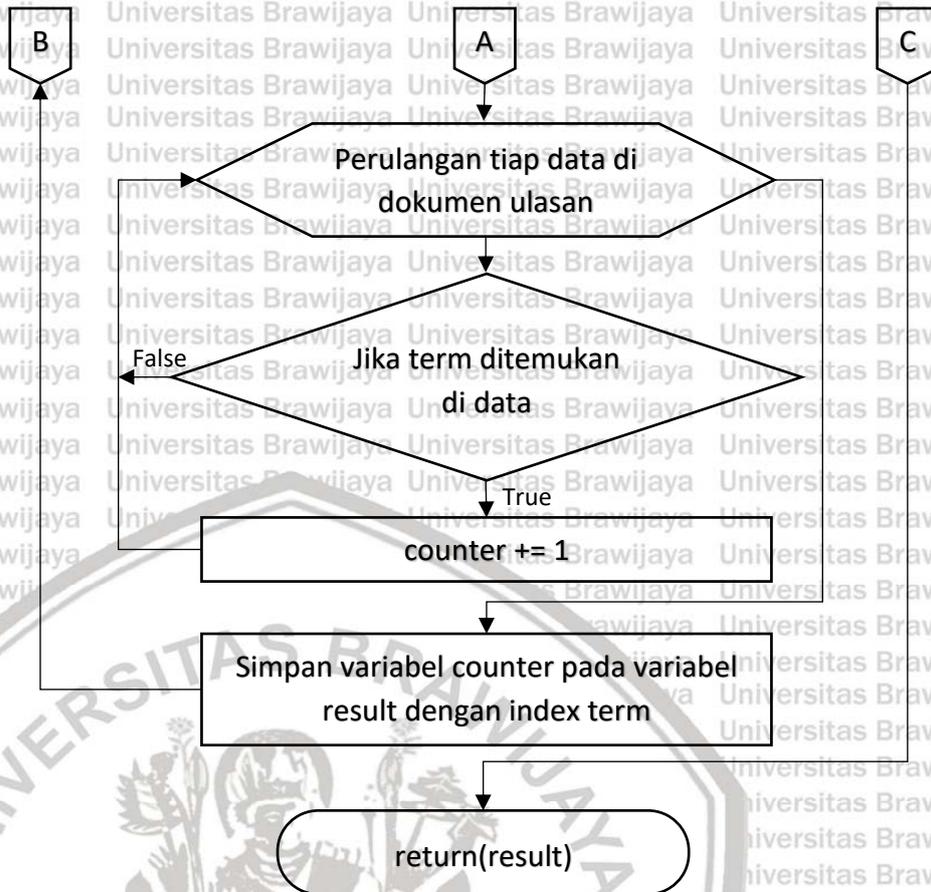
Gambar 4.9 Diagram Alir Proses *Testing* (Lanjutan)

4.3.1 Perhitungan *Document Frequency*

Langkah pertama dari metode BM25 adalah perhitungan nilai DF. Perhitungan ini bertujuan untuk mengetahui berapa banyak ulasan data latih yang mengandung *term* dari ulasan data uji. Apabila terdapat suatu *term* pada ulasan maka nilai DF dari *term* tersebut akan bertambah 1. Meskipun *term* yang ditemukan pada suatu ulasan lebih dari 1, nilai DF dari *term* tersebut tetap hanya bernilai 1. Diagram alir dari proses perhitungan DF dapat dilihat pada Gambar 4.10.



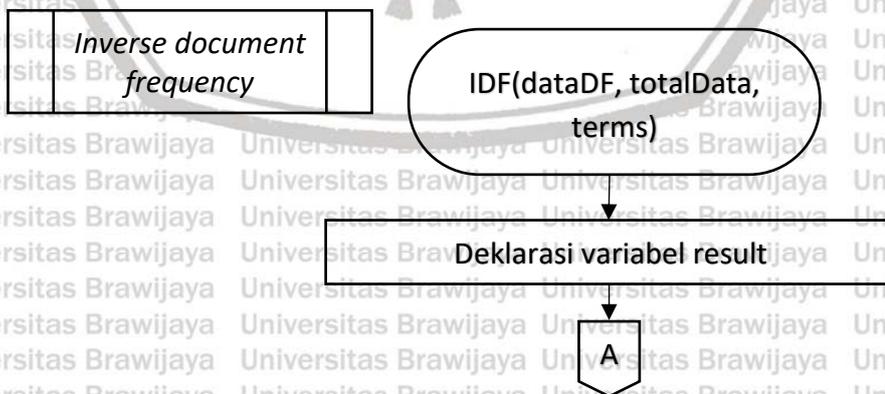
Gambar 4.10 Diagram Alir Proses *Document Frequency*



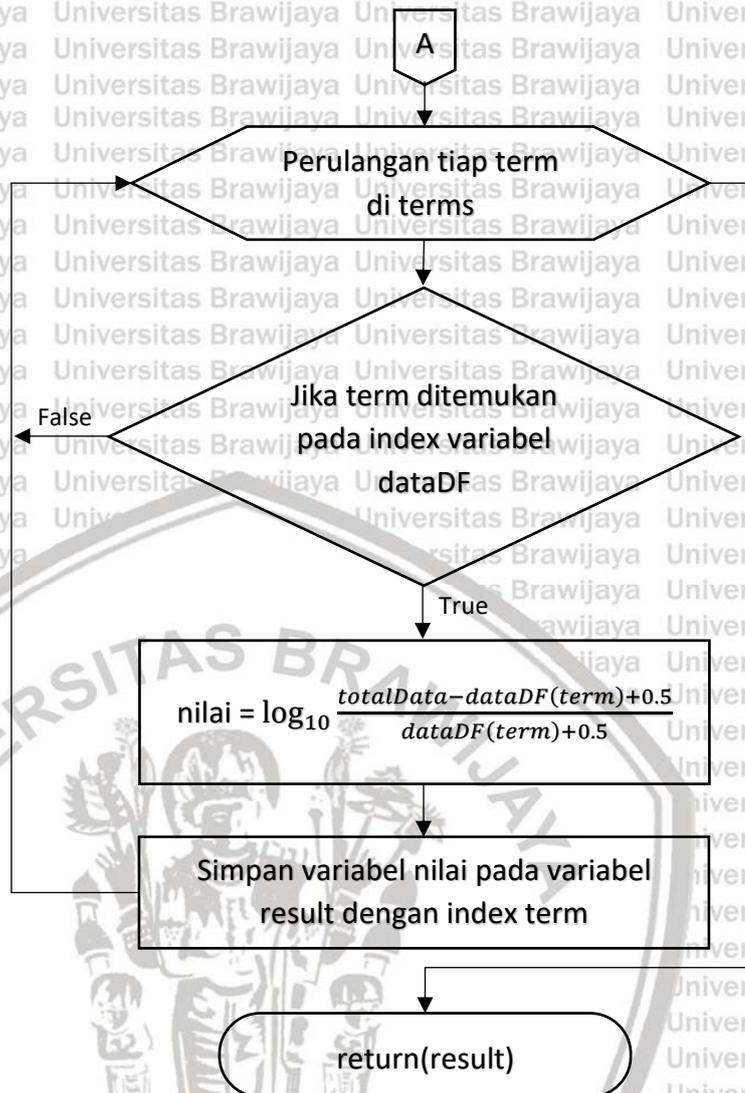
Gambar 4.10 Diagram Alir Proses *Document Frequency* (Lanjutan)

4.3.2 Perhitungan *Inverse Document Frequency*

Langkah selanjutnya adalah proses perhitungan nilai IDF. Proses ini bertujuan untuk mengetahui nilai penyebaran *term* pada ulasan yang digunakan. Semakin sedikit frekuensi *term* muncul pada suatu ulasan maka nilainya akan semakin besar. Diagram alir dari proses perhitungan IDF dapat dilihat pada Gambar 4.11.



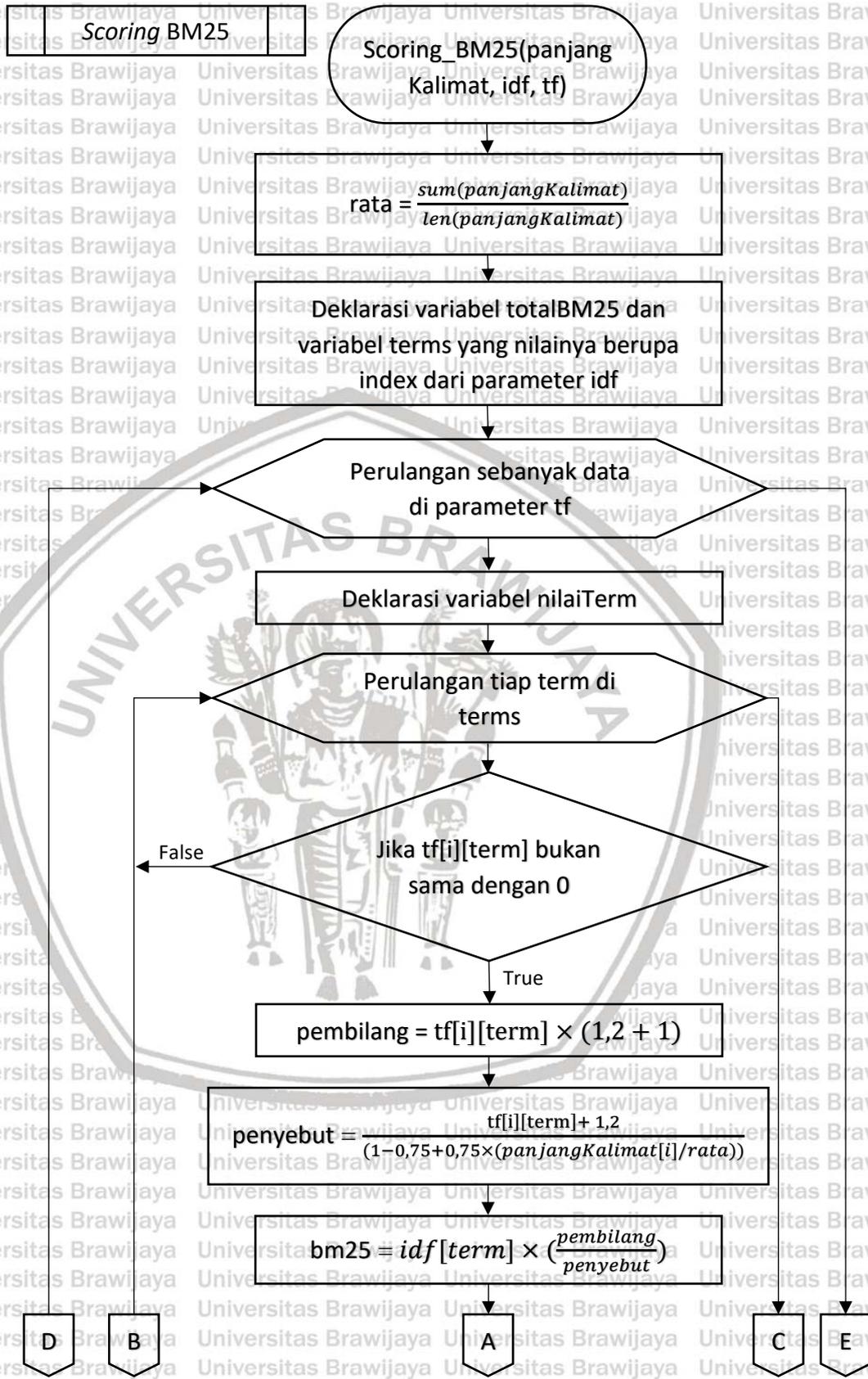
Gambar 4.11 Diagram Alir Proses *Inverse Document Frequency*



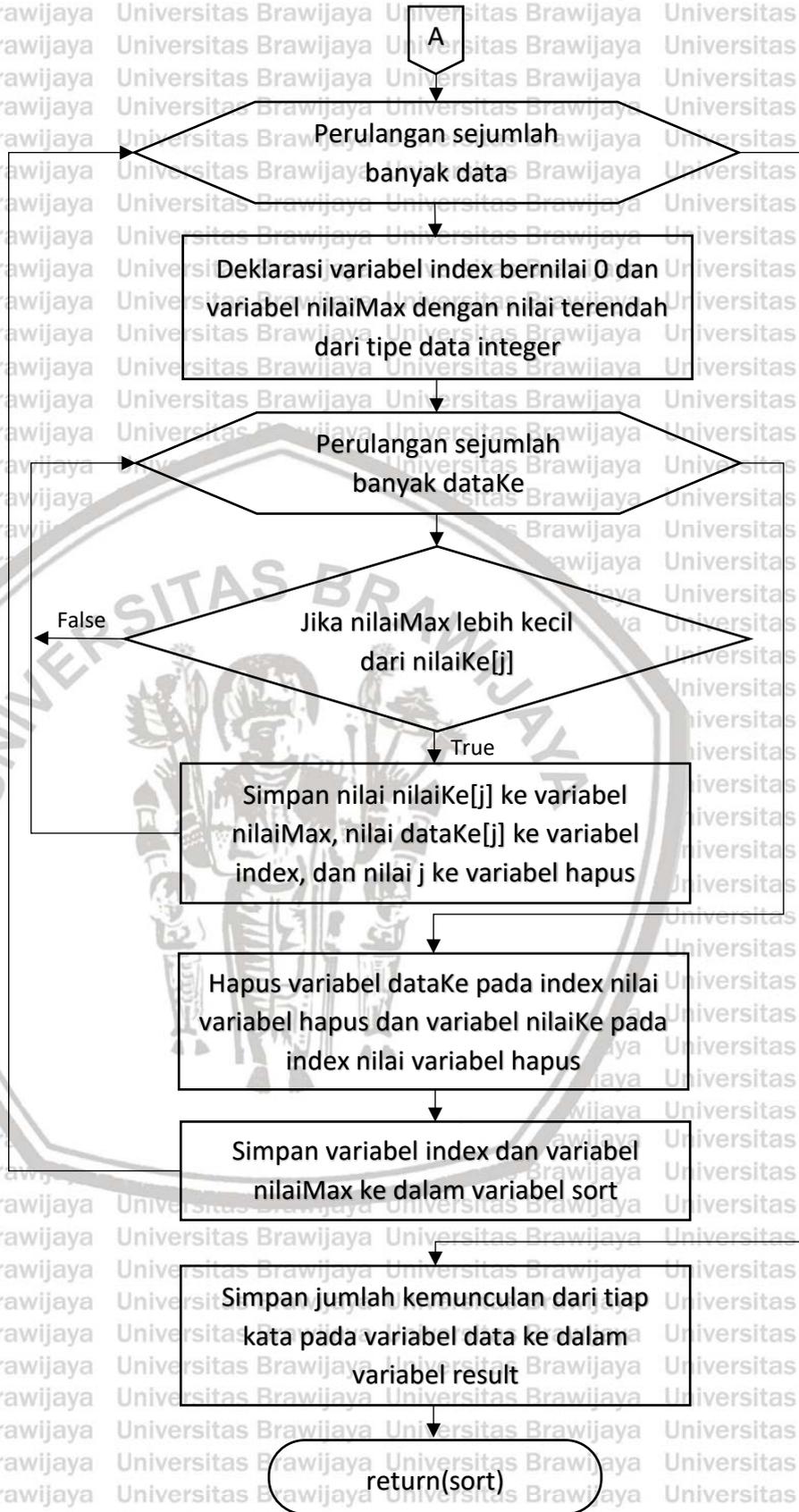
Gambar 4.11 Diagram Alir Proses *Inverse Document Frequency* (Lanjutan)

4.3.3 Scoring BM25

Setelah didapatkan nilai berupa panjang tiap ulasan, TF, dan IDF maka dilanjutkan dengan proses *scoring* BM25. Proses *scoring* BM25 bertujuan untuk mendapatkan nilai bobot kemiripan untuk tiap ulasan yang digunakan. Proses ini dilakukan berdasarkan nilai kemiripan antara data latih dan data uji. Setiap kali ulasan pada data uji diproses, maka semua ulasan pada data latih akan memiliki bobot baru sesuai dengan kemiripan dengan data uji yang sedang diproses tersebut. Diagram alir proses perhitungan *scoring* BM25 dapat dilihat pada Gambar 4.12.



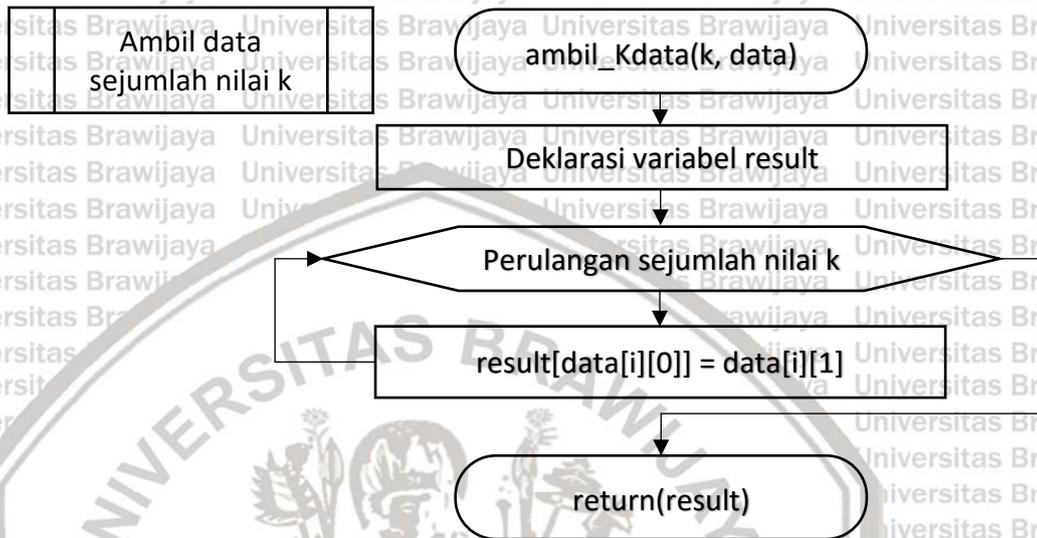
Gambar 4.12 Diagram Alir Proses Scoring BM25



Gambar 4.13 Diagram Alir Proses Pemeringkatan Ulasan (Lanjutan)

4.3.5 Ambil Data Sejumlah Nilai K

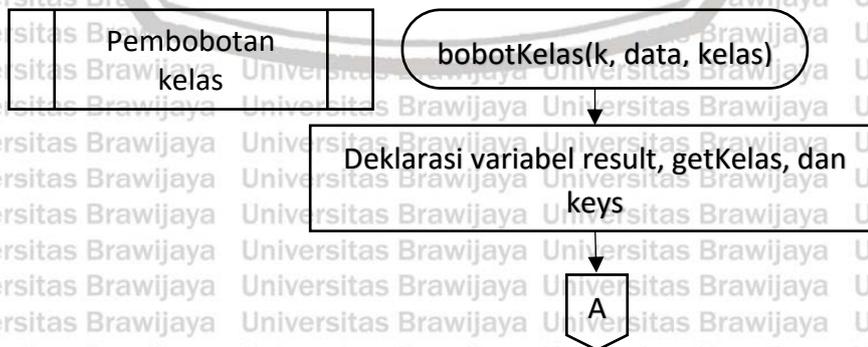
Tahapan pertama dalam proses klasifikasi menggunakan metode NWKNN adalah mengambil sejumlah data latih yang telah diurutkan nilai kemiripannya dengan data uji sebanyak nilai k. Beberapa data yang diambil tersebut akan menjadi penentu hasil klasifikasi kelas untuk data uji yang sedang diproses. Diagram alir dari proses pengambilan data sejumlah nilai k dapat dilihat pada Gambar 4.14.



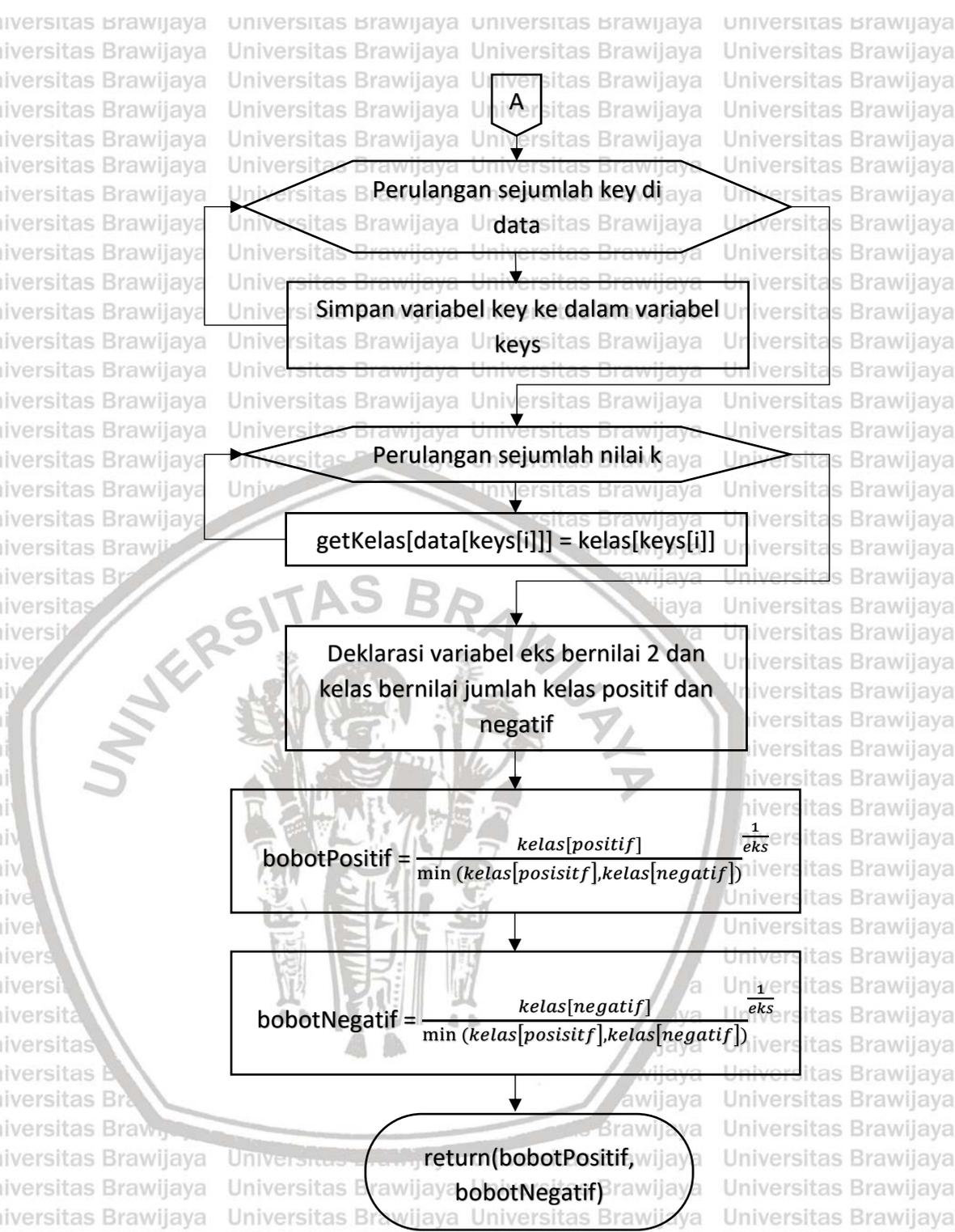
Gambar 4.14 Diagram Alir Proses Pengambilan Data Sejumlah Nilai K

4.3.6 Pembobotan Kelas

Tahapan pembobotan kelas bertujuan untuk melakukan perhitungan terhadap nilai dari kelas yang digunakan, pada penelitian ini maka yang dihitung adalah nilai dari kelas positif dan negatif. Alasan dari pemberian bobot untuk masing – masing kelas karena kelas dari data yang digunakan jumlahnya tidak seimbang, sehingga diperlukan pemberian bobot kelas supaya proses klasifikasi mendapatkan hasil yang lebih baik. Diagram alir dari proses pembobotan kelas dapat dilihat pada gambar 4.15.



Gambar 4.15 Diagram Alir Proses Pembobotan Kelas

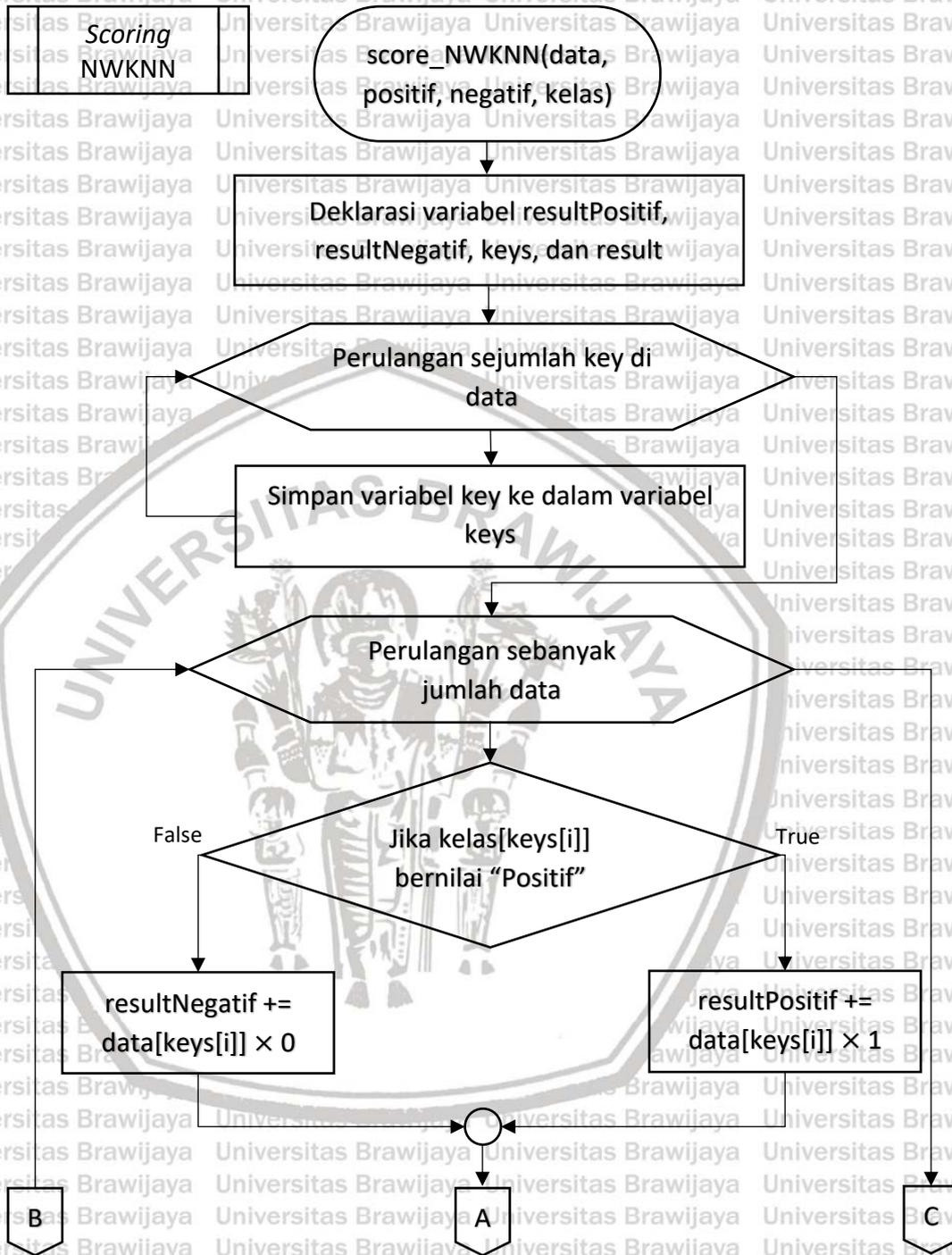


Gambar 4.15 Diagram Alir Proses Pembobotan Kelas (Lanjutan)

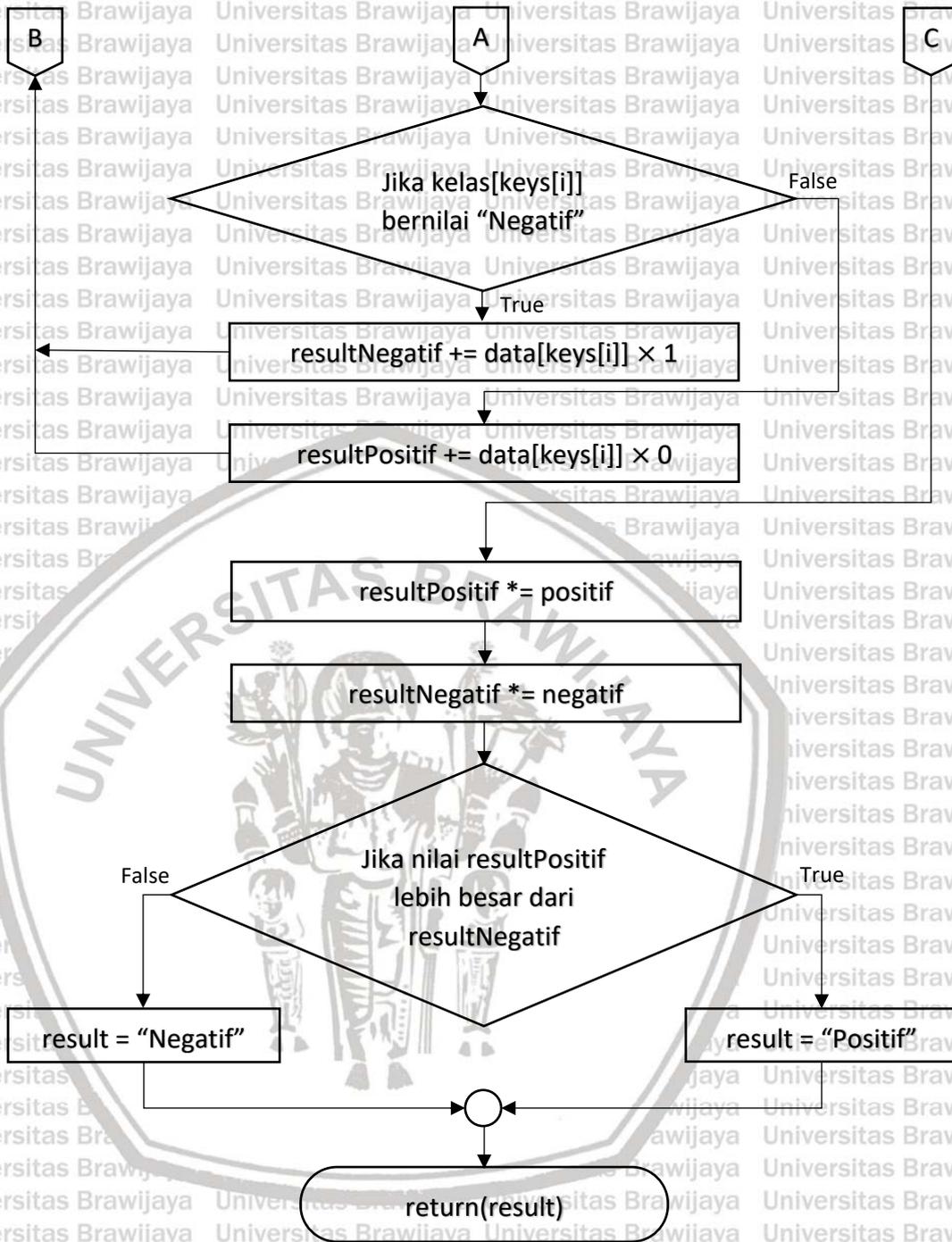
4.3.7 Scoring NWKNN

Tahapan selanjutnya dari proses klasifikasi NWKNN adalah *scoring* NWKNN. Tahapan ini merupakan langkah terakhir untuk mendapatkan hasil klasifikasi untuk tiap ulasan yang digunakan. Nilai yang dihitung merupakan nilai dari kelas yang digunakan, maka dalam penelitian ini terdapat perhitungan untuk kelas positif dan negatif. Kelas yang memiliki nilai tertinggi merupakan hasil klasifikasi

untuk data uji yang sedang diproses. Diagram alir dari proses *scoring* NWKNN dapat dilihat pada Gambar 4.16.



Gambar 4.16 Diagram Alir Proses *Scoring* NWKNN



Gambar 4.16 Diagram Alir Proses Scoring NWKNN (Lanjutan)

4.4 Perhitungan Manual

Bagian ini akan menjelaskan perhitungan manual secara keseluruhan dari penerapan metode BM25 dan NWKNN yang digunakan pada penelitian ini terhadap data latih dan data uji ulasan pengunjung SLG Kediri. Perhitungan manual meliputi *preprocessing* pada proses *training* hingga *scoring* NWKNN pada proses *testing*. Secara garis besar tahapan ini digunakan sebagai melihat tolak ukur tingkat keberhasilan dari sistem yang dibuat serta menguji kecocokan metode

terhadap data yang digunakan. Perhitungan manual ini akan menggunakan 10 data latih yang dapat dilihat pada Tabel 4.1 dan 1 data uji yang dapat dilihat pada Tabel 4.2.

Tabel 4.1 10 Data Latih

No.	Ulasan	Kelas
1.	Udara bersih tempat nya juga bersih banget, kalau mau sepi hindari hari libur.	Positif
2.	Nyaman damai seperti yang dulu saat pertama kali mengunjungi nya	Positif
3.	Salah satu ikon Kab. kediri. Tempat yang cocok murah meriah untuk rekreasi dan refreshing cari angin kencang.	Positif
4.	Letak dan bentuknya yang menjadi daya tarik ,singgah lah sejenak ketika lewat disini,merupakan salah satu bangunan yang menjadi ikon Kabupaten Kediri yang bentuknya menyerupai Arc de Triomphe yang berada di Paris, Prancis.	Positif
5.	Bingung apa yang bisa dinikmati. Panas toiletnya kotor dan bau	Negatif
6.	Lumayan lah,bisa bawa keluarga jalan" menikmati suasana segar	Positif
7.	Tempat yang nyaman dan sangat bagus untuk bersantai bersama keluarga atau teman	Positif
8.	Sangat bagus dan buat view foto foto sangat indah dan Asri	Positif
9.	Tempat ini sangat indah Cocok untuk liburan bersama keluarga	Positif
10.	Perlu ditambah tenaga kebersihan, terutama untuk memeriksa taman karena masih banyak sampah berserakan	Negatif

Tabel 4.2 1 Data Uji

No.	Ulasan
1.	Kurang tersedianya tempat sampah, sayang tempat yang indah nanti jadi kotor karena ulah orang - orang yang kurang tanggung jawab

4.4.1 Perhitungan Manual *Preprocessing*

Preprocessing merupakan tahapan pertama dalam pemrosesan teks yang terdiri atas beberapa bagian. Pada penelitian ini ulasan data latih dan data uji harus melalui *preprocessing* terlebih dahulu sebagai langkah pertama dalam melakukan analisis sentimen. Penjelasan untuk tiap tahapan akan dijelaskan pada sub bab selanjutnya.

4.4.1.1 *Cleaning*

Cleaning merupakan tahapan pertama dalam *preprocessing*. Tahapan ini berfungsi untuk menghilangkan semua karakter bukan huruf yang dianggap tidak memiliki bobot nilai untuk menentukan sebuah ulasan bernilai positif atau negatif. Hasil *cleaning* dapat dilihat pada Tabel 4.3 dan 4.4.

Tabel 4.3 Hasil *Cleaning* 10 Data Latih

No.	Ulasan
1.	Udara bersih tempat nya juga bersih banget kalau mau sepi hindari hari libur
2.	Nyaman damai seperti yang dulu saat pertama kali mengunjunginya
3.	Salah satu ikon Kab Kediri Tempat yang cocok murah meriah untuk rekreasi dan refreshing cari angin kencang
4.	Letak dan bentuknya yang menjadi daya tarik singgah lah sejenak ketika lewat disini merupakan salah satu bangunan yang menjadi ikon Kabupaten Kediri yang bentuknya menyerupai Arc de Triomphe yang berada di Paris Prancis
5.	Bingung apa yang bisa dinikmati Panas toiletnya kotor dan bau
6.	Lumayan lah bisa bawa keluarga jalan menikmati suasana segar
7.	Tempat yang nyaman dan sangat bagus untuk bersantai bersama keluarga atau teman
8.	Sangat bagus dan buat view foto foto sangat indah dan Asri
9.	Tempat ini sangat indah Cocok untuk liburan bersama keluarga
10.	Perlu ditambah tenaga kebersihan terutama untuk memeriksa taman karena masih banyak sampah berserakan

Tabel 4.4 Hasil *Cleaning* 1 Data Uji

No.	Ulasan
1.	Kurang tersedianya tempat sampah sayang tempat yang indah nanti jadi kotor karena ulah orang orang yang kurang tanggung jawab

4.4.1.2 Case Folding

Tahapan *case folding* berfungsi untuk melakukan perubahan pada setiap huruf kapital yang ditemukan di ulasan menjadi huruf kecil. Hasil *case folding* dapat dilihat pada Tabel 4.5 dan 4.6.

Tabel 4.5 Hasil Case Folding 10 Data Latih

No.	Ulasan
1.	udara bersih tempat nya juga bersih banget kalau mau sepi hindari hari libur
2.	nyaman damai seperti yang dulu saat pertama kali mengunjunginya
3.	salah satu ikon kab kediri tempat yang cocok murah meriah untuk rekreasi dan refreshing cari angin kencang
4.	letak dan bentuknya yang menjadi daya tarik singgah lah sejenak ketika lewat disini merupakan salah satu bangunan yang menjadi ikon kabupaten kediri yang bentuknya menyerupai arc de triomphe yang berada di paris prancis
5.	bingung apa yang bisa dinikmati panas toiletnya kotor dan bau
6.	lumayan lah bisa bawa keluarga jalan menikmati suasana segar
7.	tempat yang nyaman dan sangat bagus untuk bersantai bersama keluarga atau teman
8.	sangat bagus dan buat view foto foto sangat indah dan asri
9.	tempat ini sangat indah cocok untuk liburan bersama keluarga
10.	perlu ditambah tenaga kebersihan terutama untuk memeriksa taman karena masih banyak sampah berserakan

Tabel 4.6 Hasil Case Folding 1 Data Uji

No.	Ulasan
1.	kurang tersedianya tempat sampah sayang tempat yang indah nanti jadi kotor karena ulah orang orang yang kurang tanggung jawab

4.4.1.3 Tokenizing

Proses *tokenizing* berfungsi untuk memisahkan tiap kata dari suatu kalimat, dimana karakter yang digunakan sebagai faktor pemisah antar kata adalah karakter spasi. Hasil *tokenizing* dapat dilihat pada Tabel 4.7 dan 4.8.

Tabel 4.7 Hasil *Tokenizing* 10 Data Latih

No.	Hasil
1.	'udara', 'bersih', 'tempat', 'nya', 'juga', 'bersih', 'banget', 'kalau', 'mau', 'sepi', 'hindari', 'hari', 'libur'
2.	'nyaman', 'damai', 'seperti', 'yang', 'dulu', 'saat', 'pertama', 'kali', 'mengunjungi', 'nya'
3.	'salah', 'satu', 'ikon', 'kab', 'kediri', 'tempat', 'yang', 'cocok', 'murah', 'meriah', 'untuk', 'rekreasi', 'dan', 'refreshing', 'cari', 'angin', 'kencang'
4.	'letak', 'dan', 'bentuknya', 'yang', 'menjadi', 'daya', 'tarik', 'singgah', 'lah', 'sejenak', 'ketika', 'lewat', 'disini', 'merupakan', 'salah', 'satu', 'bangunan', 'yang', 'menjadi', 'ikon', 'kabupaten', 'kediri', 'yang', 'bentuknya', 'menyerupai', 'arc', 'de', 'triomphe', 'yang', 'berada', 'di', 'paris', 'prancis'
5.	'bingung', 'apa', 'yang', 'bisa', 'dinikmati', 'panas', 'toiletnya', 'kotor', 'dan', 'bau'
6.	'lumayan', 'lah', 'bisa', 'bawa', 'keluarga', 'jalan', 'menikmati', 'suasana', 'segar'
7.	'tempat', 'yang', 'nyaman', 'dan', 'sangat', 'bagus', 'untuk', 'bersantai', 'bersama', 'keluarga', 'atau', 'teman'
8.	'sangat', 'bagus', 'dan', 'buat', 'view', 'foto', 'foto', 'sangat', 'indah', 'dan', 'asri'
9.	'tempat', 'ini', 'sangat', 'indah', 'cocok', 'untuk', 'liburan', 'bersama', 'keluarga'
10.	'perlu', 'ditambah', 'tenaga', 'kebersihan', 'terutama', 'untuk', 'memeriksa', 'taman', 'karena', 'masih', 'banyak', 'sampah', 'berserakan'

Tabel 4.8 Hasil *Tokenizing* 1 Data Uji

No.	Hasil
1.	'kurang', 'tersedianya', 'tempat', 'sampah', 'sayang', 'tempat', 'yang', 'indah', 'nantinya', 'jadi', 'kotor', 'karena', 'ulah', 'orang', 'orang', 'yang', 'kurang', 'tanggung', 'jawab'

4.4.1.4 Filtering

Tahapan *filtering* berfungsi untuk menghilangkan kata yang dirasa memiliki makna umum dan tidak dapat merepresentasikan isi dari ulasan. Tahapan ini menghilangkan kata berdasarkan kumpulan kata pada file *stoplist*. Jika kata pada hasil proses *tokenizing* terdapat pada file *stoplist*, maka kata tersebut akan dihilangkan. Hasil *tokenizing* dapat dilihat pada Tabel 4.9 dan 4.10.

Tabel 4.9 Hasil Filtering 10 Data Latih

No.	Hasil
1.	'udara', 'bersih', 'nya', 'bersih', 'banget', 'sepi', 'hindari', 'libur'
2.	'nyaman', 'damai', 'kali', 'mengunjungi', 'nya'
3.	'salah', 'ikon', 'kab', 'kediri', 'cocok', 'murah', 'meriah', 'rekreasi', 'refreshing', 'cari', 'angin', 'kencang'
4.	'letak', 'bentuknya', 'daya', 'tarik', 'singgah', 'salah', 'bangunan', 'ikon', 'kabupaten', 'kediri', 'bentuknya', 'menyerupai', 'arc', 'de', 'triomphe', 'paris', 'prancis'
5.	'bingung', 'dinikmati', 'panas', 'toiletnya', 'kotor', 'bau'
6.	'lumayan', 'bawa', 'keluarga', 'jalan', 'menikmati', 'suasana', 'segar'
7.	'nyaman', 'bagus', 'bersantai', 'keluarga', 'teman'
8.	'bagus', 'view', 'foto', 'foto', 'indah', 'asri'
9.	'indah', 'cocok', 'liburan', 'keluarga'
10.	'ditambah', 'tenaga', 'kebersihan', 'memeriksa', 'taman', 'sampah', 'berserakan'

Tabel 4.10 Hasil Filtering 1 Data Uji

No.	Hasil
1.	'tersedianya', 'sampah', 'sayang', 'indah', 'kotor', 'ulah', 'orang', 'orang', 'tanggung'

4.4.1.5 Stemming

Tahapan *stemming* berfungsi untuk mengembalikan kata yang memiliki imbuhan menjadi kata dasar. Hasil *stemming* dapat dilihat pada Tabel 4.11 dan 4.12.

Tabel 4.11 Hasil Stemming 10 Data Latih

No.	Hasil
1.	'udara', 'bersih', 'nya', 'bersih', 'banget', 'sepi', 'hinder', 'libur'
2.	'nyaman', 'damai', 'kali', 'ujung', 'nya'
3.	'salah', 'ikon', 'kab', 'diri', 'cocok', 'murah', 'riah', 'rekreasi', 'refreshing', 'cari', 'angin', 'kencang'
4.	'letak', 'bentuk', 'daya', 'tarik', 'singgah', 'salah', 'bangun', 'ikon', 'kabupaten', 'diri', 'bentuk', 'rupa', 'arc', 'de', 'triomphe', 'paris', 'prancis'

Tabel 4.11 Hasil Stemming 10 Data Latih (Lanjutan)

No.	Hasil
5.	'bingung', 'nikmat', 'panas', 'toilet', 'kotor', 'bau'
6.	'lumayan', 'bawa', 'keluarga', 'jalan', 'nikmat', 'suasana', 'segar'
7.	'nyaman', 'bagus', 'santa', 'keluarga', 'teman'
8.	'bagus', 'view', 'foto', 'foto', 'indah', 'asri'
9.	'indah', 'cocok', 'libur', 'keluarga'
10.	'tambah', 'tenaga', 'bersih', 'periksa', 'taman', 'sampah', 'sera'

Tabel 4.12 Hasil Stemming 1 Data Uji

No.	Hasil
1.	'sedia', 'sampah', 'sayang', 'indah', 'kotor', 'ulah', 'orang', 'orang', 'tanggung'

4.4.2 Perhitungan Manual Metode BM25

Setelah proses *preprocessing* selesai dilakukan, maka akan didapatkan *term* yang mewakili makna pada tiap ulasan. Tahapan selanjutnya setelah mendapatkan *term* adalah melakukan perhitungan untuk mendapatkan bobot kemiripan serta pemeringkatan menggunakan metode BM25. Metode BM25 mendapatkan nilai melalui beberapa tahapan, yaitu: menghitung *term frequency* (TF) saat proses *training*, menghitung *document frequency* (DF), *inverse document frequency* (IDF), panjang tiap ulasan, rata – rata panjang ulasan, *scoring* BM25, dan pemeringkatan ulasan saat proses *testing*.

4.4.2.1 Perhitungan Manual Term Frequency

Tahapan pertama adalah menghitung berapa banyak frekuensi *term* pada ulasan data latih. Proses perhitungan manual TF dijabarkan sebagai berikut.

Term udara pada D1 hanya muncul sekali, maka bernilai $U_r = 1$

Term udara pada D2 tidak muncul sama sekali, maka bernilai $= 0$

Term bersih pada D1 muncul dua kali, maka bernilai $U_r = 2$

Term bersih pada D2 tidak muncul sama sekali, maka bernilai $= 0$

Hasil lengkap perhitungan manual TF ulasan data latih dapat dilihat pada Tabel 4.13.

Tabel 4.13 Hasil Perhitungan TF Data Latih

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
udara	1	0	0	0	0	0	0	0	0	0
bersih	2	0	0	0	0	0	0	0	0	1

Tabel 4.13 Hasil Perhitungan TF Data Latih (Lanjutan)

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
nya	1	1	0	0	0	0	0	0	0	0
banget	1	0	0	0	0	0	0	0	0	0
sepi	1	0	0	0	0	0	0	0	0	0
hindar	1	0	0	0	0	0	0	0	0	0
libur	1	0	0	0	0	0	0	0	0	0
nyaman	0	1	0	0	0	0	1	0	0	0
damai	0	1	0	0	0	0	0	0	0	0
kali	0	1	0	0	0	0	0	0	0	0
ujung	0	1	0	0	0	0	0	0	0	0
salah	0	0	1	0	0	0	0	0	0	0
ikon	0	0	1	1	0	0	0	0	0	0
kab	0	0	1	0	0	0	0	0	0	0
diri	0	0	1	1	0	0	0	0	0	0
cocok	0	0	1	0	0	0	0	0	1	0
murah	0	0	1	0	0	0	0	0	0	0
riah	0	0	1	0	0	0	0	0	0	0
rekreasi	0	0	1	0	0	0	0	0	0	0
refreshing	0	0	1	0	0	0	0	0	0	0
cari	0	0	1	0	0	0	0	0	0	0
angin	0	0	1	0	0	0	0	0	0	0
kencang	0	0	1	0	0	0	0	0	0	0
letak	0	0	0	1	0	0	0	0	0	0
bentuk	0	0	0	2	0	0	0	0	0	0
daya	0	0	0	1	0	0	0	0	0	0
tarik	0	0	0	1	0	0	0	0	0	0
salah	0	0	0	1	0	0	0	0	0	0
bangun	0	0	0	1	0	0	0	0	0	0
kabupaten	0	0	0	1	0	0	0	0	0	0
rupa	0	0	0	1	0	0	0	0	0	0



Tabel 4.13 Hasil Perhitungan TF Data Latih (Lanjutan)

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
arc	0	0	0	1	0	0	0	0	0	0
de	0	0	0	1	0	0	0	0	0	0
trionphe	0	0	0	1	0	0	0	0	0	0
paris	0	0	0	1	0	0	0	0	0	0
prancis	0	0	0	1	0	0	0	0	0	0
bingung	0	0	0	0	1	0	0	0	0	0
nikmat	0	0	0	0	1	1	0	0	0	0
panas	0	0	0	0	1	0	0	0	0	0
toilet	0	0	0	0	1	0	0	0	0	0
kotor	0	0	0	0	1	0	0	0	0	0
bau	0	0	0	0	1	0	0	0	0	0
lumayan	0	0	0	0	0	1	0	0	0	0
bawa	0	0	0	0	0	1	0	0	0	0
keluarga	0	0	0	0	0	1	1	0	1	0
jalan	0	0	0	0	0	1	0	0	0	0
suasana	0	0	0	0	0	1	0	0	0	0
segar	0	0	0	0	0	1	0	0	0	0
bagus	0	0	0	0	0	0	1	1	0	0
santa	0	0	0	0	0	0	1	0	0	0
teman	0	0	0	0	0	0	1	0	0	0
view	0	0	0	0	0	0	0	1	0	0
foto	0	0	0	0	0	0	0	2	0	0
indah	0	0	0	0	0	0	0	1	1	0
asri	0	0	0	0	0	0	0	1	0	0
libur	0	0	0	0	0	0	0	0	1	0
tambah	0	0	0	0	0	0	0	0	0	1
tenaga	0	0	0	0	0	0	0	0	0	1
periksa	0	0	0	0	0	0	0	0	0	1
taman	0	0	0	0	0	0	0	0	0	1



Tabel 4.13 Hasil Perhitungan TF Data Latih (Lanjutan)

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
sampah	0	0	0	0	0	0	0	0	0	1
sera	0	0	0	0	0	0	0	0	0	1

Pada perhitungan manual TF di Tabel 4.13 dapat dilihat bahwa ketika suatu *term* ditemukan pada suatu ulasan maka nilainya akan bertambah 1 setiap kemunculannya, sedangkan jika *term* tidak ditemukan pada suatu ulasan maka akan bernilai 0. Untuk mempermudah perhitungan manual, tiap *term* hasil dari data uji juga dicari kemunculannya pada seluruh ulasan data latih. Proses perhitungan manual TF data uji dijabarkan sebagai berikut.

Term sedia pada D10 tidak muncul sama sekali, maka bernilai = 0

Term sampah pada D10 hanya muncul sekali, maka bernilai = 1

Hasil lengkap perhitungan manual TF ulasan data uji dapat dilihat pada Tabel 4.14.

Tabel 4.14 Hasil Perhitungan TF Data Uji

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
sedia	0	0	0	0	0	0	0	0	0	0
sampah	0	0	0	0	0	0	0	0	0	1
sayang	0	0	0	0	0	0	0	0	0	0
indah	0	0	0	0	0	0	0	1	1	0
kotor	0	0	0	0	1	0	0	0	0	0
ulah	0	0	0	0	0	0	0	0	0	0
orang	0	0	0	0	0	0	0	0	0	0
tanggung	0	0	0	0	0	0	0	0	0	0

4.4.2.2 Perhitungan Manual Document Frequency

Setelah mendapatkan nilai TF untuk tiap *term* yang ada pada data uji, maka langkah selanjutnya adalah mencari nilai DF. DF digunakan untuk mengetahui jumlah ulasan data latih yang memiliki *term* tertentu dari ulasan data uji. Apabila *term* tidak ditemukan dalam data latih maka nilai DF tidak bertambah, sedangkan apabila *term* ditemukan maka nilai DF akan bertambah 1 untuk tiap ulasan data latih yang memiliki *term* tersebut. Proses perhitungan manual DF dijabarkan sebagai berikut.

Term sedia tidak muncul pada keseluruhan data latih, maka bernilai = 0

Term sampah hanya muncul sekali pada D10, maka bernilai = 1

Term indah muncul dua kali pada D8 dan D9, maka bernilai = 2

Hasil lengkap perhitungan manual DF dapat dilihat pada Tabel 4.15.

Tabel 4.15 Hasil Perhitungan DF

Term	DF
sedia	0
sampah	1
sayang	0
indah	2
kotor	1
ulah	0
orang	0
tanggung	0

4.4.2.3 Perhitungan Manual Inverse Document Frequency

Setelah mendapatkan nilai DF, maka selanjutnya adalah menghitung nilai IDF. Rumus yang digunakan untuk melakukan perhitungan IDF menggunakan Persamaan 2.2. Dimana dalam persamaan tersebut jumlah data latih dilambangkan dengan N dan nilai DF dari term dilambangkan dengan $df(q_i)$, yang nilainya dapat dilihat pada Tabel 4.15. Perhitungan manual IDF dijabarkan sebagai berikut.

$$IDF (sedia) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 1,3222$$

$$IDF (sampah) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-1+0.5}{1+0.5} = 0,8016$$

$$IDF (sayang) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 1,3222$$

$$IDF (indah) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 0,5314$$

$$IDF (kotor) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 0,8016$$

$$IDF (ulah) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 1,3222$$

$$IDF (orang) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 1,3222$$

$$IDF (tanggung) = \log_{10} \frac{N-df(q_i)+0.5}{df(q_i)+0.5} = \log_{10} \frac{10-0+0.5}{0+0.5} = 1,3222$$

Hasil lengkap perhitungan manual nilai IDF dapat dilihat pada Tabel 4.16.

Tabel 4.16 Hasil Perhitungan IDF

Term	IDF
sedia	1,3222
sampah	0,8016
sayang	1,3222
indah	0,5314
kotor	0,8016
ulah	1,3222
orang	1,3222
tanggung	1,3222

4.4.2.4 Perhitungan Manual Scoring BM25

Setelah mendapatkan *term* dari proses TF dan nilai IDF, maka dilanjutkan dengan proses *scoring* BM25 untuk mendapatkan nilai bobot kemiripan antara tiap data latih dengan data uji. Perhitungan manual *scoring* BM25 untuk mendapatkan nilai bobot dijabarkan sebagai berikut.

- 1) Hitung panjang ulasan dan panjang rata – rata ulasan data latih. Sesuai dengan Persamaan 2.1, pada rumus *scoring* BM25 diperlukan juga panjang tiap ulasan data latih dan panjang rata – rata untuk mendapatkan nilai bobot tersebut. Dimana panjang ulasan diambil berdasarkan banyaknya *term* pada tiap ulasan setelah proses preprocessing dan selanjutnya dihitung rata – rata dari keseluruhan panjang ulasan. Panjang tiap ulasan serta rata – rata panjang ulasan dapat dilihat pada Tabel 4.17.

Tabel 4.17 Hasil Perhitungan Panjang dan Rata – Rata Ulasan

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Panjang Ulasan	8	5	12	17	6	7	5	6	4	7
Rata – Rata Ulasan	7,7									

- 2) Hitung nilai bobot untuk tiap ulasan data latih. Sesuai dengan Persamaan 2.1, maka untuk proses perhitungan manual ini nilai TF menggunakan data yang ada pada Tabel 4.14, nilai IDF dari Tabel 4.16, $k_1 = 1,2$, $b = 0,75$, $|D|$ serta avgdl menggunakan nilai yang telah dihitung pada Tabel 4.17. Mengacu pada Persamaan 2.1, contoh perhitungan manual nilai bobot menggunakan metode BM25 dijabarkan sebagai berikut.

Term = sampah

Ulasan = ke-10

$$\begin{aligned}
 \text{BM25}_{(\text{sampah}, D_{10})} &= \text{idf}(q_i) \times \frac{\text{tf}(q_i, D)(k_1+1)}{\text{tf}(q_i, D) + k_1 \left(1 - b + b \frac{|D|}{\text{avgdl}}\right)} \\
 &= 0,8016 \times \frac{1 \times (1,2+1)}{1+1,2 \left(1 - 0,75 + 0,75 \frac{7}{7,7}\right)} \\
 &= 0,8016 \times \frac{2,2}{2,1181} \\
 &= 0,8016 \times 1,0386 \\
 &= 0,8325
 \end{aligned}$$

Berdasarkan contoh perhitungan bobot tersebut, maka keseluruhan perhitungan bobot untuk tiap ulasan pada data latih dapat dilihat pada Tabel 4.18 dan Tabel 4.19.

Tabel 4.18 Hasil Scoring BM25 (Ulasan ke-1 – ke-5)

Term	D1	D2	D3	D4	D5
sedia	0,0000	0,0000	0,0000	0,0000	0,0000
sampah	0,0000	0,0000	0,0000	0,0000	0,0000
sayang	0,0000	0,0000	0,0000	0,0000	0,0000
indah	0,0000	0,0000	0,0000	0,0000	0,0000
kotor	0,0000	0,0000	0,0000	0,0000	0,8812
ulah	0,0000	0,0000	0,0000	0,0000	0,0000
orang	0,0000	0,0000	0,0000	0,0000	0,0000
tanggung	0,0000	0,0000	0,0000	0,0000	0,0000
Total BM25	0,0000	0,0000	0,0000	0,0000	0,8812

Tabel 4.19 Hasil Scoring BM25 (Ulasan ke-6 – ke-10)

Term	D6	D7	D8	D9	D10
sedia	0,0000	0,0000	0,0000	0,0000	0,0000
sampah	0,0000	0,0000	0,0000	0,0000	0,8325
sayang	0,0000	0,0000	0,0000	0,0000	0,0000
indah	0,0000	0,0000	0,5842	0,6615	0,0000
kotor	0,0000	0,0000	0,0000	0,0000	0,0000
ulah	0,0000	0,0000	0,0000	0,0000	0,0000
orang	0,0000	0,0000	0,0000	0,0000	0,0000
tanggung	0,0000	0,0000	0,0000	0,0000	0,0000
Total BM25	0,0000	0,0000	0,5842	0,6615	0,8325

4.4.2.5 Pemeringkatan Ulasan

Setelah mendapatkan nilai bobot BM25, maka dilanjutkan dengan mengurutkan ulasan sesuai dengan bobot tertinggi ke bobot terendah. Hasil pemeringkatan ulasan pada proses manualisasi ini dapat dilihat pada Tabel 4.20.

Tabel 4.20 Hasil Pemeringkatan Ulasan

Urutan	Data latih ke-	Bobot BM25	Kelas
1	5	0,8812	Negatif
2	10	0,8325	Negatif
3	9	0,6615	Positif
4	8	0,5842	Positif
5	1	0,0000	Positif
6	2	0,0000	Positif
7	3	0,0000	Positif
8	4	0,0000	Positif
9	6	0,0000	Positif
10	7	0,0000	Positif

4.4.3 Perhitungan Manual Neighbor-Weighted K-Nearest Neighbor

Setelah mendapatkan nilai bobot kemiripan menggunakan metode BM25, maka langkah terakhir adalah menentukan klasifikasi kelas untuk data uji menggunakan metode NWKNN. Ada beberapa tahapan dalam metode NWKNN, yaitu: mengambil ulasan sejumlah nilai k, pembobotan tiap kelas, dan terakhir *scoring* NWKNN untuk mendapatkan hasil klasifikasi.

4.4.3.1 Mengambil Ulasan Sejumlah Nilai K

Pada perhitungan manual ini diasumsikan bahwa nilai k=5. Sesuai dengan nilai k yang diberikan, maka data latih yang telah diurutkan diambil 5 data teratas. Ulasan yang diambil sejumlah nilai k dapat dilihat pada Tabel 4.21.

Tabel 4.21 Ulasan Sejumlah Nilai K

Urutan	Data latih ke-	Bobot BM25	Kelas
1	5	0,8812	Negatif
2	10	0,8325	Negatif
3	9	0,6615	Positif
4	8	0,5842	Positif
5	1	0,0000	Positif

4.4.3.2 Pembobotan Kelas

Pada proses pembobotan tiap kelas menggunakan metode NWKNN, bobot tiap kelas berbanding terbalik dengan banyaknya kemunculan kelas tersebut. Kelas yang paling banyak muncul akan mendapatkan nilai bobot terendah, sementara kelas yang paling sedikit muncul akan mendapatkan nilai bobot tertinggi. Berikut dapat dilihat jumlah kelas pada data latih pada Tabel 4.22.

Tabel 4.22 Jumlah Kelas Positif dan Negatif

Kelas	Jumlah
Positif	8
Negatif	2

Mengacu pada Persamaan 2.3, contoh perhitungan manual pembobotan kelas pada metode NWKNN dengan nilai eksponen=2 dan jumlah kelas seperti pada Tabel 4.22 dijabarkan sebagai berikut dan ditampilkan pada Tabel 4.23.

$$Weight_{positif} = \frac{1}{\left(\frac{\text{jumlah data latih kelas "Positif"}}{\text{jumlah data latih minimum pada seluruh kelas}}\right)^{\frac{1}{exp}}} = \frac{1}{\frac{8^2}{2}} = 0,5$$

$$Weight_{negatif} = \frac{1}{\left(\frac{\text{jumlah data latih kelas "Negatif"}}{\text{jumlah data latih minimum pada seluruh kelas}}\right)^{\frac{1}{exp}}} = \frac{1}{\frac{2^2}{2}} = 1$$

Tabel 4.23 Hasil Pembobotan Kelas NWKNN

Kelas	Jumlah
Positif	0,5
Negatif	1

4.4.3.3 Scoring NWKNN

Setelah mendapatkan nilai bobot untuk tiap kelas yang digunakan, maka langkah terakhir adalah proses *scoring* NWKNN untuk mendapatkan nilai akhir serta menentukan klasifikasi kelas untuk data uji. Mengacu pada Persamaan 2.4, contoh proses perhitungan manual *scoring* NWKNN dijabarkan sebagai berikut.

$$\begin{aligned} Skor_{Positif} &= Weight_i \times \left(\sum_{d_j \in KNN(q)} Sim(q, d_j) \delta(d_j, c_i)\right) \\ &= 0,5 \times ((0,8812 \times 0) + (0,8325 \times 0) + (0,6615 \times 1) + \\ &\quad (0,5842 \times 1) + (0,000 \times 1)) \end{aligned}$$

$$\begin{aligned} &= 0,6228 \\ Skor_{Negatif} &= Weight_i \times \left(\sum_{d_j \in KNN(q)} Sim(q, d_j) \delta(d_j, c_i)\right) \\ &= 1 \times ((0,8812 \times 1) + (0,8325 \times 1) + (0,6615 \times 0) + \\ &\quad (0,5842 \times 0) + (0,000 \times 0)) \\ &= 1,7137 \end{aligned}$$

Berdasarkan perhitungan skor untuk tiap kelas yang digunakan, didapatkan hasil berupa kelas positif bernilai 1,7137 dan kelas negatif bernilai 0,6228. Maka dapat disimpulkan bahwa pada proses manualisasi ini, data uji yang digunakan termasuk dalam kelas negatif karena memiliki nilai skor yang lebih tinggi dibandingkan kelas positif. Hasil *scoring* NWKNN dapat dilihat pada Tabel 4.24.

Tabel 4.24 Hasil Perhitungan *Scoring* NWKNN

Kelas	Nilai	Hasil Klasifikasi
Positif	0,6228	Negatif
Negatif	1,7137	

4.4.4 Perhitungan Manual Nilai *Precision*, *Recall*, *F-measure*, dan *Accuracy*

Proses perhitungan ini digunakan untuk mengetahui kinerja sistem dalam melakukan klasifikasi. Perhitungan evaluasi yang digunakan berupa perhitungan *confusion matrix* yang meliputi nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Untuk memahami lebih jelas terkait penggunaan *confusion matrix* dalam proses evaluasi kinerja sistem, berikut contoh perhitungan *confusion matrix* menggunakan contoh data yang dapat dilihat pada Tabel 4.25.

Tabel 4.25 Contoh Hasil Klasifikasi

No.	Kelas Aktual	Kelas Klasifikasi
1.	Negatif	Negatif
2.	Negatif	Negatif
3.	Positif	Positif
4.	Positif	Negatif
5.	Negatif	Negatif
6.	Positif	Positif
7.	Positif	Positif
8.	Positif	Positif
9.	Negatif	Positif
10.	Positif	Positif

Setelah hasil klasifikasi diperoleh seperti pada Tabel 4.25, maka langkah selanjutnya adalah menghitung nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Untuk menghitung nilai tersebut diperlukan tabel *confusion matrix* berdasarkan hasil klasifikasi pada Tabel 4.25 yang dapat dilihat pada Tabel 4.26.

Tabel 4.26 Confusion Matrix

	Klasifikasi Positif	Klasifikasi Negatif
Aktual Positif	5 (TP)	1 (FN)
Aktual Negatif	1 (FP)	3 (TN)

Berdasarkan tabel *confusion matrix* pada Tabel 4.26 didapatkan nilai yang digunakan untuk proses mencari *f-measure*, *accuracy*, *precision*, dan *recall* menggunakan rumus pada Persamaan 2.5 hingga Persamaan 2.8. Perhitungan manual tersebut dijabarkan sebagai berikut.

$$Precision = \frac{TP}{TP+FP} = \frac{5}{5+1} = 0,8333$$

$$Recall = \frac{TP}{TP+FN} = \frac{5}{5+1} = 0,8333$$

$$F\text{-measure} = \frac{2 \times precision \times recall}{precision+recall} = \frac{2 \times 0,8333 \times 0,8333}{0,8333+0,8333} = 0,8333$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{5+3}{5+3+1+1} = 0,8$$

Berdasarkan hasil perhitungan manual tersebut didapatkan hasil berupa nilai *precision* sebesar 0,8333, *recall* sebesar 0,8333, *f-measure* sebesar 0,8333, dan *accuracy* sebesar 0,8.

4.5 Perancangan Skenario Pengujian

Perancangan skenario pengujian bertujuan untuk menjabarkan parameter apa saja yang perlu dilakukan pengujian. Skenario pengujian ini meliputi pengujian parameter nilai k1 dan nilai b pada metode BM25, dan pengujian parameter nilai k dan nilai eksponen pada metode NWKNN.

4.5.1 Pengujian Nilai K pada NWKNN

Pengujian nilai k dilakukan untuk mengetahui berapa nilai ketetangaan yang memberikan hasil terbaik pada proses klasifikasi. Pengujian ini dilakukan dengan memberikan nilai k yang berbeda beda pada proses klasifikasi menggunakan metode NWKNN. Variasi nilai k yang dilakukan pengujian yaitu 2, 3, 5, 10, 20, 25, 50, 75, 100, 150, 200, dan 250 menggunakan *5-fold cross validation*. Skenario pengujian nilai k dapat dilihat pada Tabel 4.29.

Tabel 4.27 Skenario Pengujian Nilai K

Nilai k	Precision	Recall	Accuracy	F-measure
2				
3				
5				
10				

Tabel 4.27 Skenario Pengujian Nilai K (Lanjutan)

Nilai k	Precision	Recall	Accuracy	F-measure
20				
25				
50				
75				
100				
150				
200				
250				

4.5.2 Pengujian Nilai Eksponen pada NWKNN

Pengujian nilai eksponen dilakukan untuk mengetahui nilai eksponen yang memberikan hasil terbaik pada proses klasifikasi. Pengujian ini dilakukan menggunakan nilai eksponen dalam rentang 2 – 10 menggunakan *5-fold cross validation* dengan nilai k terbaik pada pengujian nilai k. Skenario pengujian nilai eksponen dapat dilihat pada Tabel 4.28.

Tabel 4.28 Skenario Pengujian Nilai Eksponen

Nilai eksponen	Precision	Recall	Accuracy	F-measure
2				
3				
4				
5				
6				
7				
8				
9				
10				

4.5.3 Pengujian Nilai K1 pada BM25

Pengujian nilai k_1 dilakukan untuk mengetahui pengaruh perbedaan nilai k_1 pada hasil klasifikasi serta mencari nilai k_1 yang memberikan hasil terbaik. Pengujian ini mengikutsertakan metode NWKNN dengan nilai k dan nilai eksponen terbaik pada pengujian sebelumnya. Pengujian pada nilai k_1 dilakukan dalam

rentang nilai 1,2 hingga 2 dengan peningkatan tiap pengujian sebesar 0,1 dan menggunakan nilai $b=0,5$ serta *5-fold cross validation*. Alasan rentang nilai pengujian tersebut karena penelitian oleh Robertson & Zaragoza (2009) menjelaskan bahwa nilai k_1 pada rentang 1,2 – 2 memberikan hasil terbaik. Skenario pengujian nilai k_1 dapat dilihat pada Tabel 4.29.

Tabel 4.29 Skenario Pengujian Nilai K_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2				
1,3				
1,4				
1,5				
1,6				
1,7				
1,8				
1,9				
2,0				

4.5.4 Pengujian Nilai B pada BM25

Pengujian nilai b dilakukan untuk mengetahui pengaruh perbedaan nilai b pada hasil klasifikasi serta mencari nilai b yang memberikan hasil terbaik. Pengujian ini mengikutsertakan metode NWKNN dengan nilai k dan nilai eksponen terbaik pada pengujian sebelumnya. Pengujian pada nilai b dilakukan dalam rentang nilai 0,5 hingga 0,8 dengan peningkatan tiap pengujian sebesar 0,05 dan menggunakan nilai k_1 terbaik pada pengujian nilai k_1 serta *5-fold cross validation*. Alasan rentang nilai pengujian tersebut karena penelitian oleh Robertson & Zaragoza (2009) menjelaskan bahwa nilai b pada rentang 0,5 – 0,8 memberikan hasil terbaik. Skenario pengujian nilai b dapat dilihat pada Tabel 4.30.

Tabel 4.30 Skenario Pengujian Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5				
0,55				
0,6				
0,65				
0,7				
0,75				
0,8				

4.5.5 Pengujian Faktor Kesalahan Hasil Klasifikasi

Pengujian faktor kesalahan hasil klasifikasi pada sistem akan menggunakan nilai parameter terbaik dari hasil pengujian pada metode BM25 dan NWKNN. Faktor kesalahan yang dilakukan pengujian adalah adanya kata negasi “tidak” serta adanya sentimen ganda (positif dan negatif) dalam 1 ulasan. Karena dalam sistem yang dibangun tidak terdapat kamus kata yang dapat merubah kata negasi, maka pengujian kata negasi dilakukan dengan menghilangkan kata “tidak” pada *stoplist* agar tidak terhapus ketika proses *filtering*. Pengujian ini dilakukan tanpa menggunakan *5-fold cross validation*, namun hanya berfokus pada satu ulasan pada data uji yang mendapat hasil klasifikasi yang salah (mengandung kata negasi “tidak” dan sentimen ganda). Skenario pengujian faktor kesalahan hasil klasifikasi dilihat pada Tabel 4.31.

Tabel 4.31 Skenario Pengujian Faktor Kesalahan Hasil Klasifikasi

		<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
Pengujian Awal					
Negasi					
Sentimen Ganda	Positif				
	Negatif				
	Positif & Negatif				
Negasi dan Sentimen Ganda	Positif				
	Negatif				
	Positif & Negatif				

BAB 5 IMPLEMENTASI

Bab ini berisikan uraian tentang kode program yang telah dibuat sekaligus penjelasan dari tiap kode program tersebut. Uraian tersebut meliputi *preprocessing*, pemeringkatan menggunakan metode BM25, dan proses untuk mendapatkan hasil klasifikasi menggunakan metode NWKNN.

5.1 Implementasi Main Program

Implementasi *main* program merupakan bagian utama dimana program dijalankan. *Main* program berisikan pemanggilan terhadap *function - function* yang digunakan dalam sistem. Implementasi dari proses *cleaning* dapat dilihat pada Kode Sumber 5.1.

```

Algoritme: Main
1  def main():
2  #////////// DATA LATIH //////////#
3  All_dataLatih=pd.read_csv("Latih.csv")
4  dataLatih = All_dataLatih[["Ulasan","Hasil_Klasifikasi"]]
5
6  data_cleaning = cleaning(dataLatih)
7  data_caseFolding = case_folding(data_cleaning)
8  data_tokenizing = tokenizing(data_caseFolding)
9  data_filtering = filtering(data_tokenizing)
10 data_stemming, termsTiapUlasan, termsSeluruh =
11 stemming(data_filtering)
12 data_TF,lenUlasan = TF(data_stemming)
13
14 #////////// DATA UJI //////////#
15 All_dataUji=pd.read_csv("Uji.csv")
16 dataUji = All_dataUji[["Ulasan","Hasil_Klasifikasi"]]
17
18 dataUji_cleaning = cleaning(dataUji)
19 dataUji_caseFolding = case_folding(dataUji_cleaning)
20 dataUji_tokenizing = tokenizing(dataUji_caseFolding)
21 dataUji_filtering = filtering(dataUji_tokenizing)
22 dataUji_stemming, termsUji_TiapUlasan, termsUji_Seluruh =
23 stemming(dataUji_filtering)
24 dataUji_DF = DF(data_TF,termsUji_Seluruh)
25
26 hasilKlasifikasi = []
27 for i in range(len(dataUji_stemming)):
28     data_IDF = IDF(dataUji_DF,len(dataLatih),
29     termsUji_TiapUlasan[i])
30
31 data_BM25 = Scoring_BM25(lenUlasan,data_IDF,data_TF)
32 pemeringkatan_data_BM25 = Sorted_BM25(data_BM25)
33
34 nilaiK = 20
35 dataTerpilih = ambil_Kdata(nilaiK,
36 pemeringkatan_data_BM25)
37 bobotPositif,bobotNegatif =
38 bobotKelas(nilaiK,dataTerpilih,
39 dataLatih['Hasil_Klasifikasi'])
40
41 hasilKlasifikasi.append(score_NWKNN(dataTerpilih,
42 bobotPositif,bobotNegatif,dataLatih['Hasil_Klasifikasi']))

```

Kode Sumber 5.1 Main Program

Penjelasan dari Kode Sumber 5.1 dijabarkan sebagai berikut.

Baris 1 Deklarasi fungsi dengan nama *main*.

Baris 3 Membaca file bernama *Latih* dengan format *csv* dan disimpan ke dalam variabel *All_dataLatih*.

Baris 4 Mengambil data yang berada di kolom “*Ulasan*” dan “*Hasil Klasifikasi*” pada variabel *All_dataLatih* dan disimpan pada variabel *dataLatih*.

Baris 6 Proses memanggil *function cleaning* untuk mendapatkan hasil *cleaning* dari variabel *dataLatih* dan disimpan pada variabel *data_cleaning*.

Baris 7 Proses memanggil *function case_folding* untuk mendapatkan hasil *case folding* dari variabel *data_cleaning* dan disimpan pada variabel *data_caseFolding*.

Baris 8 Proses memanggil *function tokenizing* untuk mendapatkan hasil *tokenizing* dari variabel *data_caseFolding* dan disimpan pada variabel *data_tokenizing*.

Baris 9 Proses memanggil *function filtering* untuk mendapatkan hasil *filtering* dari variabel *data_tokenizing* dan disimpan pada variabel *data_filtering*.

Baris 10-11 Proses memanggil *function stemming* untuk mendapatkan hasil *stemming*, *term* pada tiap ulasan, dan *term* keseluruhan dari variabel *data_filtering* dan disimpan pada variabel *data_stemming*, *termsTiapUlasan*, *termsSeluruh*.

Baris 12 Proses memanggil *function TF* untuk mendapatkan hasil perhitungan *TF* dan panjang tiap ulasan dari variabel *data_stemming* kemudian disimpan pada variabel *data_TF*, *lenUlasan*.

Baris 15 Membaca file bernama *Uji* dengan format *csv* dan disimpan ke dalam variabel *All_dataUji*.

Baris 16 Mengambil data yang berada di kolom “*Ulasan*” dan “*Hasil Klasifikasi*” pada variabel *All_dataUji* dan disimpan pada variabel *dataUji*.

Baris 18 Proses memanggil *function cleaning* untuk mendapatkan hasil *cleaning* dari variabel *dataUji* dan disimpan pada variabel *dataUji_cleaning*.

Baris 19 Proses memanggil *function case_folding* untuk mendapatkan hasil *case folding* dari variabel *dataUji_cleaning* dan disimpan pada variabel *dataUji_caseFolding*.

Baris 20 Proses memanggil *function tokenizing* untuk mendapatkan hasil *tokenizing* dari variabel *dataUji_caseFolding* dan disimpan pada variabel *dataUji_tokenizing*.

Baris 21 Proses memanggil *function filtering* untuk mendapatkan hasil *filtering* dari variabel *dataUji_tokenizing* dan disimpan pada variabel *dataUji_filtering*.

Baris 22-23 Proses memanggil *function stemming* untuk mendapatkan hasil *stemming*, *term* pada tiap ulasan, dan *term* keseluruhan dari variabel *dataUji_filtering* dan disimpan pada variabel *dataUji_stemming*, *termsUji_TiapUlasan*, *termsUji_Seluruh*.

Baris 24 Proses memanggil *function DF* untuk mendapatkan hasil perhitungan DF dari variabel *data_TF* dan *termsUji_Seluruh* kemudian disimpan pada variabel *dataUji_DF*.

Baris 27 Inisialisasi variabel *hasilKlasifikasi*.

Baris 28 Proses perulangan sejumlah banyaknya data pada variabel *dataUji_stemming*.

Baris 29-30 Proses memanggil *function IDF* untuk mendapatkan hasil perhitungan IDF dari variabel *dataUji_DF*, jumlah data pada variabel *dataLatih*, dan data ke-i dari *termUji_TiapUlasan* kemudian disimpan pada variabel *data_IDF*.

Baris 32 Proses memanggil *function Scoring_BM25* untuk mendapatkan hasil perhitungan bobot untuk tiap ulasan menggunakan metode BM25 dari variabel *lenUlasan*, *data_IDF*, *data_TF* kemudian disimpan pada variabel *data_BM25*.

Baris 33 Proses memanggil *function Sorted_BM25* untuk mendapatkan hasil pemeringkatan bobot mulai dari yang terbesar hingga terkecil dari variabel *data_BM25* kemudian disimpan pada variabel *pemeringkatan_data_BM25*.

Baris 35 Deklarasi variabel *nilaiK* dengan nilai 20.

Baris 36-37 Proses memanggil *function ambil_Kdata* untuk mendapatkan data terpilih dari variabel *nilaiK* dan *pemeringkatan_data_BM25* kemudian disimpan pada variabel *dataTerpilih*.

Baris 38-40 Proses memanggil *function bobotKelas* untuk mendapatkan hasil perhitungan bobot untuk tiap kelas yang digunakan (positif dan negatif) dari variabel *nilaiK*, *dataTerpilih*, dan *dataLatih* pada kolom "Hasil Klasifikasi" kemudian disimpan pada variabel *bobotPositif* dan *bobotNegatif*.

Baris 41-42 Proses memanggil *function score_NWKNN* untuk mendapatkan hasil klasifikasi dari ulasan yang sedang diproses dari variabel *dataTerpilih*, *bobotPositif*, *bobotNegatif*, dan *dataLatih* pada

kolom “Hasil Klasifikasi” kemudian disimpan pada variabel *HasilKlasifikasi*.

5.2 Implementasi *Preprocessing*

Implementasi dari proses *preprocessing* terdiri atas lima tahapan yang diterapkan pada data latih maupun data uji, tahapan tersebut antara lain:

5.2.1 Implementasi *Cleaning*

Tahapan *cleaning* bertujuan untuk menghilangkan karakter dari suatu kalimat yang dirasa tidak mempunyai nilai, seperti tanda baca dan angka. Implementasi dari proses *cleaning* dapat dilihat pada Kode Sumber 5.2.

```

Algoritme: Cleaning
1  def cleaning(data_ulasan):
2      result = []
3      for ulasan in data_ulasan["Ulasan"]:
4          hasil = re.sub('[^A-Za-z]+', " ", ulasan)
5          result.append(hasil)
6      return result
    
```

Kode Sumber 5.2 *Cleaning*

Penjelasan dari Kode Sumber 5.2 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *cleaning* yang memiliki parameter *data_ulasan*.
- Baris 2 Inisialisasi variabel *result*
- Baris 3-5 Proses menghilangkan karakter bukan huruf pada keseluruhan kalimat dan menyimpan hasilnya pada variabel *result*.
- Baris 6 Mengembalikan nilai dari variabel *result*.

5.2.2 Implementasi *Case Folding*

Tahapan *case folding* bertujuan untuk merubah huruf kapital yang ada dalam kalimat menjadi huruf kecil (*lowercase*). Implementasi dari proses *case folding* dapat dilihat pada Kode Sumber 5.3.

```

Algoritme: Case Folding
1  def case_folding(data_ulasan):
2      result = []
3      for ulasan in data_ulasan:
4          hasil = ulasan.lower()
5          result.append(hasil)
6      return result
    
```

Kode Sumber 5.3 *Case Folding*

Penjelasan dari Kode Sumber 5.3 dijabarkan sebagai berikut.

Baris 1 Deklarasi fungsi dengan nama *case_folding* yang memiliki parameter *data_ulasan*.

Baris 2 Inisialisasi variabel *result*

Baris 3-5 Proses merubah setiap huruf kapital yang ditemukan pada tiap kalimat dengan huruf kecil (*lowercase*) dan menyimpan hasilnya pada variabel *result*.

Baris 6 Mengembalikan nilai dari variabel *result*.

5.2.3 Implementasi *Tokenizing*

Tahapan *tokenizing* bertujuan untuk memisahkan tiap kata pada sebuah kalimat. Implementasi dari proses *tokenizing* dapat dilihat pada Kode Sumber 5.4.

```

Algoritme: Tokenizing
1 def tokenizing(data_ulasan):
2     result = []
3     for ulasan in data_ulasan:
4         hasil = ulasan.split()
5         result.append(hasil)
6     return result
    
```

Kode Sumber 5.4 *Tokenizing*

Penjelasan dari Kode Sumber 5.4 dijabarkan sebagai berikut.

Baris 1 Deklarasi fungsi dengan nama *tokenizing* yang memiliki parameter *data_ulasan*.

Baris 2 Inisialisasi variabel *result*

Baris 3-5 Proses memisahkan tiap kata yang ada pada kalimat dengan parameter pemisah adalah karakter spasi dan menyimpan hasilnya pada variabel *result*.

Baris 6 Mengembalikan nilai dari variabel *result*.

5.2.4 Implementasi *Filtering*

Tahapan *filtering* bertujuan untuk menghilangkan kata atau token hasil dari proses *tokenizing* yang dianggap tidak penting. Implementasi dari proses *filtering* dapat dilihat pada Kode Sumber 5.5.

```

Algoritme: Filtering
1 def filtering(data_ulasan):
2     result = []
3     stopwords = open("stopword_list_tala.txt", "r").read()
4     stopword = stopwords.split("\n")
5     for teks in data_ulasan:
    
```

```

6   arrBaru = []
7   for kata in teks:
8       if kata not in stopwords:
9           arrBaru.append(kata)
10  result.append(arrBaru)
11  return result

```

Kode Sumber 5.5 Filtering

Penjelasan dari Kode Sumber 5.5 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *filtering* yang memiliki parameter *data_ulasan*.
- Baris 2 Inisialisasi variabel *result*.
- Baris 3 Membuka *file stoplist_list_tala.txt* dan menyimpan seluruh kata pada *file* tersebut dalam variabel *stopwords*.
- Baris 4 Memisahkan tiap kata yang ada dalam variabel *stopwords* berdasarkan karakter enter dan disimpan pada variabel *stopword*.
- Baris 5-10 Proses menghilangkan tiap kata yang dirasa tidak penting pada parameter *data_ulasan*, dimana kata yang dirasa tidak penting berasal dari *file stoplist_list_tala.txt* yang terpisah tiap katanya dan disimpan pada variabel *stopword*. Hasil dari proses ini disimpan pada variabel *result*.
- Baris 11 Mengembalikan nilai dari variabel *result*.

5.2.5 Implementasi Stemming

Tahapan *stemming* bertujuan untuk mengembalikan kata yang memiliki imbuhan menjadi kata dasar. Implementasi dari proses *stemming* dapat dilihat pada Kode Sumber 5.6.

```

Algoritme: Stemming
1   def stemming(data_ulasan):
2       result = []
3       termsTiapUlasan = []
4       termsSeluruh = []
5       factory = StemmerFactory()
6       stemmer = factory.create_stemmer()
7
8       for teks in data_ulasan:
9           arrBaru = []
10          term = []
11          for kata in teks:
12              kata_dasar = stemmer.stem(kata)

```

```

13 arrBaru.append(kata_dasar)
14 if kata_dasar not in term:
15     term.append(kata_dasar)
16     if kata_dasar not in termsSeluruh:
17         termsSeluruh.append(kata_dasar)
18 termsTiapUlasan.append(term)
19 result.append(arrBaru)
20
21 return result, termsTiapUlasan, termsSeluruh
    
```

Kode Sumber 5.6 Stemming

Penjelasan dari Kode Sumber 5.6 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *stemming* yang memiliki parameter *data_ulasan*.
- Baris 2 Inisialisasi variabel *result*.
- Baris 3 Inisialisasi variabel *termsTiapUlasan*.
- Baris 4 Inisialisasi variabel *termsSeluruh*.
- Baris 5 Proses memanggil fungsi *stemmer*.
- Baris 6 Menyimpan *stemmer* pada variabel *stemmer*.
- Baris 8-19 Proses mengembalikan tiap kata yang berimbuhan menjadi kata dasar lalu menyimpannya pada variabel *result*, menyimpan *term* unik untuk tiap ulasan pada variabel *termsTiapUlasan*, dan menyimpan keseluruhan *term* unik untuk seluruh data pada variabel *termsSeluruh*.
- Baris 21 Mengembalikan nilai dari variabel *result*, *termsTiapUlasan*, dan *termsSeluruh*.

5.3 Implementasi BM25

Proses pembobotan dan pemeringkatan menggunakan metode BM25 melalui beberapa tahapan, yaitu: menghitung TF pada data latih, DF, IDF, *scoring* BM25, dan pemeringkatan ulasan.

5.3.1 Implementasi Perhitungan Term Frequency

Perhitungan TF dilakukan untuk menghitung berapa banyak frekuensi kemunculan *term* pada ulasan data latih. Implementasi dari perhitungan TF dapat dilihat pada Kode Sumber 5.7.

```

Algoritme: TF
1 def TF(data ulasan):
2     result = []
3     panjangData = []
    
```

```

4 for data in data_ulasan:
5     panjangData.append(len(data))
6     result.append(Counter(data))
7 return result,panjangData

```

Kode Sumber 5.7 Perhitungan TF

Penjelasan dari Kode Sumber 5.7 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *TF* yang memiliki parameter *data_ulasan*.
- Baris 2 Inisialisasi variabel *result*.
- Baris 3 Inisialisasi variabel *panjangData*.
- Baris 3 Inisialisasi variabel *termsTiapUlasan*.
- Baris 4-6 Proses menghitung kemunculan *term* pada tiap ulasan lalu menyimpannya pada variabel *result* dan menghitung panjang ulasan berdasarkan banyaknya *term* yang dimiliki lalu menyimpannya pada variabel *panjangData*.
- Baris 21 Mengembalikan nilai dari variabel *result* dan *panjangData*.

5.3.2 Implementasi Perhitungan Document Frequency

Perhitungan DF dilakukan untuk menghitung jumlah ulasan data latih yang memiliki *term* tertentu dari ulasan data uji. Implementasi dari perhitungan DF dapat dilihat pada Kode Sumber 5.8.

```

Algoritme: DF
1 def DF(data_ulasan,terms):
2     result = dict({})
3     for term in terms:
4         counter = 0
5         for data in data_ulasan:
6             if term in data:
7                 counter+=1
8     result[term] = counter
9 return result

```

Kode Sumber 5.8 Perhitungan DF

Penjelasan dari Kode Sumber 5.8 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *DF* yang memiliki parameter *data_ulasan* dan *terms*.
- Baris 2 Inisialisasi variabel *result* yang bertipe data *dictionary*.
- Baris 3-8 Proses menghitung nilai DF tiap *term* data uji. Nilai *term* tersebut akan bertambah 1 untuk tiap ulasan data latih yang terdapat

term tersebut. Hasil proses ini disimpan pada variabel *result* dengan *key* nama *term* dan *value* berupa nilai *term* tersebut.

Baris 9 Mengembalikan nilai dari variabel *result*

5.3.3 Implementasi Perhitungan *Inverse Document Frequency*

Perhitungan IDF dilakukan setelah mendapatkan nilai DF. Tujuan perhitungan ini untuk mengetahui nilai penyebaran *term* pada ulasan yang digunakan. Implementasi dari perhitungan IDF dapat dilihat pada Kode Sumber 5.9.

```

Algorithm: IDF
1 def IDF (dataDF, totalData, terms):
2     result = dict({})
3     for term in terms:
4         if term in dataDF.keys():
5             nilai = math.log10((totalData-
6 dataDF[term]+0.5) / (dataDF[term]+0.5))
7             result[term] = nilai
8     return result
    
```

Kode Sumber 5.9 Perhitungan IDF

Penjelasan dari Kode Sumber 5.9 dijabarkan sebagai berikut.

Baris 1 Deklarasi fungsi dengan nama *IDF* yang memiliki parameter *dataDF*, *totalData*, dan *terms*.

Baris 2 Inisialisasi variabel *result* yang bertipe data *dictionary*.

Baris 3-7 Proses perhitungan nilai IDF untuk tiap *term* yang didapatkan.

Baris 8 Mengembalikan nilai dari variabel *result*

5.3.4 Implementasi Perhitungan *Scoring BM25*

Scoring BM25 dilakukan untuk mendapatkan nilai bobot untuk tiap data latihan yang digunakan. Implementasi dari *scoring* BM25 dapat dilihat pada Kode Sumber 5.10.

```

Algorithm: Scoring BM25
1 def Scoring_BM25 (panjangKalimat, idf, tf):
2     rata = sum(panjangKalimat) / len(panjangKalimat)
3     terms = idf.keys()
4     totalBM25 = {}
5     for i in range(len(tf)):
6         nilaiTerm = []
7         for term in terms:
8             if (tf[i][term] != 0):
9                 pembilang = tf[i][term] * (1.2+1)
    
```

```

10     penyebut = tf[i][term] + 1.2 * (1-
11     0.75+0.75*(panjangKalimat[i]/rata))
12     bm25 = idf[term]*(pembilang/penyebut)
13     nilaiTerm.append(bm25)
14     totalBM25[i]=sum(nilaiTerm)
15     return totalBM25

```

Kode Sumber 5.10 Perhitungan Scoring BM25

Penjelasan dari Kode Sumber 5.10 dijabarkan sebagai berikut.

Baris 1 Deklarasi fungsi dengan nama *Scoring_BM25* yang memiliki parameter *panjangKalimat*, *idf*, dan *tf*.

Baris 2 Inisialisasi variabel *rata* yang nilainya merupakan hasil bagi antara total penjumlahan pada parameter *panjangKalimat* dengan banyaknya data pada parameter *panjangKalimat*.

Baris 3 Inisialisasi variabel *terms* yang nilainya didapatkan dari *key* parameter *idf*.

Baris 4 Inisialisasi variabel *totalBM25*.

Baris 5-14 Proses *scoring* BM25 yang nilainya didapatkan dari proses perhitungan dengan menggunakan keseluruhan parameter yang ada pada fungsi ini, untuk mendapatkan bobot akhir pada tiap data latih yang digunakan.

Baris 15 Mengembalikan nilai dari variabel *totalBM25*.

5.3.5 Implementasi Pemeringkatan Ulasan

Proses pemeringkatan ulasan dilakukan setelah setiap data latih sudah memiliki nilai bobot. Proses pemeringkatan ini akan mengurutkan data latih mulai dari yang memiliki nilai bobot kemiripan terbesar hingga terkecil. Implementasi proses pemeringkatan ulasan dapat dilihat pada Kode Sumber 5.11.

```

Algoritme: Pemeringkatan ulasan
1  def Sorted_BM25(data):
2      dataKe = []
3      nilaiKe = []
4      sort = []
5      for key,val in data.items():
6          nilaiKe.append(val)
7          dataKe.append(key)
8
9      for i in range(len(data)):
10         nilaiMax = -sys.maxsize-1
11         index = 0
12         for j in range(len(dataKe)):

```

```

13     if(nilaiMax < nilaiKe[j]):
14         nilaiMax = nilaiKe[j]
15         index = dataKe[j]
16         hapus = j
17     del dataKe[hapus]
18     del nilaiKe[hapus]
19     sort.append([index, nilaiMax])
20     return sort
    
```

Kode Sumber 5.11 Pemeringkatan Ulasan

Penjelasan dari Kode Sumber 5.11 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *Sorted_BM25* yang memiliki parameter *data*.
- Baris 2 Inisialisasi variabel *dataKe*.
- Baris 3 Inisialisasi variabel *nilaiKe*.
- Baris 4 Inisialisasi variabel *sort*.
- Baris 5-7 Proses perulangan untuk mengambil *key* yang mewakili urutan ulasan data latih dan *value* yang mewakili nilai bobot data latih pada parameter *data*. Nilai *value* akan disimpan dalam variabel *nilaiKe* dan nilai *key* akan disimpan dalam variabel *dataKe*.
- Baris 9-19 Proses yang dilakukan untuk mengurutkan data berdasarkan nilai bobot terbesar ke yang terkecil. Proses ini akan membandingkan nilai bobot pada tiap data dan secara terurut menyimpan data yang bernilai paling besar ke yang paling kecil pada variabel *sort*. Proses ini akan terus berlangsung sejumlah banyaknya data latih yang digunakan.
- Baris 20 Mengembalikan nilai dari variabel *sort*.

5.4 Implementasi NWKNN

Proses klasifikasi menggunakan metode NWKNN melalui beberapa tahapan, yaitu: mengambil data sejumlah nilai k, pembobotan tiap kelas, dan terakhir *scoring* NWKNN untuk mendapatkan hasil klasifikasi.

5.4.1 Implementasi Mengambil Data Sejumlah Nilai K

Pada proses ini data yang diambil sejumlah nilai k adalah data yang telah diurutkan selama proses pembobotan dan pemeringkatan BM25. Implementasi dari pengambilan data sejumlah nilai k dapat dilihat pada Kode Sumber 5.12.

Algoritme: Mengambil data sejumlah nilai k

```

1 def ambil_Kdata(k, data):
2     result = {}
3     for i in range(k):
    
```

```

4 result[data[i][0]] = data[i][1]
5 return result

```

Kode Sumber 5.12 Mengambil Data Sejumlah Nilai K

Penjelasan dari Kode Sumber 5.12 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *ambil_kdata* yang memiliki parameter *k* dan *data*.
- Baris 2 Inisialisasi variabel *result*.
- Baris 3-4 Proses perulangan untuk mengambil data sejumlah nilai *k*.
- Baris 5 Mengembalikan nilai dari variabel *result*.

5.4.2 Implementasi Pembobotan Tiap Kelas

Proses pembobotan tiap kelas dilakukan untuk menghitung nilai bobot dari kelas yang digunakan. Kelas yang memiliki kemunculan paling banyak akan mendapatkan nilai bobot terendah, sementara kelas dengan kemunculan paling sedikit akan mendapatkan nilai bobot tertinggi. Implementasi dari pembobotan tiap kelas dapat dilihat pada Kode Sumber 5.13.

```

Algoritme: Pembobotan tiap kelas
1 def bobotKelas(k, data, kelas):
2     result = {}
3     getKelas = {}
4     keys = []
5     for key in data:
6         keys.append(key)
7     for i in range(k):
8         getKelas[data[keys[i]]] = kelas[keys[i]]
9
10    kelas = Counter(kelas)
11    eks = 2
12
13    bobotPositif = 1 / ((kelas["Positif"] /
14 (min(kelas["Positif"], kelas["Negatif"]))) ** (1/eks))
15    bobotNegatif = 1 / ((kelas["Negatif"] /
16 (min(kelas["Positif"], kelas["Negatif"]))) ** (1/eks))
17    return bobotPositif, bobotNegatif
18

```

Kode Sumber 5.13 Pembobotan Tiap Kelas

Penjelasan dari Kode Sumber 5.13 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *bobotKelas* yang memiliki parameter *k*, *data*, dan *kelas*.
- Baris 2 Inisialisasi variabel *result*.
- Baris 3 Inisialisasi variabel *getKelas*.
- Baris 4 Inisialisasi variabel *keys*.
- Baris 5-6 Proses perulangan untuk mengambil seluruh *key* dari parameter *data* dan disimpan pada variabel *keys*.
- Baris 8-9 Proses perulangan sejumlah nilai parameter *k* yang bertujuan untuk mendapatkan keseluruhan kelas dari parameter *data* yang kemudian disimpan pada variabel *getKelas*.
- Baris 11 Inisialisasi variabel *kelas* dengan nilai berupa banyaknya kelas positif dan negatif dari parameter *kelas*.
- Baris 12 Inisialisasi variabel *eks* yang merupakan nilai *eksponen*.
- Baris 14 Proses menghitung bobot untuk kelas positif yang hasilnya akan disimpan pada variabel *boborPositif*.
- Baris 15 Proses menghitung bobot untuk kelas negatif yang hasilnya akan disimpan pada variabel *bobotNegatif*.
- Baris 17 Mengembalikan nilai dari variabel *bobotPositif* dan *bobotNegatif*.

5.4.3 Implementasi Scoring NWKNN

Proses *scoring* NWKNN merupakan proses terakhir sekaligus penentu hasil klasifikasi kelas untuk data uji. Implementasi dari *scoring* NWKNN dapat dilihat pada Kode Sumber 5.14.

```

Algoritme: Scoring NWKNN
1 def score_NWKNN(data,positif,negatif,kelas):
2     resultPositif = 0
3     resultNegatif = 0
4     keys =[]
5     result = ""
6     for key in data:
7         keys.append(key)
8
9     for i in range(len(data)):
10        if(kelas[keys[i]] == "Positif"):
11            resultPositif += data[keys[i]]*1
12        else:
13            resultPositif += data[keys[i]]*0
    
```

```

14
15     if(kelas[keys[i]] == "Negatif"):
16         resultNegatif += data[keys[i]]*1
17     else:
18         resultNegatif += data[keys[i]]*0
19
20     resultPositif *= positif
21     resultNegatif *= negatif
22
23     if(resultPositif > resultNegatif):
24         result = "Positif"
25     else:
26         result = "Negatif"
27     return result

```

Kode Sumber 5.14 Scoring NWKNN

Penjelasan dari Kode Sumber 5.14 dijabarkan sebagai berikut.

- Baris 1 Deklarasi fungsi dengan nama *score_NWKNN* yang memiliki parameter *data*, *positif*, *negatif*, dan *kelas*.
- Baris 2 Inisialisasi variabel *resultPositif* dengan nilai 0.
- Baris 3 Inisialisasi variabel *resultNegatif* dengan nilai 0.
- Baris 4 Inisialisasi variabel *keys*.
- Baris 5 Inisialisasi variabel *result*.
- Baris 6-7 Proses perulangan untuk mengambil seluruh *key* dari parameter *data* dan disimpan pada variabel *keys*.
- Baris 9-18 Proses *scoring* NWKNN untuk mendapatkan skor positif dan negatif. Skor positif akan disimpan dalam variabel *resultPositif*, sementara skor negatif akan disimpan dalam variabel *resultNegatif*.
- Baris 20 Nilai dari variabel *resultPositif* dikalikan dengan nilai dari parameter positif.
- Baris 21 Nilai dari variabel *resultNegatif* dikalikan dengan nilai dari parameter negatif.
- Baris 23-26 Proses seleksi kondisi untuk mendapatkan hasil klasifikasi data uji.
- Baris 27 Mengembalikan nilai dari variabel *result*.

BAB 6 PENGUJIAN DAN ANALISIS

Bab ini berisikan penjabaran terkait hasil pengujian untuk tiap skenario yang telah ditulis pada bab perancangan. Hasil dari pengujian tersebut kemudian dilakukan analisis untuk melihat pengaruh dari pengujian tiap nilai parameter yang diujikan terhadap hasil evaluasi pada sistem yang dibuat serta mengetahui berapa nilai terbaik dari setiap parameter yang diujikan.

6.1 Pengujian Nilai K dengan 5-Fold Cross Validation

Pengujian ini dilakukan dengan membagi dataset menjadi 5 bagian *fold*, dimana 1 *fold* sebagai data uji dan 4 *fold* sebagai data latih. Pengujian pada nilai k dilakukan tanpa menggunakan metode BM25, namun hanya menggunakan pembobotan TF-IDF. Tujuannya untuk mengetahui nilai k terbaik ketika hanya menggunakan metode NWKNN saja. Setiap nilai k yang diujikan akan dicari berapa nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Nilai k terbaik didapatkan dengan menghitung rata – rata *f-measure* dan *accuracy* pada tiap nilai k yang diujikan.

6.1.1 Pengujian *Fold* ke-1

Pada *fold* ke-1 data uji yang digunakan adalah dataset pada urutan ke 1-200, sementara data latih menggunakan dataset pada urutan ke 201-1000. Hasil pengujian *fold* ke-1 dari pengujian nilai k dapat dilihat pada Tabel 6.1.

Tabel 6.1 Hasil Pengujian *Fold* ke-1 pada Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,8783	0,8552	0,8000	0,8666
3	0,8843	0,8552	0,8050	0,8695
5	0,8924	0,9276	0,8600	0,9096
10	0,8888	0,9473	0,8700	0,9171
20	0,8937	0,9407	0,8700	0,9166
25	0,9056	0,9473	0,8850	0,9260
50	0,8848	0,9605	0,8750	0,9211
75	0,8750	0,9671	0,8700	0,9187
100	0,8713	0,9802	0,8750	0,9226
150	0,8713	0,9802	0,8750	0,9226
200	0,8563	0,9802	0,8600	0,9141
250	0,8563	0,9802	0,8600	0,9141

6.1.2 Pengujian *Fold* ke-2

Pada *fold* ke-2 data uji yang digunakan adalah dataset pada urutan ke 201-400, sementara data latih menggunakan dataset pada urutan ke 1-200 dan 401-1000. Hasil pengujian *fold* ke-2 dari pengujian nilai k dapat dilihat pada Tabel 6.2.

Tabel 6.2 Hasil Pengujian *Fold* ke-2 pada Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9935	0,8423	0,85	0,9117
3	0,9712	0,9184	0,9	0,9441
5	0,9826	0,9239	0,915	0,9523
10	0,9783	0,9836	0,965	0,981
20	0,9728	0,9728	0,95	0,9728
25	0,9625	0,9782	0,945	0,9703
50	0,9731	0,9836	0,96	0,9783
75	0,9731	0,9836	0,96	0,9783
100	0,9731	0,9836	0,96	0,9783
150	0,9531	0,9945	0,95	0,9734
200	0,9432	0,9945	0,94	0,9682
250	0,9432	0,9945	0,94	0,9682

6.1.3 Pengujian *Fold* ke-3

Pada *fold* ke-3 data uji yang digunakan adalah dataset pada urutan ke 401-600, sementara data latih menggunakan dataset pada urutan ke 1-400 dan 601-1000. Hasil pengujian *fold* ke-3 dari pengujian nilai k dapat dilihat pada Tabel 6.3.

Tabel 6.3 Hasil Pengujian *Fold* ke-3 pada Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9512	0,8715	0,845	0,9096
3	0,9435	0,9329	0,89	0,9382
5	0,9555	0,9608	0,925	0,9582
10	0,9617	0,9832	0,95	0,9723
20	0,9619	0,9888	0,955	0,9752
25	0,9619	0,9888	0,955	0,9752
50	0,9619	0,9888	0,955	0,9752
75	0,9619	0,9888	0,955	0,9752

Tabel 6.3 Hasil Pengujian *Fold* ke-3 pada Nilai K (Lanjutan)

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
100	0,9619	0,9888	0,955	0,9752
150	0,9567	0,9888	0,95	0,9725
200	0,9569	0,9944	0,955	0,9753
250	0,9569	0,9944	0,955	0,9753

6.1.4 Pengujian *Fold* ke-4

Pada *fold* ke-4 data uji yang digunakan adalah dataset pada urutan ke 601-800, sementara data latih menggunakan dataset pada urutan ke 1-600 dan 801-1000. Hasil pengujian *fold* ke-4 dari pengujian nilai k dapat dilihat pada Tabel 6.4.

Tabel 6.4 Hasil Pengujian *Fold* ke-4 pada Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9183	0,931	0,89	0,9246
3	0,9	0,931	0,875	0,9152
5	0,9019	0,9517	0,89	0,9261
10	0,8967	0,9586	0,89	0,9266
20	0,8875	0,9793	0,895	0,9311
25	0,8867	0,9724	0,89	0,9276
50	0,8658	0,9793	0,875	0,919
75	0,8511	0,9862	0,865	0,9137
100	0,843	1	0,865	0,9148
150	0,8192	1	0,84	0,9006
200	0,8146	1	0,835	0,8978
250	0,8146	1	0,835	0,8978

6.1.5 Pengujian *Fold* ke-5

Pada *fold* ke-5 data uji yang digunakan adalah dataset pada urutan ke 801-1000, sementara data latih menggunakan dataset pada urutan ke 1-800. Hasil pengujian *fold* ke-5 dari pengujian nilai k dapat dilihat pada Tabel 6.5.

Tabel 6.5 Hasil Pengujian *Fold* ke-5 pada Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9160	0,8571	0,845	0,8856
3	0,9140	0,8357	0,83	0,8731

Tabel 6.5 Hasil Pengujian *Fold* ke-5 pada Nilai K (Lanjutan)

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
5	0,8911	0,9357	0,875	0,9128
10	0,8851	0,9357	0,87	0,9097
20	0,8758	0,9571	0,875	0,9146
25	0,8645	0,9571	0,865	0,9084
50	0,8616	0,9785	0,875	0,9163
75	0,8404	0,9785	0,855	0,9042
100	0,8353	0,9785	0,85	0,9013
150	0,8263	0,9857	0,845	0,899
200	0,8117	0,9857	0,83	0,8903
250	0,7976	0,9857	0,815	0,8817

6.1.6 Rata Rata Hasil Pengujian Nilai K

Setelah menghitung nilai *precision*, *recall*, *f-measure*, dan *accuracy* pada tiap nilai k menggunakan *5-fold cross validation*, maka selanjutnya nilai tersebut akan dirata – rata. Nilai hasil rata – rata dari setiap pengujian nilai k digunakan untuk mendapatkan nilai k terbaik. Hasil rata – rata pengujian nilai k dapat dilihat pada Tabel 6.6.

Tabel 6.6 Rata – Rata Hasil Pengujian Nilai K

Nilai k	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9315	0,8714	0,846	0,8996
3	0,9226	0,8946	0,86	0,908
5	0,9247	0,9399	0,893	0,9318
10	0,9221	0,9617	0,909	0,9413
20	0,9183	0,9677	0,909	0,9421
25	0,9162	0,9688	0,908	0,9415
50	0,9094	0,9781	0,908	0,942
75	0,9003	0,9808	0,901	0,938
100	0,8969	0,9862	0,901	0,9384
150	0,8853	0,9898	0,892	0,9336
200	0,8765	0,9909	0,884	0,9291
250	0,8737	0,9909	0,8809	0,9274

Berdasarkan Tabel 6.6 didapatkan bahwa nilai k terbaik terdapat pada pengujian ketika nilai k = 20, yang terbukti dengan rata – rata *f-measure* tertinggi yaitu sebesar 0,9421 diikuti dengan nilai rata – rata *accuracy* sebesar 0,909, nilai rata – rata *precision* sebesar 0,9183, dan nilai rata – rata *recall* sebesar 0,9677. Nilai k = 20 akan digunakan pada pengujian selanjutnya yaitu pengujian nilai eksponen.

6.2 Pengujian Nilai Eksponen dengan 5-Fold Cross Validation

Pengujian pada nilai eksponen sama seperti pada pengujian nilai k, dimana dataset terbagi menjadi 5 bagian *fold* dengan pembagian 1 *fold* sebagai data uji dan 4 *fold* sebagai data latih. Pengujian pada nilai eksponen juga dilakukan tanpa menggunakan metode BM25, namun hanya menggunakan metode pembobotan TF-IDF dan metode klasifikasi NWKNN. Setiap nilai eksponen yang diujikan akan dicari nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Nilai eksponen terbaik didapatkan dengan menghitung rata – rata *f-measure* dan *accuracy* pada tiap nilai eksponen yang diujikan.

6.2.1 Pengujian Fold ke-1

Pada *fold* ke-1 data uji yang digunakan adalah dataset pada urutan ke 1-200, sementara data latih menggunakan dataset pada urutan ke 201-1000. Hasil pengujian *fold* ke-1 dari pengujian nilai eksponen dapat dilihat pada Tabel 6.7.

Tabel 6.7 Hasil Pengujian *Fold* ke-1 pada Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,8937	0,9407	0,87	0,9166
3	0,8787	0,9539	0,865	0,9148
4	0,8795	0,9605	0,87	0,9182
5	0,8795	0,9605	0,87	0,9182
6	0,8639	0,9605	0,855	0,9096
7	0,8546	0,9671	0,85	0,9074
8	0,8448	0,9671	0,84	0,9018
9	0,8457	0,9736	0,845	0,9051
10	0,8457	0,9736	0,845	0,9051

6.2.2 Pengujian Fold ke-2

Pada *fold* ke-2 data uji yang digunakan adalah dataset pada urutan ke 201-400, sementara data latih menggunakan dataset pada urutan ke 1-200 dan 401-1000. Hasil pengujian *fold* ke-2 dari pengujian nilai eksponen dapat dilihat pada Tabel 6.8.

Tabel 6.8 Hasil Pengujian *Fold* ke-2 pada Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9728	0,9728	0,95	0,9728
3	0,9679	0,9836	0,955	0,9757
4	0,9528	0,9891	0,945	0,9706
5	0,9479	0,9891	0,94	0,968
6	0,9479	0,9891	0,94	0,968
7	0,9479	0,9891	0,94	0,968
8	0,9432	0,9945	0,94	0,9682
9	0,9432	0,9945	0,94	0,9682
10	0,9432	0,9945	0,94	0,9682

6.2.3 Pengujian *Fold* ke-3

Pada *fold* ke-3 data uji yang digunakan adalah dataset pada urutan ke 401-600, sementara data latih menggunakan dataset pada urutan ke 1-400 dan 601-1000. Hasil pengujian *fold* ke-3 dari pengujian nilai eksponen dapat dilihat pada Tabel 6.9.

Tabel 6.9 Hasil Pengujian *Fold* ke-3 pada Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9619	0,9888	0,955	0,9752
3	0,9567	0,9888	0,95	0,9725
4	0,9516	0,9888	0,945	0,9698
5	0,9516	0,9888	0,945	0,9698
6	0,9516	0,9888	0,945	0,9698
7	0,9516	0,9888	0,945	0,9698
8	0,9516	0,9888	0,945	0,9698
9	0,9516	0,9888	0,945	0,9698
10	0,9516	0,9888	0,945	0,9698

6.2.4 Pengujian *Fold* ke-4

Pada *fold* ke-4 data uji yang digunakan adalah dataset pada urutan ke 601-800, sementara data latih menggunakan dataset pada urutan ke 1-600 dan 801-1000. Hasil pengujian *fold* ke-4 dari pengujian nilai eksponen dapat dilihat pada Tabel 6.10.

Tabel 6.10 Hasil Pengujian *Fold* ke-4 pada Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,8875	0,9793	0,895	0,9311
3	0,8421	0,9931	0,86	0,9113
4	0,8421	0,9931	0,86	0,9113
5	0,8323	0,9931	0,85	0,9056
6	0,8135	0,9931	0,83	0,8944
7	0,8135	0,9931	0,83	0,8944
8	0,8135	0,9931	0,83	0,8944
9	0,8135	0,9931	0,83	0,8944
10	0,81	1	0,83	0,895

6.2.5 Pengujian *Fold* ke-5

Pada *fold* ke-5 data uji yang digunakan adalah dataset pada urutan ke 801-1000, sementara data latih menggunakan dataset pada urutan ke 1-800. Hasil pengujian *fold* ke-5 dari pengujian nilai eksponen dapat dilihat pada Tabel 6.11.

Tabel 6.11 Hasil Pengujian *Fold* ke-5 pada Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,8758	0,9571	0,875	0,9146
3	0,8509	0,9785	0,865	0,9102
4	0,8466	0,9857	0,865	0,9108
5	0,8414	0,9857	0,86	0,9078
6	0,8263	0,9857	0,845	0,899
7	0,8263	0,9857	0,845	0,899
8	0,8235	1	0,85	0,9032
9	0,8235	1	0,85	0,9032
10	0,8235	1	0,85	0,9032

6.2.6 Rata Rata Hasil Pengujian Nilai Eksponen

Setelah menghitung nilai *f-measure*, *accuracy*, *precision*, dan *recall* pada tiap nilai eksponen menggunakan *5-fold cross validation*, maka selanjutnya nilai tersebut akan dirata – rata. Nilai hasil rata – rata dari setiap pengujian nilai eksponen digunakan untuk mendapatkan nilai eksponen terbaik. Hasil rata – rata pengujian nilai eksponen dapat dilihat pada Tabel 6.12.

Tabel 6.12 Rata – Rata Hasil Pengujian Nilai Eksponen

Nilai eksponen	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
2	0,9183	0,9677	0,909	0,9421
3	0,8992	0,9796	0,8989	0,9369
4	0,8945	0,9834	0,8969	0,9362
5	0,8905	0,9834	0,893	0,9339
6	0,8806	0,9834	0,883	0,9282
7	0,8788	0,9847	0,882	0,9277
8	0,8753	0,9887	0,8809	0,9275
9	0,8755	0,99	0,882	0,9281
10	0,8748	0,9914	0,882	0,9283

Berdasarkan Tabel 6.12 didapatkan bahwa nilai eksponen terbaik terdapat pada pengujian ketika nilai eksponen = 2, yang terbukti dengan rata – rata *f-measure* tertinggi yaitu sebesar 0,9421 diikuti dengan nilai rata – rata *accuracy* sebesar 0,909, nilai rata – rata *precision* sebesar 0,9183, nilai rata – rata *recall* sebesar 0,9677. Nilai k = 20 dan nilai eksponen = 2 akan digunakan pada pengujian selanjutnya yaitu pengujian pada nilai k_1 dan b pada metode BM25.

6.3 Pengujian Nilai k_1 dengan *5-Fold Cross Validation*

Pengujian pada nilai k_1 juga menggunakan *5-fold cross validation*, dimana terbagi menjadi 5 bagian *fold* dengan pembagian 1 *fold* sebagai data uji dan 4 *fold* sebagai data latih. Proses pengujian nilai k_1 pada metode BM25 akan menggunakan nilai k dan eksponen terbaik pada metode NWKNN yang didapatkan berdasarkan pengujian sebelumnya. Setiap nilai k_1 yang diujikan akan dicari nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Nilai k_1 terbaik didapatkan dengan menghitung rata – rata *f-measure* dan *accuracy* pada tiap nilai k_1 yang diujikan.

6.3.1 Pengujian *Fold* ke-1

Pada *fold* ke-1 data uji yang digunakan adalah dataset pada urutan ke 1-200, sementara data latih menggunakan dataset pada urutan ke 201-1000. Hasil pengujian *fold* ke-1 dari pengujian nilai k_1 dapat dilihat pada Tabel 6.13.

Tabel 6.13 Hasil Pengujian *Fold* ke-1 pada Nilai K_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,929	0,9473	0,905	0,9381
1,3	0,929	0,9473	0,905	0,9381
1,4	0,929	0,9473	0,905	0,9381
1,5	0,929	0,9473	0,905	0,9381
1,6	0,929	0,9473	0,905	0,9381
1,7	0,929	0,9473	0,905	0,9381
1,8	0,929	0,9473	0,905	0,9381
1,9	0,929	0,9473	0,905	0,9381
2,0	0,929	0,9473	0,905	0,9381

6.3.2 Pengujian *Fold* ke-2

Pada *fold* ke-2 data uji yang digunakan adalah dataset pada urutan ke 201-400, sementara data latih menggunakan dataset pada urutan ke 1-200 dan 401-1000. Hasil pengujian *fold* ke-2 dari pengujian nilai k_1 dapat dilihat pada Tabel 6.14.

Tabel 6.14 Hasil Pengujian *Fold* ke-2 pada Nilai K_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,9779	0,9619	0,945	0,9698
1,3	0,9725	0,9619	0,94	0,9672
1,4	0,9722	0,951	0,93	0,9615
1,5	0,9777	0,9565	0,94	0,967
1,6	0,9777	0,9565	0,94	0,967
1,7	0,9777	0,9565	0,94	0,967
1,8	0,9777	0,9565	0,94	0,967
1,9	0,9777	0,9565	0,94	0,967
2,0	0,9776	0,951	0,935	0,9641

6.3.3 Pengujian *Fold* ke-3

Pada *fold* ke-3 data uji yang digunakan adalah dataset pada urutan ke 401-600, sementara data latih menggunakan dataset pada urutan ke 1-400 dan 601-1000. Hasil pengujian *fold* ke-3 dari pengujian nilai k_1 dapat dilihat pada Tabel 6.15.

Tabel 6.15 Hasil Pengujian *Fold* ke-3 pada Nilai K_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,9723	0,9832	0,96	0,9777
1,3	0,9723	0,9832	0,96	0,9777
1,4	0,9723	0,9832	0,96	0,9777
1,5	0,9723	0,9832	0,96	0,9777
1,6	0,9723	0,9832	0,96	0,9777
1,7	0,9723	0,9832	0,96	0,9777
1,8	0,9723	0,9832	0,96	0,9777
1,9	0,9723	0,9832	0,96	0,9777
2,0	0,9723	0,9832	0,96	0,9777

6.3.4 Pengujian *Fold* ke-4

Pada *fold* ke-4 data uji yang digunakan adalah dataset pada urutan ke 601-800, sementara data latih menggunakan dataset pada urutan ke 1-600 dan 801-1000. Hasil pengujian *fold* ke-4 dari pengujian nilai k_1 dapat dilihat pada Tabel 6.16.

Tabel 6.16 Hasil Pengujian *Fold* ke-4 pada Nilai K_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,9319	0,9448	0,91	0,9383
1,3	0,9383	0,9448	0,915	0,9415
1,4	0,9319	0,9448	0,91	0,9383
1,5	0,9319	0,9448	0,91	0,9383
1,6	0,9319	0,9448	0,91	0,9383
1,7	0,9319	0,9448	0,91	0,9383
1,8	0,9315	0,9379	0,905	0,9347
1,9	0,9315	0,9379	0,905	0,9347
2,0	0,9315	0,9379	0,905	0,9347

6.3.5 Pengujian *Fold* ke-5

Pada *fold* ke-5 data uji yang digunakan adalah dataset pada urutan ke 801-1000, sementara data latih menggunakan dataset pada urutan ke 1-800. Hasil pengujian *fold* ke-5 dari pengujian nilai k_1 dapat dilihat pada Tabel 6.17.

Tabel 6.17 Hasil Pengujian *Fold* ke-5 pada Nilai k_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,9436	0,9571	0,93	0,9503
1,3	0,9432	0,95	0,925	0,9466
1,4	0,9366	0,95	0,92	0,9432
1,5	0,93	0,95	0,915	0,9399
1,6	0,93	0,95	0,915	0,9399
1,7	0,9236	0,95	0,91	0,9366
1,8	0,923	0,9428	0,905	0,9328
1,9	0,923	0,9428	0,905	0,9328
2,0	0,9225	0,9357	0,9	0,929

6.3.6 Rata Rata Hasil Pengujian Nilai k_1

Setelah menghitung nilai *f-measure*, *accuracy*, *precision*, dan *recall* pada tiap nilai k_1 menggunakan *5-fold cross validation*, maka selanjutnya nilai tersebut akan dirata – rata. Nilai hasil rata – rata dari setiap pengujian nilai k_1 digunakan untuk mendapatkan nilai k_1 terbaik. Hasil rata – rata pengujian nilai k_1 dapat dilihat pada Tabel 6.18.

Tabel 6.18 Rata – Rata Hasil Pengujian Nilai k_1

Nilai k_1	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
1,2	0,9509	0,9589	0,93	0,9548
1,3	0,9511	0,9574	0,9289	0,9542
1,4	0,9484	0,9553	0,925	0,9518
1,5	0,9482	0,9563	0,9259	0,9522
1,6	0,9482	0,9563	0,9259	0,9522
1,7	0,9469	0,9563	0,925	0,9515
1,8	0,9467	0,9535	0,923	0,95
1,9	0,9467	0,9535	0,923	0,95
2,0	0,9466	0,951	0,921	0,9487

Berdasarkan Tabel 6.18 didapatkan bahwa nilai k_1 terbaik terdapat pada pengujian ketika nilai $k_1 = 1,2$, yang terbukti dengan rata – rata *f-measure* tertinggi yaitu sebesar 0,9548 diikuti dengan nilai rata – rata *accuracy* sebesar 0,93, nilai rata – rata *precision* sebesar 0,9509, nilai rata – rata *recall* sebesar 0,9589. Nilai $k = 20$, nilai eksponen = 2, dan nilai $k_1 = 1,2$ akan digunakan pada pengujian selanjutnya yaitu pengujian pada nilai b pada metode BM25.

6.4 Pengujian Nilai B dengan 5-Fold Cross Validation

Pengujian pada nilai b juga menggunakan *5-fold cross validation*, dimana terbagi menjadi 5 bagian *fold* dengan pembagian 1 *fold* sebagai data uji dan 4 *fold* sebagai data latih. Proses pengujian nilai b pada metode BM25 akan menggunakan nilai k dan eksponen terbaik pada metode NWKNN serta nilai k_1 terbaik pada metode BM25 yang didapatkan berdasarkan pengujian sebelumnya. Setiap nilai b yang diujikan akan dicari nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Nilai b terbaik didapatkan dengan menghitung rata – rata *f-measure* dan *accuracy* pada tiap nilai b yang diujikan.

6.4.1 Pengujian *Fold* ke-1

Pada *fold* ke-1 data uji yang digunakan adalah dataset pada urutan ke 1-200, sementara data latih menggunakan dataset pada urutan ke 201-1000. Hasil pengujian *fold* ke-1 dari pengujian nilai b dapat dilihat pada Tabel 6.19.

Tabel 6.19 Hasil Pengujian *Fold* ke-1 pada Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,929	0,9473	0,905	0,9381
0,55	0,929	0,9473	0,905	0,9381
0,6	0,929	0,9473	0,905	0,9381
0,65	0,9285	0,9407	0,9	0,9346
0,7	0,9285	0,9407	0,9	0,9346
0,75	0,929	0,9473	0,905	0,9381
0,8	0,929	0,9473	0,905	0,9381

6.4.2 Pengujian *Fold* ke-2

Pada *fold* ke-2 data uji yang digunakan adalah dataset pada urutan ke 201-400, sementara data latih menggunakan dataset pada urutan ke 1-200 dan 401-1000. Hasil pengujian *fold* ke-2 dari pengujian nilai b dapat dilihat pada Tabel 6.20.

Tabel 6.20 Hasil Pengujian *Fold* ke-2 pada Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,9779	0,9619	0,945	0,9698
0,55	0,9777	0,9565	0,94	0,967
0,6	0,9775	0,9456	0,93	0,9613
0,65	0,9775	0,9456	0,93	0,9613
0,7	0,9775	0,9456	0,93	0,9613
0,75	0,9776	0,951	0,935	0,9641
0,8	0,9776	0,951	0,935	0,9641

6.4.3 Pengujian *Fold* ke-3

Pada *fold* ke-3 data uji yang digunakan adalah dataset pada urutan ke 401-600, sementara data latih menggunakan dataset pada urutan ke 1-400 dan 601-1000. Hasil pengujian *fold* ke-3 dari pengujian nilai b dapat dilihat pada Tabel 6.21.

Tabel 6.21 Hasil Pengujian *Fold* ke-3 pada Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,9723	0,9832	0,96	0,9777
0,55	0,9723	0,9832	0,96	0,9777
0,6	0,9723	0,9832	0,96	0,9777
0,65	0,9723	0,9832	0,96	0,9777
0,7	0,9723	0,9832	0,96	0,9777
0,75	0,9723	0,9832	0,96	0,9777
0,8	0,9722	0,9776	0,955	0,9749

6.4.4 Pengujian *Fold* ke-4

Pada *fold* ke-4 data uji yang digunakan adalah dataset pada urutan ke 601-800, sementara data latih menggunakan dataset pada urutan ke 1-600 dan 801-1000. Hasil pengujian *fold* ke-4 dari pengujian nilai b dapat dilihat pada Tabel 6.22.

Tabel 6.22 Hasil Pengujian *Fold* ke-4 pada Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,9319	0,9448	0,91	0,9383
0,55	0,9319	0,9448	0,91	0,9383
0,6	0,9315	0,9379	0,905	0,9347
0,65	0,9251	0,9379	0,9	0,9315
0,7	0,9189	0,9379	0,895	0,9283
0,75	0,9189	0,9379	0,895	0,9283
0,8	0,9189	0,9379	0,895	0,9283

6.4.5 Pengujian *Fold* ke-5

Pada *fold* ke-5 data uji yang digunakan adalah dataset pada urutan ke 801-1000, sementara data latih menggunakan dataset pada urutan ke 1-800. Hasil pengujian *fold* ke-5 dari pengujian nilai b dapat dilihat pada Tabel 6.23.

Tabel 6.23 Hasil Pengujian *Fold* ke-5 pada Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,9436	0,9571	0,93	0,9503
0,55	0,93	0,95	0,915	0,9399
0,6	0,9361	0,9428	0,915	0,9395
0,65	0,9361	0,9428	0,915	0,9395
0,7	0,9285	0,9285	0,9	0,9285
0,75	0,929	0,9357	0,905	0,9323
0,8	0,929	0,9357	0,905	0,9323

6.4.6 Rata Rata Hasil Pengujian Nilai B

Setelah menghitung nilai *f-measure*, *accuracy*, *precision*, dan *recall* pada tiap nilai b menggunakan 5-*fold cross validation*, maka selanjutnya nilai tersebut akan dirata – rata. Nilai hasil rata – rata dari setiap pengujian nilai b digunakan untuk mendapatkan nilai b terbaik. Hasil rata – rata pengujian nilai b dapat dilihat pada Tabel 6.24.

Tabel 6.24 Rata – Rata Hasil Pengujian Nilai B

Nilai b	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
0,5	0,9509	0,9589	0,93	0,9548
0,55	0,9482	0,9563	0,9259	0,9522
0,6	0,9493	0,9514	0,923	0,9502
0,65	0,9479	0,95	0,921	0,9489
0,7	0,9451	0,9472	0,917	0,9461
0,75	0,9453	0,9497	0,9189	0,9474
0,8	0,9452	0,9486	0,9179	0,9468

Berdasarkan Tabel 6.24 didapatkan bahwa nilai b terbaik terdapat pada pengujian ketika nilai b = 0,5, yang terbukti dengan rata – rata *f-measure* tertinggi yaitu sebesar 0,9548 diikuti dengan nilai rata – rata *accuracy* sebesar 0,93, nilai rata – rata *precision* sebesar 0,9509, nilai rata – rata *recall* sebesar 0,9589.

6.5 Pengujian Faktor Kesalahan Hasil Klasifikasi

Proses pengujian kata negasi dan sentimen ganda pada faktor kesalahan hasil klasifikasi akan menggunakan nilai k dan eksponen terbaik pada metode NwKNN serta nilai k_1 dan b terbaik pada metode BM25 yang didapatkan berdasarkan pengujian sebelumnya. Pengujian ini akan menggunakan data uji dan data latih pada *fold* ke-3 karena memiliki nilai *confusion* matrix yang lebih baik dibandingkan

fold yang lain dan berfokus pada ulasan data latih ke-95. Penjabaran ulasan ke-95 yang menjadi fokus pengujian ini dapat dilihat pada Tabel 6.25.

Tabel 6.25 Ulasan yang Mendapatkan Hasil Klasifikasi Salah

Ulasan	Negasi	Sentimen Ganda	
		Positif	Negatif
Monumen berada tepat di tengah-tengah simpang lima gumul dan disediakan akses masuk dari tempat parkir melewati terowongan bawah tanah, tapi pada saat saya kesana beberapa fasilitas seperti toilet dan pusat informasi tidak berfungsi padahal musim liburan bukan weekday	Tidak berfungsi	Monumen berada tepat di tengah-tengah simpang lima gumul dan disediakan akses masuk dari tempat parkir melewati terowongan bawah tanah	Beberapa fasilitas seperti toilet dan pusat informasi tidak berfungsi padahal musim liburan bukan weekday

Setiap pengujian faktor kesalahan hasil klasifikasi akan dicari nilai *f-measure*, *accuracy*, *precision*, dan *recall*. Hasil pengujian faktor kesalahan sistem dapat dilihat pada Tabel 6.26.

Tabel 6.26 Hasil Pengujian Faktor Kesalahan Hasil Klasifikasi

		<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F-measure</i>
Pengujian Awal		0,9723	0,9832	0,9600	0,9777
Negasi		0,9723	0,9832	0,9600	0,9777
Sentimen Ganda	Positif	0,9779	0,9833	0,9650	0,9806
	Negatif	0,9723	0,9832	0,9600	0,9777
	Positif & Negatif	0,9725	0,9833	0,9602	0,9779
Negasi dan Sentimen Ganda	Positif	0,9777	0,9832	0,9650	0,9805
	Negatif	0,9777	0,9832	0,9650	0,9805
	Positif & Negatif	0,9779	0,9833	0,9651	0,9806

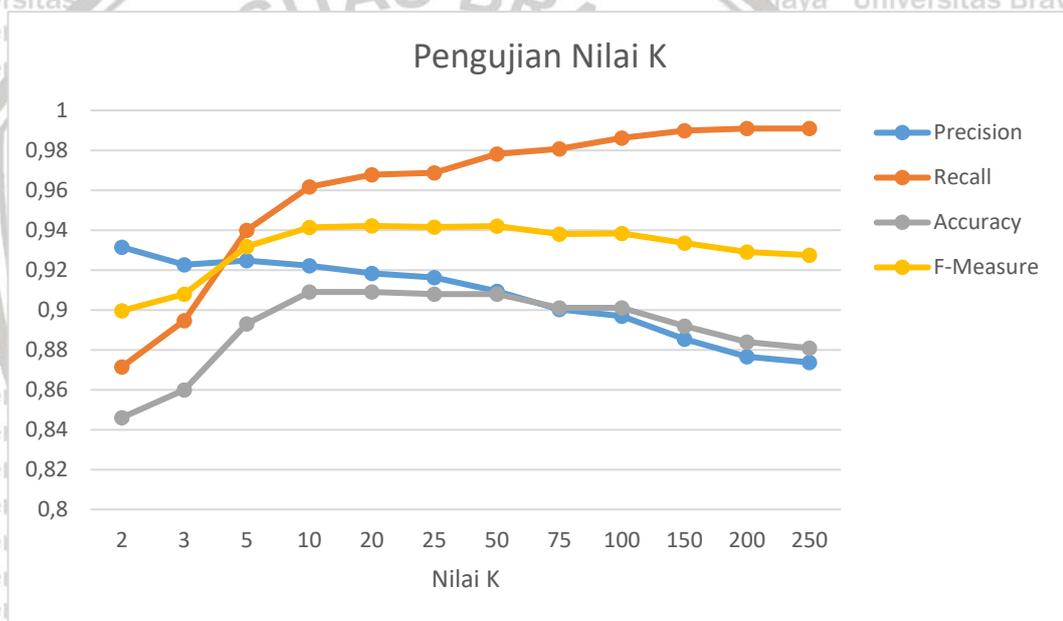
Berdasarkan Tabel 6.26 didapatkan bahwa perbaikan pada ulasan yang memiliki kata negasi dan sentimen ganda dapat menghasilkan nilai *confusion matrix* yang lebih baik serta menandakan bahwa tingkat keberhasilan sistem dalam proses klasifikasi juga ikut meningkat.

6.6 Analisis Pengujian

Berdasarkan hasil dari keseluruhan pengujian yang dilakukan pada parameter k_1 dan b pada BM25 serta parameter k dan eksponen pada NWKNN, didapatkan hasil bahwa nilai terbaik untuk nilai $k_1 = 1,2$, $b = 0,5$, $k = 20$, dan eksponen = 2. Hasil pengujian yang didapatkan dengan menggunakan kombinasi nilai terbaik tersebut adalah rata – rata nilai *precision* sebesar 0,9509, *recall* sebesar 0,9589, *accuracy* sebesar 0,93, dan *f-measure* sebesar 0,9548. Kesalahan dalam proses klasifikasi untuk ulasan pengunjung SLG Kediri disebabkan karena terdapat kata yang mengandung negasi, seperti kata “tidak bersih” yang seharusnya bisa diganti menjadi kata “kotor” serta terdapat ulasan yang mengandung lebih dari 1 sentimen (terdapat pendapat yang bersifat positif dan negatif bersamaan), seperti “Tempat itu sangat sejuk, tetapi banyak sampah berserakan”

6.6.1 Analisis Pengujian Nilai K

Grafik rata – rata hasil evaluasi sistem untuk pengujian nilai k pada keseluruhan *fold* dapat dilihat pada Gambar 6.1.



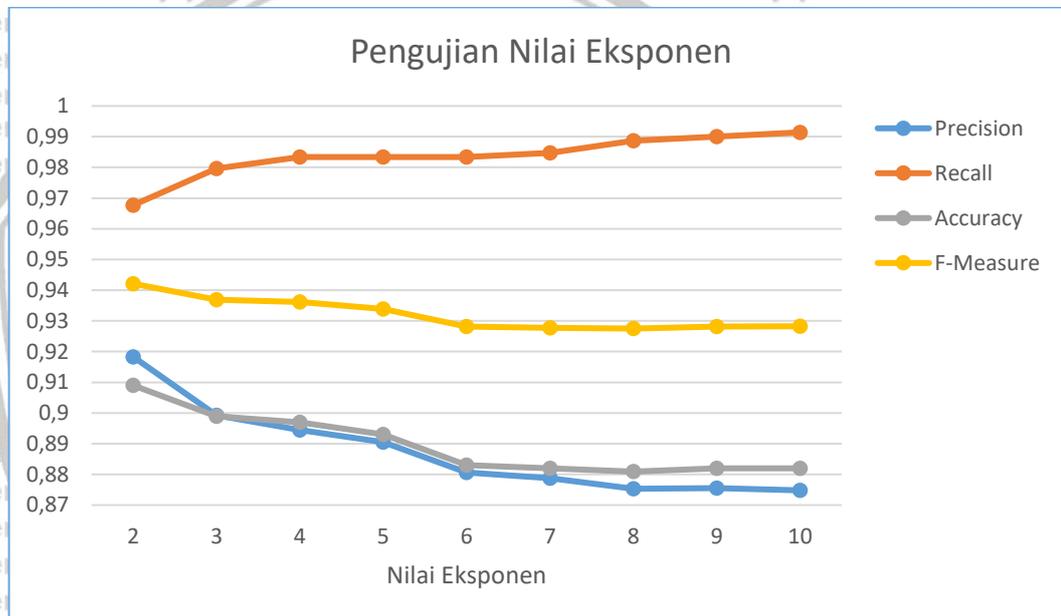
Gambar 6.1 Grafik Rata – Rata Pengujian pada Nilai K

Berdasarkan grafik pada Gambar 6.1 dapat dilihat bahwa pengujian pada nilai k didapatkan hasil berupa nilai k terbaik pada $k=20$ karena memiliki nilai rata – rata *accuracy* dan *f-measure* tertinggi. Hasil evaluasi sistem pada $k=20$ adalah nilai rata – rata *precision* sebesar 0,9183, *recall* sebesar 0,9677, *accuracy* sebesar 0,909, dan *f-measure* sebesar 0,9421. Pada grafik tersebut juga dapat dilihat bahwa semakin besar nilai k , maka nilai *precision* semakin kecil sedangkan nilai *recall* semakin besar. Sementara nilai *accuracy* dan *f-measure* mengalami kenaikan nilai secara bertahap ketika $k=2$ hingga $k=20$, namun setelah lebih dari 20 mulai terjadi penurunan secara terus menerus. Hasil tersebut disebabkan karena data yang digunakan pada penelitian ini tidak seimbang, dimana kelas positif lebih dominan

dibandingkan kelas negatif sehingga penggunaan nilai k yang besar dapat menarik jumlah data latih dengan kelas positif yang sangat banyak. Kemunculan data latih dengan kelas positif yang banyak dapat menyebabkan nilai akhir untuk kelas positif lebih tinggi dibandingkan kelas negatif, sehingga hasil klasifikasi untuk data uji yang seharusnya negatif bisa mendapatkan hasil klasifikasi positif karena banyaknya data kelas positif yang ikut dalam proses perhitungan. Kesalahan klasifikasi pada data uji yang seharusnya bernilai negatif tersebut mengakibatkan menurunnya nilai *precision*, *accuracy*, dan *f-measure* sementara nilai *recall* akan semakin naik karena jumlah kesalahan hasil klasifikasi pada data yang seharusnya positif semakin kecil.

6.6.2 Analisis Pengujian Nilai Eksponen

Grafik rata – rata hasil evaluasi sistem untuk pengujian nilai eksponen pada keseluruhan *fold* dapat dilihat pada Gambar 6.2.



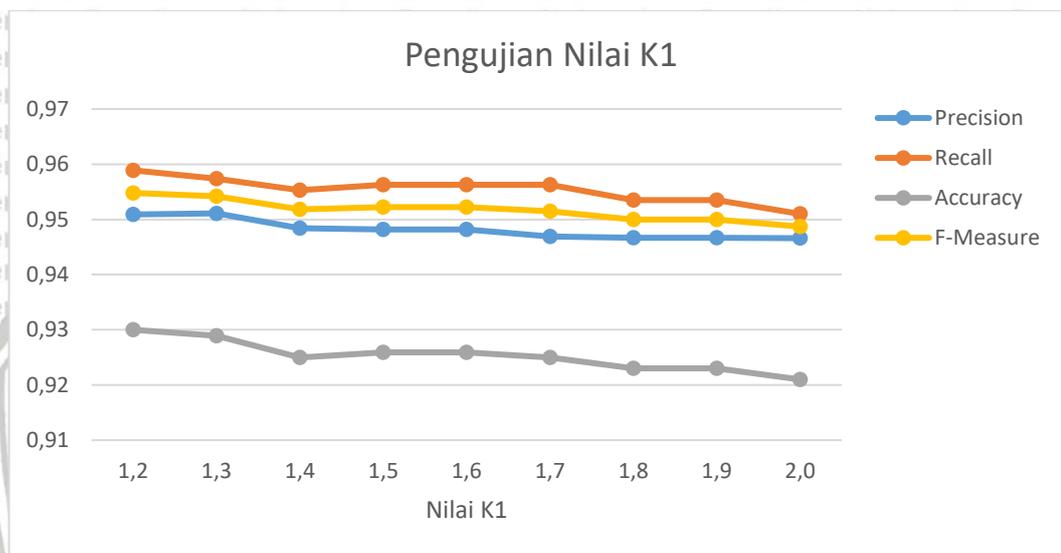
Gambar 6.2 Grafik Rata – Rata Pengujian pada Nilai Eksponen

Berdasarkan grafik pada Gambar 6.2 dapat dilihat bahwa pengujian pada nilai eksponen didapatkan hasil berupa nilai eksponen terbaik ketika bernilai 2, karena memiliki nilai rata – rata *accuracy* dan *f-measure* tertinggi. Hasil evaluasi sistem pada eksponen=2 adalah nilai rata – rata *precision* sebesar 0,9183, *recall* sebesar 0,9677, *accuracy* sebesar 0,909, dan *f-measure* sebesar 0,9421. Pada grafik tersebut dapat dilihat bahwa semakin tinggi nilai eksponen, maka nilai *precision*, *accuracy* dan *f-measure* akan semakin menurun. Penurunan tersebut terjadi karena semakin tinggi nilai eksponen, maka selisih nilai bobot untuk tiap kategori kelas pada metode NWKNN semakin kecil. Selisih nilai bobot yang semakin kecil mengakibatkan terjadinya penurunan kemampuan sistem untuk melakukan proses klasifikasi. Alasan terjadinya penurunan kemampuan sistem karena nilai akhir dari metode NWKNN untuk kelas positif (kelas yang dominan) akan cenderung lebih tinggi dibandingkan dengan kelas negatif, sehingga kemungkinan

sistem untuk salah dalam memberikan hasil klasifikasi untuk kelas yang seharusnya negatif semakin tinggi. Hal tersebut mengakibatkan terjadinya penurunan pada nilai *precision*, *accuracy*, dan *f-measure*, sementara nilai *recall* semakin tinggi karena kesalahan untuk hasil klasifikasi kelas yang seharusnya positif semakin kecil.

6.6.3 Analisis Pengujian Nilai K1

Grafik rata – rata hasil evaluasi sistem untuk pengujian nilai k_1 pada keseluruhan *fold* dapat dilihat pada Gambar 6.3.

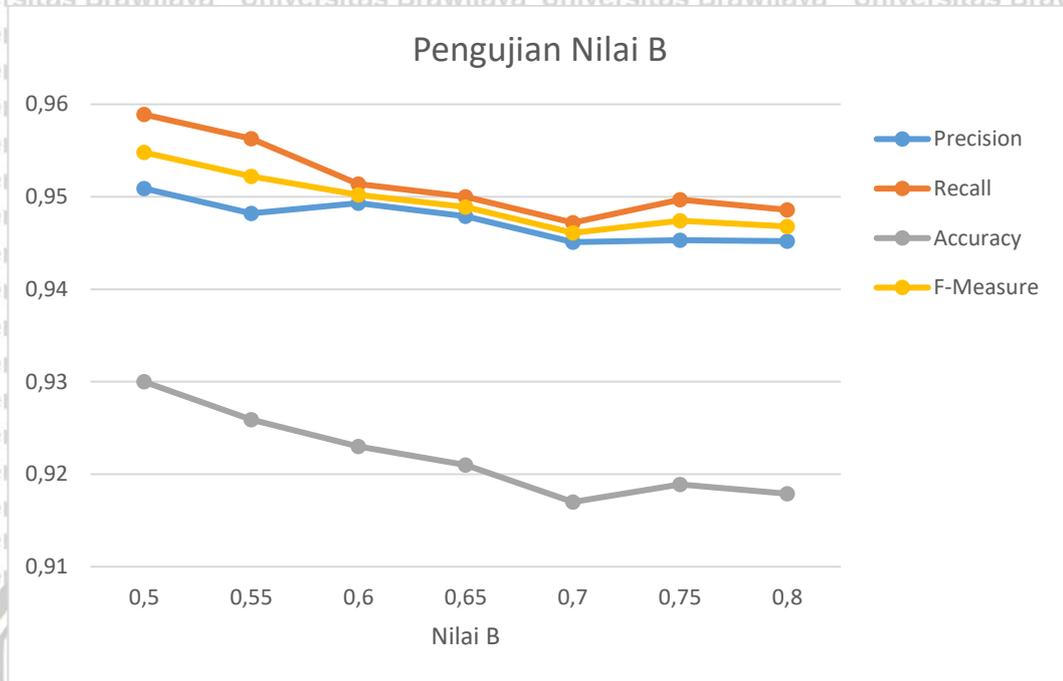


Gambar 6.3 Grafik Rata – Rata Pengujian pada Nilai K1

Berdasarkan grafik pada Gambar 6.3 dapat dilihat bahwa pengujian pada nilai k_1 didapatkan hasil berupa nilai k_1 terbaik ketika bernilai 1,2. Nilai $k_1=1,2$ merupakan nilai terbaik karena memiliki nilai rata – rata *accuracy* dan *f-measure* tertinggi. Hasil evaluasi sistem pada $k_1=1,2$ adalah nilai rata – rata *precision* sebesar 0,9509, *recall* sebesar 0,9589, *accuracy* sebesar 0,93, dan *f-measure* sebesar 0,9548. Pada grafik tersebut dapat dilihat bahwa semakin tinggi nilai k_1 , maka keseluruhan nilai hasil evaluasi semakin menurun. Meskipun terjadi penurunan nilai pada hasil evaluasi sistem, namun tingkat penurunan tersebut sangat kecil. Terbukti bahwa pada pengujian $k_1=2$ yang merupakan angka pengujian terbesar mendapatkan nilai rata – rata *precision* sebesar 0,9466, *recall* sebesar 0,9510, *accuracy* sebesar 0,921, dan *f-measure* sebesar 0,9487. Hasil nilai evaluasi sistem antara $k_1=1,2$ dengan $k_1=2$ hanya berbeda kurang dari 0,01 untuk tiap rata – rata nilai evaluasi. Secara garis besar bisa disimpulkan bahwa besarnya nilai k_1 pada metode BM25 tidak memberikan pengaruh yang signifikan terhadap hasil evaluasi karena memiliki hasil akhir yang relatif sama.

6.6.4 Analisis Pengujian Nilai B

Grafik rata – rata hasil evaluasi sistem untuk pengujian nilai b pada keseluruhan *fold* dapat dilihat pada Gambar 6.4.



Gambar 6.4 Grafik Rata – Rata Pengujian pada Nilai B

Berdasarkan grafik pada Gambar 6.4 dapat dilihat bahwa pengujian pada nilai b didapatkan hasil berupa nilai b terbaik ketika bernilai 0,5. Nilai $b=0,5$ merupakan nilai terbaik karena memiliki nilai rata – rata *accuracy* dan *f-measure* tertinggi. Hasil evaluasi sistem pada $b=0,5$ adalah nilai rata – rata *precision* sebesar 0,9509, *recall* sebesar 0,9589, *accuracy* sebesar 0,93, dan *f-measure* sebesar 0,9548. Pada grafik tersebut dapat dilihat bahwa semakin tinggi nilai b, maka keseluruhan nilai hasil evaluasi memiliki kecenderungan semakin menurun. Meskipun terjadi penurunan nilai pada hasil evaluasi sistem, namun pada beberapa pengujian nilai b terjadi kenaikan nilai evaluasi meskipun kenaikan yang terjadi tidak terlalu besar. Apabila pengujian nilai b terbaik adalah 0,5, maka pengujian nilai b dengan hasil evaluasi sistem paling kecil adalah pada $b=0,8$. Pada $b=0,8$ nilai rata – rata *precision* sebesar 0,9452, *recall* sebesar 0,9486, *accuracy* sebesar 0,9179, dan *f-measure* sebesar 0,9468. Dari hasil evaluasi sistem antara $b=0,5$ dan $b=0,8$ dapat dilihat bahwa perbedaan nilainya sangat kecil, yaitu kurang dari 0,015 untuk tiap rata – rata nilai evaluasi. Oleh karena itu, bisa disimpulkan bahwa besarnya nilai b pada metode BM25 juga tidak memberikan pengaruh yang signifikan terhadap hasil evaluasi karena memiliki hasil akhir yang relatif sama.

Meskipun demikian penggunaan metode BM25 bersamaan dengan metode klasifikasi NWKNN menghasilkan nilai evaluasi sistem yang lebih baik dibandingkan hanya menggunakan metode NWKNN. Hal tersebut bisa dilihat ketika hanya menggunakan metode klasifikasi NWKNN dan menggunakan pembobotan TF-IDF, nilai *accuracy* dan *f-measure* yang didapatkan dengan nilai k

dan eksponen terbaik adalah sebesar 0,909 dan 0,9421. Sementara nilai *accuracy* dan *f-measure* yang didapatkan dengan nilai k dan eksponen terbaik pada metode NWKNN dan menggunakan nilai k_1 dan b terbaik pada metode BM25 adalah sebesar 0,93 dan 0,9548.

6.6.5 Analisa Pengujian Faktor Kesalahan Hasil Klasifikasi

Berdasarkan pengujian yang telah dilakukan didapatkan nilai *f-measure*, *accuracy*, *precision*, dan *recall* seperti pada pada Tabel 6.26. Tabel tersebut memperlihatkan bahwa hasil *confusion matrix* antara pengujian awal dengan perbaikan kata negasi terhadap ulasan pada Tabel 6.25 menghasilkan nilai yang sama, meskipun demikian terjadi perbedaan pada nilai akhir untuk kelas positif dan negatif pada saat proses klasifikasi. Pada perbaikan kata negasi, kelas positif bernilai = 40,6271 dan kelas negatif bernilai 29,8342, sementara pada pengujian awal, kelas positif bernilai = 40,3139 dan kelas negatif bernilai 23,8465. Sementara pada perbaikan sentimen ganda dapat dilihat bahwa nilai *f-measure*, *accuracy*, *precision*, dan *recall* sedikit lebih tinggi dibandingkan pada pengujian awal. Hal ini disebabkan karena sentimen positif berhasil diklasifikasikan sebagai kelas positif karena nilai akhir kelas positif lebih tinggi 10 poin dibandingkan nilai akhir kelas negatif. Sementara pada sentimen negatif masih mendapatkan klasifikasi yang salah karena nilai akhir kelas negatif lebih rendah 5 poin dibandingkan nilai akhir kelas positif. Nilai evaluasi sistem tertinggi diperoleh ketika perbaikan kata negasi serta sentimen ganda diterapkan bersamaan, 2 faktor tersebut yang apabila terpisah hanya memiliki pengaruh kecil dalam proses klasifikasi namun menjadi faktor yang sangat berpengaruh ketika diterapkan bersamaan. Hal tersebut terbukti dari ulasan pada Tabel 6.25 yang dipecah menjadi 2 ulasan berhasil diklasifikasikan secara benar sehingga membuat nilai *f-measure*, *accuracy*, *precision*, dan *recall* meningkat. Hasil tersebut membuktikan bahwa kata negasi dan sentimen ganda merupakan faktor dalam terjadinya kesalahan hasil klasifikasi, maka dari itu diperlukan perbaikan pada tiap ulasan agar hasil klasifikasi dapat lebih baik lagi.

BAB 7 PENUTUP

Bab ini berisikan penjabaran terkait kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang dapat digunakan untuk penelitian selanjutnya.

7.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat ditarik kesimpulan dari sistem analisis sentimen pada ulasan pengunjung Simpang Lima Gumul Kediri sebagai berikut.

1. Pengujian yang dilakukan pada sistem analisis sentimen ulasan pengunjung Simpang Lima Gumul Kediri menggunakan metode BM25 dan NWKNN mendapatkan hasil terbaik berupa rata – rata nilai *precision* = 0,9509, *recall* = 0,9589, *accuracy* = 0,93, dan *f-measure* = 0,9548. Hasil tersebut didapatkan menggunakan nilai parameter optimal, yaitu $k_1 = 1,2$, $b = 0,5$, $k = 20$, dan eksponen = 2. Selain itu hasil pengujian juga menunjukkan bahwa penggunaan metode BM25 menggantikan proses pembobotan dan pemeringkatan TF-IDF pada NWKNN menghasilkan nilai evaluasi sistem yang lebih baik dibandingkan menggunakan TF-IDF.
2. Parameter nilai k_1 dan b pada metode BM25 yang semakin besar menyebabkan penurunan nilai akurasi dan *f-measure*. Meskipun demikian hal tersebut tidak memberikan pengaruh yang signifikan pada kinerja sistem karena tiap pengujian yang dilakukan hanya memiliki perbedaan nilai yang sangat kecil.
3. Parameter nilai k yang terlalu kecil atau terlalu besar serta nilai eksponen yang terlalu besar menyebabkan penurunan nilai akurasi dan *f-measure*. Kedua parameter tersebut memiliki pengaruh yang signifikan terhadap kinerja sistem karena menjadi penentu dalam hasil proses klasifikasi, sehingga setiap pengujian memiliki selisih nilai yang besar.

7.2 Saran

Saran yang dapat diberikan dari penelitian ini untuk dikembangkan pada penelitian berikutnya adalah sebagai berikut.

1. Diperlukan proses perbaikan kata yang mengandung negasi sebelum melakukan *preprocessing*, sehingga makna dari kata bernegasi dalam ulasan tidak hilang ketika proses *preprocessing* dilakukan. Sebagai contoh apabila terdapat kalimat “Tempat ini tidak bersih”, kata “tidak bersih” dalam kalimat tersebut dapat diganti menjadi kata “kotor”.
2. Memisahkan ulasan yang memiliki sentimen ganda (terdapat kalimat dan negatif dalam satu ulasan). Hal ini dilakukan untuk meminimalisir tingkat kesalahan sistem dalam memberikan hasil klasifikasi pada tiap ulasan.
3. Proses pelabelan pada setiap ulasan seharusnya dilakukan oleh seorang pakar bahasa.

DAFTAR REFERENSI

Averiani, 2019. *Simpang Lima Gumul, L'Arc de Triomphe ala Kediri*. [Online] Available at: <https://ntdindonesia.com/wisata/simpang-lima-gumul-larc-de-triomphe-ala-kediri/> [Accessed 2 Maret 2021].

Desta, E. Y. & Gurmessa, T. T., 2019. Analysis and Result of Classification Algorithm on Email Classification. *International Journal of Computer Engineering Research*, VIII(1), pp. 1-9.

Haq, F. U. & Rachmat, H., 2020. Penggunaan Google Review sebagai Penilaian Kepuasan Pengunjung dalam Pariwisata. *Journal of Sustainable Tourism Research*, II(1), pp. 10-12.

Harjanto, D. S., Endah, S. N. & Bahtiar, N., 2012. Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF). *Jurnal Sains dan Matematika*, XX(3), pp. 64-70.

Indriati & Ridok, A., 2016. Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). *Journal of Environmental Engineering & Sustainable Technology*, III(1), pp. 23-32.

Kusuma, L., 2019. *5 Fakta Simpang Lima Gumul yang Tidak Banyak Diketahui*. [Online] Available at: <https://agtvnews.com/2019/08/5-fakta-simpang-lima-gumul-kediri.html> [Accessed 28 November 2020].

Martha, M, V. C., Naga, D. S. & Rompas, P., 2018. Perbandingan Pengklasifikasi k-Nearest Neighbor dan Neighbor-Weighted k-Nearest Neighbor Pada Sistem Analisis Sentimen dengan Data Microblog. *Jurnal Sains dan Teknologi, Universitas Negeri Manado*, I(1), pp. 81-90.

Mubarok, I., 2018. *Cara Daftar dan Verifikasi Google Bisnisku*. [Online] Available at: <https://www.niagahoster.co.id/blog/google-bisnisku/> [Accessed 30 November 2020].

Mujtaba, H., 2020. *What is Cross Validation in Machine learning? Types of Cross Validation*. [Online] Available at: <https://www.mygreatlearning.com/blog/cross-validation/> [Accessed 4 Maret 2021].

Novantiran, A., Sabariah, M. K. & Effendy, V., 2015. Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *e-Proceeding of Engineering*, II(1), pp. 1177-1182.

Pardede, J., Husada, M. G. & Riansyah, R., 2018. Implementasi dan Perbandingan Metode Okapi BM25 dan PLSA pada Aplikasi Information Retrieval. pp. 1-10.

Pradhana, M. O., Indriati & Adinugroho, S., 2020. Analisis Sentimen Masyarakat terhadap Uji Coba LRT Jakarta Menggunakan Improved K-Nearest Neighbor

dan Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, IV(6), pp. 1888-1896.

Pristiyanti, R. I., Fauzi, M. A. & Muflikhah, L., 2018. Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor. *International Journal of Engineering & Technology*, Volume VII, pp. 1499-1501.

Putri, F. O., Indriati & Wihandika, R. C., 2020. Analisis Sentimen pada Ulasan Pengguna MRT Jakarta Menggunakan Metode Neighbor-Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, VI(7), pp. 2195-2203.

Robertson, S. & Zaragoza, H., 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends® in Information Retrieval*, III(4), p. 333–389.

Sahara, S. & Wahyudi, M., 2015. K-Nearest Neighbors Sebagai Analisis Sentiment Review Produk Appstore For Android. *Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIT)*, pp. 77-82.

Septrinas, E., I. & Soebroto, A. A., 2019. Klasifikasi Berita Olahraga Berbahasa Indonesia menggunakan Metode BM25 dan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, III(10), pp. 9762-9769.

Talib, R., Hanif, M. K., Ayesha, S. & Fatima, F., 2016. Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications (IJACSA)*, VII(11), pp. 414-418.

Tan, S., 2005. Neighbor-weighted K-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications: An International Journal*, XXVIII(4), pp. 667-671.

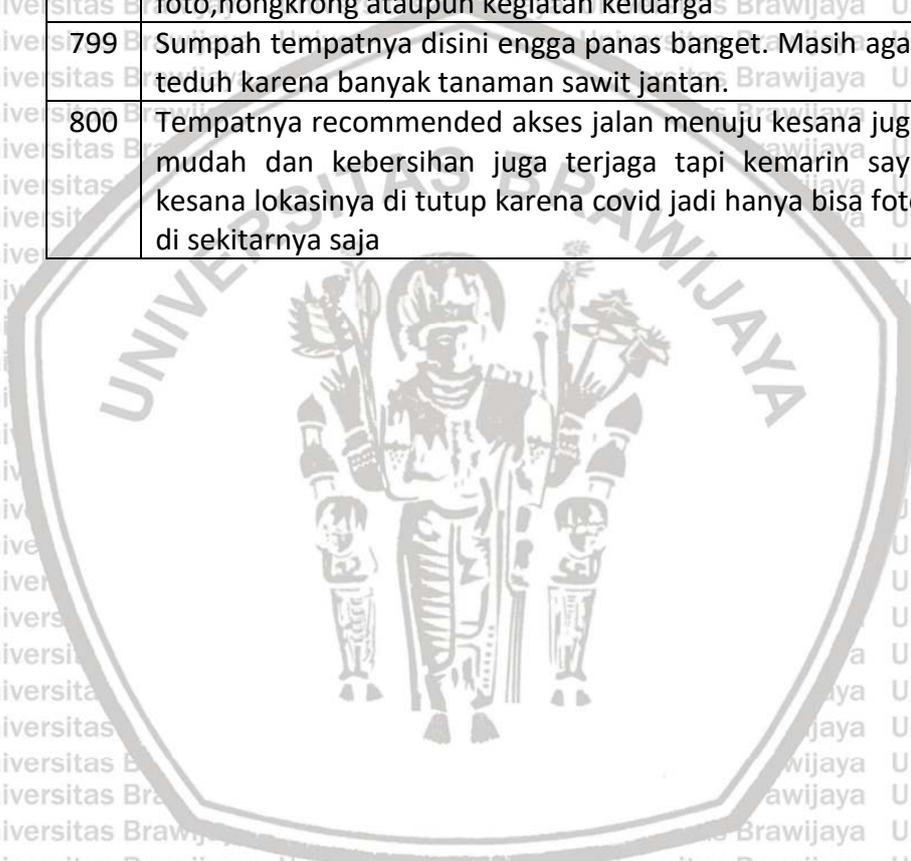
Yang, C.-Z., Du, H.-H., Wu, S.-S. & Chen, I.-X., 2012. Duplication Detection for Software Bug Reports based on BM25 Term Weighting. *Conference on Technologies and Applications of Artificial Intelligence*, pp. 33-38.

LAMPIRAN A DATA

No.	Ulasan	Kelas Aktual
1	Tempat keren ..kediri punya	Positif
2	Bersih, terawat. Pandemi masih tutup.	Positif
3	Indah, bersih ... Tempat instagramable	Positif
4	Enak buat nyantai	Positif
5	Bagus kalo malem hari	Positif
6	Udara bersih tempat nya juga bersih banget, kalau mau sepi hindari hari libur.	Positif
7	Suasana senja ketika berada disini sangat bagus	Positif
8	Kalo ke kediri wajib mampir di simpang lima gumul	Positif
9	Tempat yang indah untuk gowes	Positif
10	Bagus untuk bermain sekeluarga	Positif
11	Pertama kali main di tulungagung, menyempatkan maen ke simpang lima gumul kediri. Pengalaman pertama kali datang ke kediri dan mampir hanya ke simpang 5 gumul. Cukup amaze dengan simpang 5 gumul, karena desain dan tata letaknya yang bagus sekali menurut saya. Disana sangat ramah sekali untuk pejalan kaki dan diberikan tempat parkir yang luas dan bagus. Mungkin untuk teman teman yang belum pernah dtg ke simpang 5 gumul ini, simpang 5 gumul punya beberapa pintu masuk lewat jalur bawah tanah. Jadi buat temen temen yang bawa kendaraan bisa parkir di tempat parkir disetiap sisi yang sudah disediakan	Positif
12	Jalanan disana cukup besar dan ramai, jadi sangat nyaman untuk pejalan kaki ketika melewati jalan bawah tanah	Positif
13	Bagus si apa lagi buat ngopi ngopi santai	Positif
14	Bersih dan tertata	Positif
15	Tempat nya memposana	Positif
16	Tempat favorit keluarga	Positif
17	Tempatnya enak buat foto2.. parkirnya di sebrang jalan gumul	Positif

No.	Ulasan	Kelas Aktual
18	Nyaman damai seperti yang dulu saat pertama kali mengunjungi nya	Positif
19	Nyaman dan pas buat ikon..... Bisa mengangkat citra Kediri jadi lebih tinggi	Positif
20	Salah satu ikon Kab. kediri. Tempat yang cocok murah meriah untuk rekreasi dan refreshing cari angin kencang.	Positif
...		
775	Walaupun gerimis tetap aman dong kesana karena lewat terowongan, dan di sepanjang terowongan banyak foto-foto yang menunjukkan keindahan serta sejarah Kediri.	Positif
776	Ikon kota kediri, bagus dan indah	Positif
777	Tempat yg bagus ga pernah ada bosennya kesini	Positif
778	IKON kediri, bagi yg masih baru di kediri dan ngga sering main bisa bingung. Kl pas ga rame cocok buat jogging. Ada tamannya juga, rindang. Parkiran luas, ada parkir 1 dan 2,	Positif
779	Dengan adanya pembatasan pengunjung menjadikan suasana lebih asri	Positif
780	Lokasinya mudah dicari, dan bangunanya bagus. Ada parkiran disekitar persimpangan untuk motor dan mobil. Bisa masuk ke lokasi dan jalan" disekitaran lokasi itu, buat foto" epic banget!	Positif
781	Pusat Nongkrong Oke, Fotografi Ok	Positif
782	Tempat ikonik di Kediri. Terlihat bagus dan megah. Sayang, jauh-jauh dari Sby pas ke sini gak bisa foto-foto langsung on the spot. Mungkin sdg ditutup selama pandemi.	Positif
783	Tempat yg menarik untuk jogging pagi hari	Positif
784	Senang bisa gowes di sekitar SLG	Positif
785	Lumayan lah,bisa bawa keluarga jalan" menikmati suasana segar	Positif
786	Bagus bgt tempatnya, kesini sama temen dan solo having fun	Positif
787	Tempat yang nyaman	Positif
788	Simbol kediri yg anggun	Positif
789	Tempat yang sangat cocok untuk bersantai	Positif
790	Sangat estetik terasa seperti di manca negara	Positif
791	Tempat yang cantik dan tertata rapi.sangat di anjurkan buat libutan pas weekend	Positif
792	Tempat iconik di Kediri, luas dan banyak tempat instagramable	Positif

No.	Ulasan	Kelas Aktual
793	Sementara pandemi masih dilarang masuk.Hny sempat berfoto beberapa saat saja , tempat yg keren	Positif
794	Memang sangatlah istimewa dan jd ikonik kota Kediri dgn Gudang Garamnya... Bagus buat spot foto	Positif
795	Tempat yang nyaman dan sangat bagus untuk bersantai bersama keluarga atau teman	Positif
796	Bersih Aman Indah	Positif
797	Bagus, bersih, nyaman, aman	Positif
798	Tempat yang asik buat semua kegiatan, buat hunting foto,nongkrong ataupun kegiatan keluarga	Positif
799	Sumpah tempatnya disini engga panas banget. Masih agak teduh karena banyak tanaman sawit jantan.	Positif
800	Tempatnya recommended akses jalan menuju kesana juga mudah dan kebersihan juga terjaga tapi kemarin saya kesana lokasinya di tutup karena covid jadi hanya bisa foto di sekitarnya saja	Positif



LAMPIRAN B HASIL KLASIFIKASI SISTEM

No.	Ulasan	Kelas Aktual	Kelas Sistem
1	Sangat direkomendasikan apabila untuk liburan keluarga terutama jika ada anak kecil, maka berkeliling area slg terutama bermain di taman akan sangat menyenangkan untuk anak	Positif	Positif
2	Tempat nya bagus buat nyantai di sore atau pagi hari, lumayan sepi di lalui kendaraan besar, recomended buat hunting karena lumayan banyak spot	Positif	Positif
3	Tempat yang cocok apabila punya anak yang aktif berlarian karena disini area yg ramah anak	Positif	Positif
4	Lokasi sangat strategis, Mudah dijangkau, dan lahan parkir yg luas di seputaran simpang lima gumul.	Positif	Positif
5	Parkir murah, 2000 rupiah saja.. Klo mau k tugu simpang lima gumul.	Positif	Positif
6	Tempat parkir resmi, bukan kaleng-kaleng jadi merasa aman dan tidak perlu khawatir	Positif	Positif
7	Tempat yang bagus untuk menyegarkan diri	Positif	Positif
8	Bersih, fasilitas lengkap parkir luas disemua arah masuk Simpang Lima	Positif	Positif
9	Mantap, bagi kami yg dari lampung tengah, udara pagi cukup dingin. Tapi all semuanya mantap pak.InsyAllah klo ke Kediri akan kami sempatkan ke simpang 5.	Positif	Positif
10	Tempat yang sangat bagus di Kediri walaupun di luar kota tapi cukup dekat dari pusat kota Kediri.	Positif	Positif
11	Jika Anda suka berjalan atau jogging di sore atau malam hari, tempat ini akan memberi Anda banyak kesenangan.	Positif	Positif
12	Persimpangan yg bagus buat foto2 di malam hari	Positif	Positif
13	Keren banget cocok buat wisata keluarga	Positif	Positif
14	Dari sekian banyak waktu ketika saya berkunjung kesini, favorit saya adalah ketika	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
	pagi hari untuk jogging dan malam hari untuk menikmati udara malam. Suasana nya sangat mendukung untuk dapat dinikmati dengan nikmat		
15	Tempat menarik buat liburan keluarga dan belanja cemilan apalagi ketika cfd	Positif	Positif
16	Bagus nih disini.. kalau kamu ke Kediri sempatin kesini. Jangan lupa berswa foto juga disini. Kurang mantap kalau ke Kediri nggak berfoto di icon nya kota ini. Oiya, jalan dikit ada pasar malamnya lho, jual makanan2 kaki lima yang lumayan2 enak buat yg pingin aman dikantong	Positif	Positif
17	Tempatnya bagus skali buat photo2, area parkir kendaraannya juga cukup luas.... Pengunjung bs datang kesini kapan saja, dan banyak terdapat makanan khas dr jawa timurnya...	Positif	Positif
18	Spot photo bagus dari segala sudut walaupun sama sekali tidak mencirikan kondisi budaya kabupaten Kediri	Positif	Positif
19	Penikmat senja wajib kesini kalau mau menikmati indahnya suasanya yang menyenangkan di kediri	Positif	Positif
20	Salah satu sudut yang indah di Kediri. Cocok buat yang suka selfi	Positif	Positif
21	Tempat yang bagus untuk berkumpul dengan teman dan keluarga	Positif	Positif
22	Bagus banget.. bersih pula	Positif	Positif
23	Santai sejenak menikmati pemandangan disini memang paling nikmat	Positif	Positif
24	Berulang kali kesini tidak pernah merasa bosan dengan suasana nya	Positif	Positif
25	Bagus buat foto, lumayanlah kalo mau hangout kesini aja.	Positif	Positif
26	Suasana mantap, bagus kalo malem sih lebih aestethic buat foto.	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
27	1 kata buat Monumen Gumul the best walaupun jauh pas liat Monumen ini yg berdiri kokoh di tengah kota membuat rasa lelah ini hilang seketika dgn keistimewaan dari Monumen tersebut	Positif	Positif
28	Keindahan yang luar biasa pada malam hari selalu membuat saya terkesima dengan bangunan ini	Positif	Positif
29	Wah bagus banget udah serasa di perancis hehe	Positif	Positif
30	Tempat bersejarah yang wajib dikunjungi kapanpun kita berada di Kediri, hukumnya wajib untuk merasakan keindahan monumen ini kalau di Kediri	Positif	Positif
31	Salah satu monumen di kediri yang mirip salah satu kota Paris. Tempat yang bagus, terutama masuk gratis.	Positif	Positif
32	Kalau mau kesini saran saya harus datang malam atau malam, biar tidak panas. Selain itu, saya memikirkan pemandangan yang lebih baik di sore atau malam hari.	Positif	Positif
33	Banyak hal yang ada di area monumen ini, ada taman, tempat bermain, tempat makan, dll jadi kalau kesini pasti betah deh untuk menjelajah setiap sudut yang ada	Positif	Positif
34	Monumen SLG sangat indah..	Positif	Positif
35	Tempatnya bersih, indah...	Positif	Positif
36	Tidak sia-sia jauh-jauh dari Gunung Kawi ke Simpang Lima Gumul satu keluarga. Pemandangannya indah kalau malam hari. Pemandangan di sana bagus-bagus buat hunting atau foto-foto.	Positif	Positif
37	Indah di sore hari	Positif	Positif
38	Area publik yang anginnya semilir sejuk nikmat	Positif	Positif
39	Pemandangan yang indah selalu membuat mata terasa nyaman melihat nya	Positif	Positif
40	Tempat yang fantastis	Positif	Negatif

No.	Ulasan	Kelas Aktual	Kelas Sistem
41	Tempat wisata keluarga yg luas dan memiliki panorama indah serta udara yg bersih.	Positif	Positif
42	Bagus tempatnya. suasana kota sekali. mau foto pagi siang atau malem tetep aja bagus	Positif	Positif
43	Pokok nya bagus ini tempatnya, salam dari Malang yang rela kesini karena pengen menikmati suasana nya	Positif	Positif
44	Salah satu maskotnya kediri yg tetap terjaga keasriannya hingga saat ini Klo ke kota kediri wajib mampir kesini biarpun hanya sekedar selfi sebentar	Positif	Positif
45	Simpang lima gumul tempat yg sangat cocok untuk represing keluarga, saya anjurkan yg belum pernah kesana silahkan datang kesana.	Positif	Positif
46	Untuk segi jalanan sangat bagus dan sangat diperhatikan membuat pengendara yg lewat sana merasa nyaman	Positif	Negatif
47	Suasananya nyaman di sini, enak buat wisata kota, parkiran luas.	Positif	Positif
48	Cantik dan indah buat berfoto	Positif	Positif
49	Ngopi disini enak	Positif	Positif
50	Jika kalian ke Kediri, jangan lupa sempatkan untuk singgah di simpang lima gumul Kediri. Bangunan yang berada di tengah-tengah kota kediri ini sangat menarik perhatian siapa saja yang melewatinya. Bentuk bangunannya yang menyerupai Arc De Triomphe di Paris, Francis lho.	Positif	Positif
51	Spot foto yg keren, berasa di paris. Ada rumput yg bisa buat tempat duduk, enak kalau cuacanya gak panas	Positif	Positif
52	Yang kayak tempat duduk warna putih di sekitar monumen banyak jejak kaki yang jadi nuansanya kotor	Negatif	Negatif
53	Ada beberapa fandalisme di tempat parkir, sama tidak terawat temboknya jadi terkesan jorok	Negatif	Negatif

No.	Ulasan	Kelas Aktual	Kelas Sistem
54	Alhamdulillah pas kesana pas mendung, jadi adem. Tempatnya enak, lebih enak buat foto2 aja si. Kalo mau jajan kudu turun ke pasar nya. Kalo kesana pas weekend, beneran deh, super banyak yg Dateng, super banyak jual, kayanya tu apa aja ada disana.	Positif	Positif
55	Perawatan dan fasilitas TDK dioptimalkan, bangunan banyak yang terabaikan, banyak ruang kosong dibawah tanah dan minim petunjuk arah. Utk ikon kota sebaiknya di maksimalkan pengelolaanya jd bisa terjaga dan terawat dgn baik.	Negatif	Negatif
56	Cocok buat main bareng teman, keluarga. Cuma bayar parkir. Beloknya kadang bikin beberapa orang bingung mau belok ke mana. Buat olahraga juga cocok	Positif	Positif
57	Kebanggaan orang kediri harta berharga banget buat kita orang kediri... Masuk gratis cuma bayar parkir 2000... Laper ada pasar kuliner nya.. Setiap minggu ada car free day nya... Tempat luas... Angin ew sepoi" duduk di rumput bikin betah gak mau pulang... Bersih kok...	Positif	Positif
58	Petunjuk tempat parkir susah terlihat jadi bikin bingung orang yg baru kesini	Negatif	Negatif
59	Bagus, sayang ditutup akibat pandemi.	Negatif	Positif
60	Nyaman dan strategis	Positif	Positif
61	Disini juga ada taman yang lumayan luas dan penjual, baik makanan maupun oleh-oleh khas kediri meskipun perlu jalan agak jauh	Positif	Positif
62	Wifi corner di area taman SLG kenapa gak ada yang jaga ya? Mau wifi an disana bingung beli kupon dimana	Negatif	Negatif
63	Harga makanan dan minuman di area bawah murah meriah rek, dijamin kenyang dan hemat makan disini mah	Positif	Positif
64	Bangunan yg bagus buat berfoto..	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
65	Tempatnya yang sejuk dan indah menambah suasana lebih tenang begitu berkesan kota kediri dan bagus untuk foto bersama	Positif	Positif
66	Salah satu tempat teramai di Jawa Timur yang bernuansa seperti negeri sebelah, Tidak ada tiket masuk, parkir mobil hanya Rp. 5000 saja, Tempat yang nyaman, sejuk, jika kesini pada saat malam hari	Positif	Positif
67	Bagi saya, tempat ini hanyalah kurang perawatan pada lampu hiasan dan kebersihan pada dindingnya saja, karena pada diatas dindingnya banyak sarang burung walet.	Negatif	Negatif
68	The real paris van java .. Nuansanya enak dan terawat tempatnya	Positif	Positif
69	Tempat menyenangkan untuk jalan2 sore. Tempatnya bersih	Positif	Positif
70	Ada beberapa fasilitas yang lepas dari pengawasan sehingga terkesan kotor dan kusam	Negatif	Negatif
71	Banyak burung disini, walet dan merpati. Harus hati hati jika tiba terkena kotoran burung terutama di monumen SLG nya, tapi tidak ada penanda terkait keberadaan burung burung disini, kalau lagi apes ya bisa kena karna kurang hati hati	Negatif	Negatif
72	Gak ada kuliner	Negatif	Positif
73	Tempat ini bagus dikunjungi pagi, siang sore, malam. Jika untuk berfoto sangat bagus disore hari karna matahari tidak begitu terik. Dimalam hari dekat parkiran terdapat pasar malam yg buka setiap hari dari pukul 6-12 malam. Selain itu juga terdapat beberapa taman yg bagus untuk digunakan objek berfoto.	Positif	Positif
74	Udaranya segar, halamannya luas.	Positif	Positif
75	Miris karena banyak pengunjung yang buang sampah sembarangan bikin pandangan terganggu.	Negatif	Negatif

No.	Ulasan	Kelas Aktual	Kelas Sistem
76	Waktu CFD area sekitar monumen ini pasti jadi kotor, ntah dari pengunjung dan dari penjual makanan disana masih kurang sadar diri untuk menjaga kebersihan! MIRIS!	Negatif	Negatif
77	Indah sekali. Cocok untuk piknik keluarga. Tapi karna malam ini anginnya kencang jadi super dingin sekali	Positif	Positif
78	Sudah disediakan tempat parkir kendaraan yang luas lengkap dengan mushola +toilet. Untuk menuju area bangunan, wisatawan diharuskan melewati lorong bawah tanah yang sudah dipersiapkan. Untuk memasuki area ini tidak dipungut bangunan alias gratis.	Positif	Positif
79	Salah satu monumen kebanggaan masyarakat kediri.. Banyak pengunjung. Parkir luas.	Positif	Positif
80	Agak kecewa sama penjaga tempat parkir karena helm saya pernah hilang waktu parkir diarea parkir, perlu ditambah kewaspadaannya lagi biar gak ada yng senasib seperti saya	Negatif	Negatif
81	Desain arsitektur monumen yang juga mirip Arc De Triomphe di Prancis ini menurut saya membawa kesan dan makna tersendiri terhadap kota kediri.	Positif	Positif
82	Wisata gratis murah meriah Tempatnya bagus salah satu ikon Kota Kediri Ada stand khusus untuk souvenir dan jajanan Terlebih lagi ada jajan khas yang jarang dijumpai di tempat lain yaitu Cenil (klanting, lopis, klepon)	Positif	Positif
83	Tempat sederhana tapi bikin nyaman karna suasananya yang enak	Positif	Positif
84	Wisata bersih menurut saya dibanding gor yang kotor tidak terawat	Positif	Negatif
85	Area hijau yang ada di sekitar area SLG sangat menyejukkan	Positif	Positif
86	Suka. Tempatnya terbuka dengan angin semilir. Keren banget buat Selfi	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
87	Tempatnya nyaman sllu sedap dipandang mata,kalau pagi hari buat olah raga udaranya bagus	Positif	Positif
88	Kebersihan toilet maupun kebersihan di area monument seperti diabaikan karna terlihat kotor sekali, kesadaran dari pengunjung sangat kurang dan dari pihak pemkab sangat sedikit menyediakan tong sampah dan baiknya menempatkan petugas tambahan. Masih mengecewakan sarana prasarana pendukungnya	Negatif	Negatif
89	Terowongan penyeberangannya kotor	Negatif	Negatif
90	Keren...terowongan dibawah jalan raya	Positif	Positif
91	Enak buat nongkrong karena di kelilingi jalan berputar dengan simpang 5 nya serta di bawah monumen ada terowongan dari dan menuju tempat parkir serta ada foto2 jadul di lorong terowongan tsb	Positif	Positif
92	Lokasinya terlalu luas untuk dijelajahi. Apalagi sekarang area sekitar monumen harus steril dari pedagang kaki lima sehingga cukup menyulitkan pengunjung yang ingin jajan atau belanja.	Negatif	Positif
93	Satu yang kurang menurut saya dari kawasan simpang lima gumul adalah papan petunjuk arah. Karena buat yang pertama berkunjung akan kesulitan dan mengira bahwa untuk ke monumen harus menyeberanh jalan. Padahal bisa lewat terowongan.	Negatif	Positif
94	Monumen berada tepat di tengah-tengah simpang lima gumul dan disediakan akses masuk dari tempat parkir melewati terowongan bawah tanah, tapi pada saat saya kesana beberapa fasilitas seperti toilet dan pusat informasi tidak berfungsi padahal musim liburan bukan weekday	Negatif	Positif
95	Saya waktu liburan kesana sedikit kecewa karna toilet dan pusat informasi tidak terurus dan lebih seperti diabaikan	Negatif	Negatif

No.	Ulasan	Kelas Aktual	Kelas Sistem
96	Bersih dan nyaman, penempatan tumbuhan juga rapi sehingga enak di pandang	Positif	Positif
97	Tempatnya bagus , luas , ada terowongan menuju simpang lima gumul dari tempat parkir	Positif	Positif
98	Perlu di berikan tempat berteduh di area parkir dan area lokasi simpang lima gumul , kalau pas panas , areanya panas sekali	Negatif	Negatif
99	Parkiran luas.. toilet tersedia.. di salah satu area parkir juga tersedia kulineran yg beraneka ragam salah satu nya yg khas dari kediri yakni sate bekicot.	Positif	Positif
100	Sayangnya terowongan disana kurang terawat.. semoga pemerintah setempat mampu memperbaikinya dan untuk parawisatawan agar bisa menjaga nya bersama. untuk terowongan sulit untuk di gunakan oleh pengguna kursi roda	Negatif	Negatif
101	Belum ke Kediri kalau belum ke Simpang Lima Gumul ini, monumen kota kediri ini ikonik banget.	Positif	Positif
102	Bangunannya seperti bangunan di luar negeri. Kalau mau ke sini dr area parkir ada jalan bawah tanah menuju ke monumen ini, jadi tidak mengganggu lalu lintas yg ada disitu. Bagus buat foto foto hehe	Positif	Positif
103	Tempat yang keren buat selfie, wefie, ikon Kota Kediri yang keren abizzz.. Gratis pula, cuma bayar parkir aja,. Apalagi kalau malam, tambah keren aja...	Positif	Positif
104	Beberapa kali menyempatkan diri untuk mampir sekedar duduk duduk sebentar di kawasan simpang lima ini biar berasa sah ada di kota kediri. Kawasannya tertata rapi dan bersih. Wajib mampir buat siapapun yang lagi dolan ke Kediri	Positif	Positif
105	Nyaman buat bersantai dengan keluarga.	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
106	Simpang Lima Gumul bagus, strategis, dan dengan fasilitas terowongan bawah tanah, toilet, info pariwisata, dan sovenir shop.	Positif	Positif
107	Dingin banget udaranya tapi dingin nya terasa sejuk jadi enak	Positif	Positif
108	Lokasi mudah dijangkau, bangunannya megah dan indah .. parkir cukup luas ..	Positif	Positif
109	Tempat wisata yg ramai di kunjungi orang, banyak tempat jajan	Positif	Positif
110	Tempat yang terlihat sederhana namun bagus dan penuh makna	Positif	Positif
111	Cocok nih buat foto, waktu santai. Paling bagus di kunjungi menurut saya waktu malam. Lampu kota menghiasi setiap sudutnyaaa	Positif	Positif
112	Tempat ini bagus..buat foto"..bangunanya indah..ada perpustakaan keliling..ada assesoris dan mewahnya adalah jalur menuju bangunan ini lewat bawah tanah karena posisi bangunan di tengah simpang lima.	Positif	Positif
113	Indahnya kebudayaan Indonesia salah satunya tercermin di monumen slg ini	Positif	Positif
114	Tempat bagus untuk berfoto dan bertamasya	Positif	Positif
115	Ayuk kita singgah ke Gumul yang suasanya adem romantis, dan berbagai macam sudut pandang seseorang hilir mudik ketempat wisata ini	Positif	Positif
116	Salah satu tempat wisata di kediri. Aksesnya hrus melewati terowongan bawah tanah. Jadi parkir kendaraan di tempat yg sudah ditentukan, kmudian kita jalan kaki melewati terowongan bawah tanah. Ingat2 kita masuk dr pintu mana krn itu nanti akan jd jalan kita pulang,, jd jgn smp lupa..	Positif	Positif
117	Jika Anda meliriknyaa, monumen itu benar-benar identik dengan salah satu monumen paling terkenal di Prancis. Namun tempat ini punya sejarah dan cerita tersendiri di balik	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
	bangunan tersebut yang membuat bangunan ini penuh akan makna		
118	Sungguh indah tempatnya, terutama di malam hari, dan sebagai salah satu ikon kab. Kediri.	Positif	Positif
119	Memiliki nilai sejarah yang sangat tinggi. Saat ini dijadikan sebagai salah satu tempat wisata yang menarik. Di situ kita bisa belajar dan melihat lebih banyak lagi saksi saksi bisu perjuangan bangsa Indonesia pada masa lalu.	Positif	Positif
120	Suasannya enak, mudah dijangkau baik dengan kendaraan besar atau kecil.	Positif	Positif
121	Bagus. Parkirnya luas	Positif	Positif
122	Cukup menarik, hanya berfoto-foto landmark Kediri, tidak pernah sepi pengunjung pada malam hari, banyak terdapat pedagang kaki lima di area Pasar Tugu, hari sabtu dan minggu pagi juga ramai oleh pengunjung pasar, rekreasi dan berolahraga lari pagi di Tugu.	Positif	Positif
123	Tempat yang nyaman untuk melepas penat, suasana tenang, tempatnya jg bersih	Positif	Positif
124	Tempat yang bagus untuk membawa anak belajar tentang sejarah Kediri dan apa yang ada di Kediri, semua potensinya terutama potensi wisata. Pemerintah sudah menyediakan fasilitas bagi pengunjung. Area parkir yang luas baik utk kendaraan roda dua ataupun kendaraan roda empat dan seterusnya. Ada lorong unik dan kreatif yang dibangun sebagai penghubung antara tempat parkir dan monumen. Jadi pengunjung tidak usah khawatir untuk menyeberang dari tempat parkir menuju monumen. Ada lorong bawah tanah yang akan mengantar pengunjung dari area parkir menuju monumen, begitu juga sebaliknya.	Positif	Positif
125	Tempat yang sejuk jadi bisa jadi tempat destinasi yang nyaman. Tempatnya luas lagi	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
126	Sering sih ke Gumul Kediri, makin hari makin bagus dan juga sekarang makin dirawat banget. Jadi pengen KEDIRI LAGI	Positif	Positif
127	Disana ada tempat yang cocok untuk anak belajar ttg sejarah, dan taman yang sangat luas untuk bermain.	Positif	Positif
128	Lokasi parkir cukup luas, tempatnya juga sejuk	Positif	Positif
129	Simpang lima gumul anginnya terasa sejuk klo di waktu pagi hari di sa'at mentari mulai naik dari timur.	Positif	Positif
130	Tempat yang bagus untuk dikunjungi ketika berkunjung di Kabupaten Kediri. Disekitar Simpang Lima Gumul juga banyak tempat makan, bermain, taman, tempat belanja(Alfamart, Indomart dan Toko Kelontong), Kolam Renang dan lainnya.	Positif	Positif
131	Tempatnya bagus seperti di paris, buat objek foto bagus, saat saya kesini juga banyak bule yang berfoto foto ria, walau panas.	Positif	Positif
132	Gedung yang Indah dan taman sekitar yang cocok rehat bersama keluarga lengkap dengan fasilitas toilet dan mushola	Positif	Positif
133	Bagus. Ramai. Megah suasana malam lebih indah..	Positif	Positif
134	Tempatnya luas, bersih dan pemandangannya indah	Positif	Positif
135	Enak klo buat nyantai sore sore	Positif	Positif
136	Sangat bagus tempatnya,,, lokasinya strategis sekali..	Positif	Positif
137	Tempat yang bagus untuk bertamasya bersama keluarga	Positif	Positif
138	Tempat kece nya kediri	Positif	Positif
139	Monumen simpang lima gumul sangat direkomendasikan untuk pendatang baru. Anda bisa berwisata kuliner disini. cukup untuk membayar parkir dengan harga 2000 hingga 5000 rupiah, anda sudah bisa memanjakan	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
	mata dengan pemandangan yang indah dan spot foto yang indah		
140	Simpang lima gumul yang terletak di tengah kota kediri disamping bangunannya yang indah, simpang lima gumul banyak dikunjungi wisatawan karena memiliki daya tarik tersendiri, menawarkan bangunan yang indah, simpang lima gumul juga menawarkan pemandangan yang indah dan harga tiket masuk yang terjangkau. cukup terjangkau.	Positif	Positif
141	Hhhh Tempat favorit ber swafoto karena pemandangan yang bagus	Positif	Positif
142	Bersih banget tertata rapi	Positif	Positif
143	Gk tau knapa jauh2 dari malang cuma pingin foto2 disini. Mungkin karena suasana disini membuatku terpana jadi punya keinginan buat kesini terus	Positif	Positif
144	Bangunan monumen nya mempesona mata saya	Positif	Positif
145	Kebersihan sangat terjaga dan tempat yang bagus	Positif	Positif
146	Simpang Lima Gumul tempatnya mudah dijangkau, dekat dengan pusat kuliner, biasanya selalu rame apalagi di hari libur	Positif	Positif
147	Spot foto yang menarik...	Positif	Positif
148	Tempat rekreasi murah meriah namun tetap dengan kualitas yang baik	Positif	Positif
149	Monumen sudah diperbaiki dan sekarang sangat bagus, salah satu destinasi wisata di kediri yg wajib dikunjungi.	Positif	Positif
150	Lampu yang digunakan tidak berlebihan namun hanya menggunakan perpaduan lampu putih dan oren di beberapa tempat sehingga membangun kesan megah pada bangunan ini	Positif	Positif
151	Tempatnya nyaman bersih indah	Positif	Positif
152	Tempatnya sejuk, ada ruang bermain anak	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
153	Bagus buat swaphoto... Apalagi yg mau prewed bagus juga,, hehe.. Suasana nya lebih enak malam,, sejuk anginnya..	Positif	Positif
154	Tempat melepas penat untuk keluarga	Positif	Positif
155	Menarik dikunjungi untuk keluarga, bisa bersantai dengan semilir angin yang sejuk	Positif	Positif
156	Tempat Bagus untuk bersantai dan berswafoto, kunjungi saat malam jika berkenan,	Positif	Positif
157	Bagus...buat liburan keluarga murah meriah	Positif	Positif
158	Tempat yang bagus untuk mengambil foto	Positif	Positif
159	Cocok buat jalan jalan dengan keluarga	Positif	Positif
160	Tempatnya bagus, sejuk, menarik, cocok untuk semua kalangan, tua, muda, anak-anak..	Positif	Positif
161	Parkir 2000 bisa sepuasnya menikmati keindahan di area SLG	Positif	Positif
162	Tempat yang bagus untuk menghabiskan liburan Anda di kediri, tempat ini direkomendasikan untuk Anda yang ingin mengambil foto dengan latar belakang yang baik.	Positif	Positif
163	Baguss .. Angin semilir kadang kenceng , asyik buat nongkrong sama teman , pacar , istri	Positif	Positif
164	Destinasi wisata yang menarik, karena monumennya seperti arc de triomphe, jadi berasa seperti di Paris	Positif	Positif
165	Icon kota kediri yng selalu menjadi jujukan para wisatawan luar kota untuk sport selfie dll	Positif	Positif
166	Tempat cukup nyaman untuk foto2. ada terowongan bawah tanah dari parkiran menuju lokasi slg. Tempat yg harus dikunjungi jika berkunjung ke kediri.	Positif	Positif
167	Monumen megah nan indah, cocok untuk berswafoto,,,"	Positif	Positif
168	Bagus buat foto disaat week day bukan week end soalnya gak terlalu rame. Kalau malam bisa	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
	duduk di rumputnya sambil ngobrol, nyemil tapi jangan lupa bawa jaket karena dingin. Kalo siang bagus juga buat foto.		
169	Lokasi strategis, tempat bersantai bersama keluarga.	Positif	Positif
170	Saya suka tempat ini, ada street food nya, lengkap banget, terus monumen nya juga bagus banget, tempat bersih cocok untuk santai, nongkrong dan kumpul keluarga. Dan juga masuk monumen gratis alias tidak dipungut biaya, hanya parkir saja, kalau motor 3k kalau mobil kurang tau siihh	Positif	Positif
171	Dekat rumah....pemandangan nya jg keren...buat foto2 its ok	Positif	Positif
172	Menikmati ramainya slg dan bisa jadi pilihan liburan rakyat gratis	Positif	Positif
173	Tempat yang cocok untuk piknik bersama keluarga	Positif	Positif
174	Tempat keren buat SwaFoto ...Lokasi Wisata yg bagus. Gunakan Jaket, karena daerahnya sangat berangin dan dingin.	Positif	Positif
175	Menjadi ikon kabupaten Kediri, terletak di jalan antara Kediri-Pare, tempatnya bagus dan selalu ramai dikunjungi banyak orang	Positif	Positif
176	Tempatnya luas dan enak buat cuci mata sama melepas penat	Positif	Positif
177	Pemandangan dan suasana disini adalah yang nomer 1 pokoknya	Positif	Positif
178	Arek bermain anak yang nyaman dan aman, jadi sebagai orang tua tidak perlu khawatir berlebihan bun.	Positif	Positif
179	Tempat yang bagus untuk bersantai	Positif	Positif
180	Udaranya sejuk buat bersantai	Positif	Positif
181	Area luas rapi, ada tempat parkir, sejuk, nyaman	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
182	Tempat bermain dan nongkrong untuk orang tua dan anak hehe	Positif	Positif
183	Taman gumul ini sesuai namanya, terletak di dekat monumen simpang lima gumul. Di beberapa hari taman ini terlihat sepi dan hanya segelintir orang atau pedagang yang erada di taman ini. Taman ini memiliki banyak wahana bermain juga untuk anak anak merasa betah berlama lama disini	Positif	Positif
184	Selain di monumen SLG nya, tamannya juga memberikan nuansa sejuk	Positif	Positif
185	Tempat nya seru kalo malam banyak penjual makanan dan mainan	Positif	Positif
186	Terdapat monumen yg terkenal Simpang Lima Gumul, tempat yang bersejarah ada terowongan bawah tanah untuk mencapai monumen tsb. Saat malam hari anginya cukup kencang dan sejuk.. ditambah lampu" yang cantik..	Positif	Positif
187	Tempat yg cukup dapat membuat inspirasi kemajuan jaman, enak buat nongkrong.	Positif	Positif
188	Selalu ramai dikunjungi oleh wisatawan ..	Positif	Positif
189	Tempat yang bagus untuk Anda yang ingin bersantai jika mengunjungi kediri	Positif	Positif
190	Cocok untuk bersantai bersama teman pacar atau keluarga	Positif	Positif
191	Tersedia taman kecil di dekat monumen simpang lima gumul yang cocok untuk rekreasi keluarga. Terbukti anak saya nyaman main disana	Positif	Positif
192	Di bagian bawah parkir SLG ada taman yang indah dan bersih. Banyak tempat bermain anak anak. Ada ayunan, prosotan dan tempat duduk. Tersedia banyak bak sampai di pinggirnya. Toilet juga ada. Jika mau kesini anda bisa parkir di parkir selatan tugu gumul. Bayar parkir hanya 2.000 saja. Tempat	Positif	Positif

No.	Ulasan	Kelas Aktual	Kelas Sistem
	parkir luas. Bagi anda yang kelaparan, banyak juga jajanan dan minuman disana		
193	Banyak pedagang mulai makanan minuman hingga peralatan rumah tangga dan Tamannya Indah	Positif	Positif
194	Tempatnya sejuk, teduh dan rindang... Cocok untuk berolah raga	Positif	Positif
195	Taman taman yang tersedia di area SLG lumayan bagus untuk tmpat berkumpul bersama teman dan keluarga dengan fasilitas arena bermain anak	Positif	Positif
196	Mendinginkan pikiran gk perlu mahal. Cukup ke SLG semua stress bisa hilang dan kita jadi lebih tenang	Positif	Positif
197	Cocok dijadikan salah satu pilihan wisata keluarga	Positif	Positif
198	Lokasi muda dijangkau	Positif	Positif
199	Enak,adem,banyak jajanan,cocok buat liburan murah meriah sma kluarga	Positif	Positif
200	Nyaman, adem, tempat bersantai, aman untuk anak-anak	Positif	Positif