

**ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA
MENGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR
INFORMATION GAIN**

TUGAS AKHIR

Sebagai salah satu syarat untuk memperoleh gelar sarjana S-1 di Program Studi Informatika,
Jurusan Teknik Informatika, Fakultas Teknik Industri, Universitas Pembangunan Nasional
“Veteran” Yogyakarta



Disusun Oleh :

RAHMATSYAH HABIBIE HARAHAP

NIM : 123140135

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
YOGYAKARTA**

2021

HALAMAN PENGESAHAN PEMBIMBING

**ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA
MENGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR
INFORMATION GAIN**

Disusun oleh :

Rahmatsyah Habibie Harahap
123140135

Telah diuji dan dinyatakan lulus oleh pembimbing
pada tanggal:

Menyetujui,
Pembimbing I

Pembimbing II

Oliver S. Simanjuntak, S.Kom., M.Eng.
NIK. 2 8305 11 0300 1

Wilis Kaswidjanti, S.Si., M.Kom.
NIK. 2 7604 00 0226 1

Mengetahui,
Ketua Program Studi

Dr. Heriyanto, A.Md., S.Kom., M.Cs
NIK. 2 7706 11 0301 1

HALAMAN PENGESAHAN PEMBIMBING

ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA
MENGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR
INFORMATION GAIN

Disusun oleh :

Rahmatsyah Habibie Harahap
123140135

Telah diuji dan dinyatakan lulus
pada tanggal: oleh :

Menyetujui,
Penguji I

Penguji II

Oliver S. Simanjuntak, S.Kom., M.Eng
NIK. 2 8305 11 0300 1

Wilis Kaswidjanti, S.Si., M.Kom.
NIK. 2 7604 00 0226 1

Penguji III

Penguji IV

Mangaras Yanu F., S.T., M.Eng.
NIK. 2 8201 13 0425 1

Juwairiah, S.Si., M.T.
NIK. 2 7607 00 0230 1

SURAT PERNYATAAN
KARYA ASLI TUGAS AKHIR

Sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknik Industri Universitas Pembangunan Nasional “Veteran” Yogyakarta, yang bertanda tangan dibawah ini, saya :

Nama : Rahmatsyah Habibie Harahap
No. Mhs : 123140135

Menyatakan bahwa karya ilmiah saya yang berjudul :

**ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA
MENGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR
INFORMATION GAIN**

Merupakan karya asli saya dan belum pernah dipublikasikan dimanapun. Apabila dikemudian hari, karya saya disinyalir bukan merupakan karya asli saya, maka saya bersedia menerima konsekuensi apapun yang diberikan Program Studi Informatika Fakultas Teknik Industri Universitas Pembangunan Nasional “Veteran” Yogyakarta kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta

Pada tanggal : 21 Juni 2021

Yang Menyatakan



Rahmatsyah Habibie Harahap
123140135

PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini :

Nama : Rahmatsyah Habibie Harahap
No. Mhs : 123140135
Fakultas/Prodi : Teknik Industri/Informatika

Dengan ini saya menyatakan bahwa judul Tugas Akhir

ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA MENGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR INFORMATION GAIN

Adalah hasil kerja saya sendiri dan benar bebas dari plagiasi kecuali cuplikan serta ringkasan yang terdapat di dalamnya telah saya jelaskan sumbernya (sitasi) dengan jelas. Apabila pernyataan ini terbukti tidak benar maka saya bersedia menerima sanksi sesuai peraturan Mendiknas RI No. 17 Tahun 2010 dan Peraturan Perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan penuh tanggung jawab.

Yogyakarta,

Yang membuat Pernyataan



Rahmatsyah Habibie Harahap
123140135

HALAMAN PERSEMBAHAN

Sujud syukur saya persembahkan kepadaMu ya Allah, Tuhan Yang Maha Agung dan Maha Tinggi. Atas takdirmu saya bisa menjadi pribadi yang berpikir, berilmu, beriman dan bersabar. Semoga keberhasilan ini menjadi satu langkah awal untuk masa depan saya, dalam meraih citacita.

Dengan ini saya persembahkan karya ini untuk, Kedua orang tua saya. Terima kasih atas kasih sayang yang berlimpah dari mulai saya lahir, hingga saya sudah sebesar ini, terima kasih juga atas limpahan doa yang tak berkesudahan. Serta segala hal terbaik yang telah dilakukan.

ABSTRAK

Melalui Kementerian Pendidikan dan Kebudayaan Pemerintah telah melarang perguruan tinggi untuk melaksanakan perkuliahan tatap muka (konvensional) dan memerintahkan untuk menyelenggarakan perkuliahan atau pembelajaran secara *daring* (Surat Edaran Kemendikbud Dikti No. 1 tahun 2020). *Twitter* kerap dimanfaatkan untuk meninjau atau mengambil berbagai pendapat atau opini para pengguna yang nantinya akan diolah menjadi sebuah data, yang kemudian data tersebut akan diolah kembali untuk menemukan sebuah solusi dari suatu permasalahan yang ada. Analisis sentimen dilakukan untuk mengetahui pendapat atau opini yang dikemukakan dalam kalimat atau dokumen tersebut, apakah akan bernilai positif, negatif atau netral. Untuk menganalisis suatu sentimen diperlukan adanya metode yang bisa mengelompokkan suatu sentimen, Maximum Entropy merupakan metode yang digunakan pada penelitian ini dengan metode evaluasi yang digunakan adalah Confusion Matrix yang nantinya akan menghitung nilai Macro dan Micro averaging dari nilai evaluasi yang dihasilkan. Hasil evaluasi yang dilakukan dalam penelitian ini menghasilkan nilai akurasi Macro yang cukup tinggi yaitu sebesar 68% dengan nilai precision 72% dan recall sebesar 92% . Pengujian dilakukan dengan menguji sebanyak 250 tweet.

Kata Kunci : twitter, analisis sentimen, kuliah online, maximum entropy, information gain

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah, dan inayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul

“ANALISIS SENTIMEN PEMBELAJARAN DARING PADA MAHASISWA MENGGUNAKAN METODE MAXIMUM ENTROPY DENGAN SELEKSI FITUR INFORMATION GAIN”.

Penyusunan ini merupakan salah satu syarat yang harus dilaksanakan oleh setiap mahasiswa dalam menyelesaikan jenjang pendidikan S1 di Program Studi Teknik Informatika Fakultas Teknik Industri Universitas Pembangunan Nasional “Veteran” Yogyakarta. Didalam penyusunan tugas akhir ini, penulis menyadari bahwa penulisan tugas akhir ini tidak lepas dari bimbingan, bantuan dan dukungan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, penulis sangat mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Kepada Allah SWT yang senantiasa memberikan petunjuk, kemudahan dan kesehatan.
2. Kepada kedua orang tua penulis yang selalu mendoakan, memberikan semangat dan pengorbanannya baik dari segi moril maupun materi kepada penulis, sehingga penulis dapat menyelesaikan skripsi ini.
3. Kepada Bapak Oliver S. Simanjuntak, S.Kom., M.Eng. dan Ibu Wilis Kaswidjanti, S.Si., M.Kom. selaku dosen pembimbing yang telah bersedia membimbing dan mengarahkan penulis selama menyusun skripsi ini dan memberikan banyak ilmu serta solusi dari setiap permasalahan dalam menyelesaikan penulisan skripsi ini.
4. Kepada Bapak Budi Santosa, S.Si., M.T. selaku dosen wali.
5. Kepada Bapak Mangaras Yanu F., S.T., M.Eng. dan Ibu Juwairiah, S.Si., M.T. selaku dosen penguji yang telah bersedia memberikan masukan dalam menyelesaikan penulisan skripsi ini.

6. Kepada Bapak/Ibu dosen dan staff di Program Stud Teknik Informatika, yang telah banyak membantu penulis dan memberikan ilmu pengetahuan selama menempuh pendidikan di Universitas Pembangunan Nasional “Veteran” Yogyakarta.
7. Kepada seluruh teman-teman yang turut membantu dan memberikan semangat kepada saya dalam menyusun skripsi ini.

Semoga laporan tugas akhir ini dapat bermanfaat bagi semua pihak. Sekali lagi penulis juga mengucapkan terimakasih kepada semua pihak yang telah membantu. Semoga Allah SWT selalu meridhoi semua umatnya, Aamiin Ya Robbal Alamiin.

Wassalammu alaikum warahmatullahi wabarakatuh

Yogyakarta, 18 Juni 2021

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN PEMBIMBING	ii
HALAMAN PENGESAHAN PENGUJI	iii
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR	iv
SURAT PERNYATAAN BEBAS PLAGIAT	v
HALAMAN PERSEMBAHAN	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR MODUL	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian.....	4
1.6. Metode Penelitian.....	5
1.6.1. Metode Pengumpulan Data.....	5
1.6.2. Metode Pengembangan Sistem	6
1.7. Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA	8
2.1. <i>Text Mining</i>	8
2.2. Analisis Sentimen.....	8
2.2.1. Level Analisis Sentimen	9
2.3. Media Sosial.....	10
2.4. <i>Web Scraping</i>	10
2.5. <i>Machine Learning</i>	11
2.6. <i>Text Preprocessing</i>	11
2.4.1. <i>Cleansing & Case Folding</i>	12
2.4.3. <i>Tokenizing</i> (Tokenisasi).....	12
2.4.4. <i>Stopword Removal</i>	12
2.4.5. <i>Stemming</i>	12
2.7. Seleksi Fitur	16
2.8. <i>Information Gain</i>	17
2.9. <i>Maximum Entropy</i>	18
2.10. Pengujian Klasifikasi.....	20
2.7. Penelitian Sebelumnya	21
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM	30
3.1. Metodologi Penelitian dan Pengembangan Sistem	30
3.1.1. Studi Literatur	31
3.1.2. Pengumpulan Data Sekunder	31
3.1.3. Analisis Masalah	31
3.1.4. Metode Pengembangan Sistem	31

3.2. Analisis Kebutuhan	32
3.2.1. Kebutuhan Data <i>Input</i>	33
3.2.2. Kebutuhan Proses Sistem.....	34
3.2.3. Kebutuhan Proses <i>Output</i>	34
3.2.4. Kebutuhan Perangkat Lunak	35
3.3. Perancangan Sistem.....	35
3.3.1. <i>Data Flow Diagram</i> (DFD)	38
3.3.2. <i>Flowchart</i>	38
3.4. Seleksi Fitur <i>Information Gain</i>	42
3.5. Metode <i>Maximum Entropy</i>	45
3.6. Perancangan Basis Data.....	52
3.6.1. <i>Entity Relations Diagram</i> (ERD)	52
3.6.2. Perancangan Struktur Tabel	53
3.6.3. Relasi Antar Tabel (RAT).....	56
3.7. Perancangan Antarmuka Sistem	57
3.7.1. Perancangan Antarmuka Admin	57
BAB IV HASIL DAN PEMBAHASAN.....	64
4.1. Implementasi Aplikasi.....	64
4.1.1. Halaman <i>Login</i>	64
4.1.2. Halaman Beranda	65
4.1.3. Halaman Dataset.....	66
4.1.4. Halaman Karakteristik Kata.....	67
4.1.5. Halaman <i>Stopword</i>	68
4.1.6. Halaman Kata Baku.....	70
4.1.7. Halaman <i>Information Gain</i>	72
4.1.8 Halaman <i>Training</i>	81
4.1.9. Halaman <i>Testing</i>	82
4.1.10. Halaman <i>Analysis</i>	88
4.1.11. Halaman Grafik Visualisasi	94
4.2. Pengujian	96
4.2.1. Pengujian <i>Confusion Matrix</i>	96
4.3. Pembahasan.....	97
BAB V PENUTUP.....	99
5.1. Kesimpulan	99
5.2. Saran	99
DAFTAR PUSTAKA.....	100

DAFTAR GAMBAR

Gambar 3.1. Metodologi Penelitian.....	30
Gambar 3.2. <i>Data Flow Diagram Level 0</i>	36
Gambar 3.3. <i>Data Flow Diagram Level 1</i>	37
Gambar 3.4. <i>Flowchart Preprocessing</i>	40
Gambar 3.5. <i>Flowchart Information Gain</i>	41
Gambar 3.6. <i>Flowchart Maximum Entropy</i>	42
Gambar 3.7. <i>Entity Relations Diagram</i>	53
Gambar 3.8. Relasi Antar Tabel (RAT).....	56
Gambar 3.9. Rancangan Halaman <i>Login Admin</i>	57
Gambar 3.10. Rancangan Halaman Beranda	58
Gambar 3.11. Rancangan Halaman Dataset	58
Gambar 3.12. Rancangan Halaman <i>Input Dataset</i>	59
Gambar 3.13. Rancangan Halaman <i>Stopwords</i>	59
Gambar 3.14. Rancangan Halaman <i>Input Stopwords</i>	60
Gambar 3.15. Rancangan Halaman Kata Baku.....	60
Gambar 3.16. Rancangan Halaman <i>Input Kata Baku</i>	61
Gambar 3.17. Rancangan Halaman <i>Information Gain</i>	61
Gambar 3.18. Rancangan Halaman <i>Training</i>	62
Gambar 3.19. Rancangan Halaman <i>Testing</i>	62
Gambar 3.20. Rancangan Halaman <i>Analysis</i>	63
Gambar 3.21. Rancangan Halaman Grafik	63
Gambar 4.1. Halaman <i>Login</i>	64
Gambar 4.2. Halaman Beranda	65
Gambar 4.3. Halaman Dataset.....	66
Gambar 4.4. Halaman Karakteristik Kata.....	67
Gambar 4.5. Halaman <i>Stopword</i>	68
Gambar 4.6. Halaman Kata Baku.....	70
Gambar 4.7. Halaman <i>Information Gain</i>	72
Gambar 4.8. Halaman <i>Training</i>	81
Gambar 4.9. Halaman <i>Testing</i>	82
Gambar 4.10. Halaman <i>Analysis</i>	89
Gambar 4.11. Halaman Grafik Visualisasi	94

DAFTAR TABEL

Tabel 2.1. Afiks Yang Tidak Dibolehkan.....	15
Tabel 2.2. Pengujian Klasifikasi	20
Tabel 2.3. Lanjutan Penelitian Sebelumnya	26
Tabel 2.4. Lanjutan Penelitian Sebelumnya	27
Tabel 2.5. Lanjutan Penelitian Sebelumnya	29
Tabel 3.1. Contoh Case Folding.....	38
Tabel 3.2. Contoh Tokenizing.....	39
Tabel 3.3. Contoh Stopwords Removal.....	39
Tabel 3.4. Contoh Stemming	40
Tabel 3.5. Total Dataset.....	43
Tabel 3.6. Data Komentar.....	43
Tabel 3.7. Data Term.....	43
Tabel 3.8. Lanjutan Data Term	43
Tabel 3.9. Hasil Nilai Entropy Kelas Sentimen	46
Tabel 3.10. Lanjutan Hasil Nilai Entropy Kelas Sentimen.....	47
Tabel 3.11. Hasil Nilai Z	47
Tabel 3.12. Lanjutan Hasil Nilai Z.....	48
Tabel 3.13. Hasil Nilai Probabilitas	49
Tabel 3.14. Hasil Nilai Probabilitas Sentimen.....	52
Tabel 3.15. tbl_user	53
Tabel 3.16. tbl_dataset.....	54
Tabel 3.17. tbl_stopword	54
Tabel 3.18. tbl_katadasar	54
Tabel 3.19. tbl_tidakbaku	54
Tabel 3.20. tbl_info_word	55
Tabel 3.21. tbl_tfidf.....	55
Tabel 3.22. tbl_tfidfhasil.....	55
Tabel 3.23. tbl_komentar_uji.....	56
Tabel 3.24. tbl_analysis	56
Tabel 4.1. <i>Confusion Matrix</i>	96
Tabel 4.2. Hasil Pengujian.....	97

DAFTAR MODUL

Modul Program 4.1. <i>Source Code</i> Halaman <i>Login</i>	64
Modul Program 4.2. Lanjutan <i>Source Code</i> Halaman <i>Login</i>	65
Modul Program 4.3. <i>Source Code</i> Halaman Beranda	65
Modul Program 4.4. Lanjutan <i>Source Code</i> Halaman Beranda	66
Modul Program 4.5. <i>Source Code</i> Halaman Dataset	67
Modul Program 4.6. <i>Source Code</i> Halaman <i>Stopwords</i>	68
Modul Program 4.7. Lanjutan <i>Source Code</i> Halaman <i>Stopwords</i>	69
Modul Program 4.8. <i>Source Code</i> Halaman Kata Baku	70
Modul Program 4.9. Lanjutan <i>Source Code</i> Halaman Kata Baku.....	71
Modul Program 4.10. <i>Source Code</i> Halaman <i>Information Gain</i>	72
Modul Program 4.11. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	73
Modul Program 4.12. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	74
Modul Program 4.13. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	75
Modul Program 4.14. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	76
Modul Program 4.15. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	77
Modul Program 4.16. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	78
Modul Program 4.17. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	79
Modul Program 4.18. Lanjutan <i>Source Code</i> Halaman <i>Information Gain</i>	80
Modul Program 4.19. <i>Source Code</i> Halaman <i>Training</i>	81
Modul Program 4.20. Lanjutan <i>Source Code</i> Halaman <i>Training</i>	82
Modul Program 4.21. <i>Source Code</i> Halaman <i>Testing</i>	82
Modul Program 4.22. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	83
Modul Program 4.23. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	84
Modul Program 4.24. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	85
Modul Program 4.25. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	86
Modul Program 4.26. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	87
Modul Program 4.27. Lanjutan <i>Source Code</i> Halaman <i>Testing</i>	88
Modul Program 4.28. <i>Source Code</i> Halaman <i>Analysis</i>	89
Modul Program 4.29. Lanjutan <i>Source Code</i> Halaman <i>Analysis</i>	90
Modul Program 4.30. Lanjutan <i>Source Code</i> Halaman <i>Analysis</i>	91
Modul Program 4.31. Lanjutan <i>Source Code</i> Halaman <i>Analysis</i>	92
Modul Program 4.32. Lanjutan <i>Source Code</i> Halaman <i>Analysis</i>	93
Modul Program 4.33. Lanjutan <i>Source Code</i> Halaman <i>Analysis</i>	94
Modul Program 4.34. <i>Source Code</i> Halaman Grafik Visualisasi.....	94
Modul Program 4.35. Lanjutan <i>Source Code</i> Halaman Grafik Visualisasi	95
Modul Program 4.36. Lanjutan <i>Source Code</i> Halaman Grafik Visualisasi	95

BAB I

PENDAHULUAN

1.1. Latar Belakang

Wabah *corona virus disease* 2019 (Covid-19) yang telah melanda dunia, memberikan tantangan tersendiri bagi lembaga-lembaga yang ada, khususnya pada lembaga pendidikan di Indonesia. Dalam mencegah penyebaran dari virus Covid-19 ini, pemerintah telah melarang kegiatan yang menyebabkan terjadinya kerumuman, adanya pembatasan sosial (*social distancing*) dan menjaga jarak fisik (*physical distancing*). Melalui Kementerian Pendidikan dan Kebudayaan Pemerintah telah melarang perguruan tinggi untuk melaksanakan perkuliahan tatap muka (konvensional) dan memerintahkan untuk menyelenggarakan perkuliahan atau pembelajaran secara *daring* (Surat Edaran Kemendikbud Dikti No. 1 tahun 2020). Perguruan tinggi dituntut untuk dapat menyelenggarakan pembelajaran secara *daring* atau *online* (Firman, F., 2020).

Banyak mahasiswa yang mengemukakan pendapat atau beropini tentang pembelajaran *daring* yang dilakukan di masa pandemi ini. Sebagian besar para mahasiswa berkomentar atau berpendapat lewat akun pribadi media sosial mereka. Sehingga memunculkan beragam pendapat atau opini yang bersifat netral, positif, maupun negatif. Perkembangan media yang begitu pesat, memunculkan banyak media online dari media berita sampai media sosial. Contohnya dari *Facebook*, *Twitter*, *Path*, *Instagram*, *Google+*, *Tumblr*, *LinkedIn* dan sebagainya masih banyak lagi (Rofqoh et al., n.d.). Pada media sosial khususnya *twitter*, kerap dimanfaatkan sebagai tempat mengemukakan pendapat atau opini maupun mengeluarkan keluh kesah atau keresahan yang dihadapi oleh banyak mahasiswa. Tidak jarang juga *twitter* menjadi tempat untuk melakukan diskusi tentang berbagai hal yang menimbulkan pro dan kontra. *Twitter* juga kerap dimanfaatkan untuk meninjau atau mengambil berbagai pendapat atau opini para pengguna yang nantinya akan diolah menjadi

sebuah data, yang kemudian data tersebut akan diolah kembali untuk menemukan sebuah solusi dari suatu permasalahan yang ada. Analisis sentimen dilakukan untuk mengetahui pendapat atau opini yang dikemukakan dalam kalimat atau dokumen tersebut, apakah akan bernilai positif, negatif atau netral.

Analisis sentimen merupakan ilmu yang mempelajari bagaimana menganalisa opini, pesan, sentimen, evaluasi, penilaian, emosi dari suatu entitas berupa produk, pelayanan, organisasi, permasalahan, topik dan atribut (Liu, 2012). Melalui analisis sentimen dapat diketahui sifat dari opini, komentar atau sentimen yang disampaikan oleh mahasiswa, apakah akan bersifat positif, negatif atau bahkan netral. Pada proses analisis sentimen dilakukan dengan menggunakan penerapan dari algoritma *machine learning* yang dapat membantu proses klasifikasi. Salah satu algoritma machine learning yang bisa digunakan adalah *maximum entropy* dan dengan menggunakan tambahan seleksi fitur *information gain* yang mampu menambah akurasi dari analisis.

Sebelumnya pernah dilakukan penelitian yang berjudul analisis sentimen pemilihan presiden 2019 pada *twitter* menggunakan metode *maximum entropy*. Penelitian ini menganalisis opini yang berasal dari ungkapan seseorang melalui media sosial *twitter* tentang pemilihan presiden pada tahun 2019. Hasil dari analisis sentimen pada penelitian ini berupa kesimpulan apakah sentimen bernilai positif atau negatif. Pada penelitian ini juga akan digunakan sebagai prediksi elektabilitas pasangan calon presiden yang akan maju pada pemilihan presiden pada tahun 2019. Teknik pengumpulan data yang dilakukan pada penelitian ini adalah dengan mengambil *tweet* berbahasa Indonesia yang membahas tentang Pilpres 2019. Pengumpulan data dilakukan dengan cara menyaring *tweet* dengan beberapa *keywords* yang berkaitan dengan Pilpres 2019. Data yang telah diolah akan digunakan sebagai dataset yang nantinya akan menjadi data latih dan data uji dengan jumlah yang seimbang pada masing-masing kelas positif dan negatif. Data latih yang digunakan sebanyak

300 data *tweet* yang mana data tersebut memiliki jumlah 150 *tweet* positif dan 150 *tweet* negatif. Metode *maximum entropy* yang digunakan dalam penelitian ini untuk menganalisis data sentimen memiliki hasil akurasi sebesar 89,16% dengan nilai *precision* 100%, *recall* 89,16% dan *f-measure* sebesar 94,27%. Hasil evaluasi dari penelitian ini dengan menggunakan metode *maximum entropy* menghasilkan nilai *macro* dan *micro* yang sama nilai evaluasi bisa ditingkatkan dengan menambahkan jumlah data latih yang digunakan. Semakin banyak data latih maka nilai evaluasi sistem akan semakin baik (Sabily et al., 2019).

Berdasarkan permasalahan yang ditemukan, inilah yang menjadi ide dasar dalam penelitian ini untuk mengetahui pendapat atau opini para mahasiswa yang berkomentar di *twitter* tentang pembelajaran *daring*. Pengumpulan data akan dilakukan dengan cara menyaring *tweet* dengan *keywords* yang berkaitan dengan pembelajaran *daring*. Data yang telah diambil akan diolah menjadi *dataset* yang akan menjadi data latih dan data uji. Proses klasifikasi pada metode *maximum entropy* hanya akan menggunakan informasi kemunculan dari suatu fitur dalam sebuah dokumen (Anggraeni, 2008). Secara garis besar *maximum entropy* akan mencari probabilitas yang paling sama dengan menggunakan asumsi minimal. Pengklasifikasian akan dibagi menjadi 3 kategori yaitu positif, negatif dan netral. Seleksi fitur *information gain* pada proses klasifikasi akan digunakan untuk mengukur seberapa sering muncul suatu fitur dalam kelas tertentu yang akan dibandingkan seberapa sering fitur tersebut muncul dalam seluruh kelas yang ada. Fitur *information gain* akan menghitung seberapa banyak *term* yang dipakai untuk mengetahui *lexical* teks dalam proses klasifikasi (Lei, 2002).

Penggunaan metode *maximum entropy* dengan seleksi fitur *information gain* untuk melakukan klasifikasi komentar sudah pernah diterapkan, namun hanya menggunakan dua klasifikasi sentimen, yaitu sentimen positif dan negatif, belum terdapat sentimen yang bersifat netral. Oleh karena itu, penelitian yang akan dilakukan menggunakan metode

maximum entropy dengan seleksi fitur *information gain* ini akan dibagi menjadi 3 kelas sentimen pada klasifikasi komentar yaitu, positif, negatif dan netral. Berdasarkan uraian diatas, maka akan dilakukan penelitian tentang analisis sentimen pembelajaran *daring* pada mahasiswa berdasarkan *tweet* pada *twitter* menggunakan metode *maximum entropy* dengan seleksi fitur *information gain*.

1.2. Rumusan Masalah

Rumusan dari permasalahan diatas adalah “Bagaimana cara mengklasifikasikan komentar atau persepsi mahasiswa terhadap proses pembelajaran daring melalui komentar pada *Twitter* dengan pencarian *keywords* terkait ?“

1.3. Batasan masalah

Agar masalah yang dibahas menjadi fokus dan lebih jelas dalam mencapai sasaran, maka dibuat batasan dari perumusan masalah di atas, diantaranya adalah :

1. Analisis dilakukan berdasarkan *keywords* terkait kuliah online ataupun belajar daring yang ada pada media *twitter*.
2. Data komentar yang digunakan merupakan data komentar berbahasa Indonesia.
3. Data sentimen akan dikategorikan dalam 3 kelas yaitu, positif, negatif dan netral.

1.4. Tujuan Penelitian

Dari rumusan masalah di atas, maka tujuan dari penulisan tugas akhir ini adalah :

1. Mengimplementasikan metode *maximum entropy* dengan seleksi fitur *information gain* untuk menghasilkan ekstraksi komentar *twitter* pada analisis sentimen.
2. Menghasilkan sebuah sistem yang dapat digunakan untuk menemukan informasi terkait sifat dari pendapat mahasiswa/i terkait kuliah *online*.

1.5. Manfaat Penelitian

Adapun manfaat pada penelitian ini adalah sebagai berikut :

1. Dapat memberikan informasi terkait performa *maximum entropy* dengan seleksi fitur *information gain* dalam melakukan analisis sentimen untuk melakukan klasifikasi komentar dan opini mahasiswa pada komentar *twitter*.
2. Memberikan informasi pendapat mahasiswa/i dalam menghadapi kuliah secara *online* atau pembelajaran *daring*.
3. Penggunaan metode *maximum entropy* dan fitur *information gain* untuk analisis sentimen secara kontinu di masa yang akan datang.

1.6. Metode Penelitian

Metodologi yang digunakan dalam penelitian ini terdiri atas metode pengumpulan data dan metode pengembangan sistem.

1.6.1. Metode Pengumpulan Data

a. Studi Literatur

Tahap studi literatur merupakan tahapan dalam mencari, mengumpulkan dan membaca referensi yang bertujuan untuk memperdalam pembahasan mater yang berkaitan dengan fitur *information gain* dan metode *maximum entropy*. Sumber studi literatur yang digunakan berasal dari jurnal penelitian ilmiah, artikel online, e-book atau penelitian yang berkaitan.

b. Teknik Pengumpulan Data

Teknik pengumpulan data yang akan digunakan di dapatkan dari data yang diperoleh melalui aplikasi sosial media *Twitter*. Data yang diambil merupakan data komentar tentang kuliah *online*.

c. Teknik Analisa Data

Setelah melakukan pengumpulan data, maka akan dilakukan tahap analisa data. Pada tahap analisa data dilakukan dengan cara memilah data untuk dijadikan sebagai data latih.

1.6.2. Metode Pengembangan Sistem

Metode pengembangan sistem yang akan digunakan dalam penelitian ini yaitu menggunakan metode waterfall. Yang mana metode pengembangan sistem ini menyediakan dalam pendekatan alur hidup perangkat lunak secara terurut atau linear, dengan tahapan-tahapannya sebagai berikut (Pressman, 2012) :

a. Requirements Analysis

Tahap ini merupakan proses analisa atau pengumpulan data-data yang berkaitan dengan sistem yang akan dibuat.

b. Design

Tahap ini merupakan proses pembangunan struktur data serta perancangan *interface* dari sistem yang akan dibuat.

c. Coding & Testing

Pada tahap ini akan dilakukan penggabungan dari modul-modul yang telah dibuat dan dilakukam pengujian untuk mengetahui apakah ada kesalahan yang terjadi atau tidak.

d. Implementation

Tahap ini merupakan proses dari modul-modul yang sudah diintegrasikan mejadi sistem yang lengkap.

e. Maintenance

Pada tahap ini akan dilakukan pemeliharaan sistem yang telah dibuat dan telah diterapkan, maka tahap ini merupakan tahapan akhir dari metode waterfall.

1.7. Sistematika Penulisan

Sistematika penulisan pada penelitian disusun dalam lima bab yang masing-masing bab akan dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Dalam bab ini menjelaskan tentang teori-teori yang menjadi landasan penelitian yang akan dilakukan dalam mendukung penelitian ini, seperti tinjauan literatur mengenai karya ilmiah sebagai referensi, gambaran pembelajaran, serta pemodelan yang akan menggunakan *Data Flow Diagram* (DFD), metode *Maximum Entropy* dan fitur *Information Gain* dalam membangun sistem analisis sentimen tentang kuliah *online*.

BAB III METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM

Pada bab ini berisi mengenai metode penelitian, pengumpulan data dan analisis data. Analisis data dibuat dengan mengkombinasikan pengumpulan data yang diperoleh. Serta cara bagaimana aktifitas kerangka kerja dalam tiap tahapnya serta gambaran umum perancangan sistem.

BAB IV HASIL, PENGUJIAN DAN PEMBAHASAN

Bab empat berisi secara rinci penerapan dari analisis data dan pengolahan data yang dibuat menjadi sistem dokumen manajemen dengan menampilkan antarmuka program disertai cara kerja dan penggunaan program.

BAB V PENUTUP

Bab lima merupakan penutup dari penjelasan yang berisi tentang kesimpulan dan saran dari hasil sistem.

BAB II

TINJAUAN PUSTAKA

2.1. Text Mining

Text mining adalah proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut (Aditya, 2015).

Kunci dari proses pada *text mining* adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber (Hearst, 2003). Proses *text mining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep/entitas, produksi taksonomi granular, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas, yaitu pembelajaran hubungan antar entitas (Bridge, 2011).

2.2. Analisis Sentimen

Analisis sentimen atau dapat disebut juga sebagai *opinion mining* adalah bidang studi yang menganalisa pendapat, sentimen, evaluasi, penilaian, dan emosi banyak orang mengenai entitas seperti produk, pelayanan, organisasi, individu, permasalahan, acara, topik, dan atributnya (Liu, 2012). Analisis sentimen dilakukan dengan tujuan untuk meninjau atau melihat pendapat terhadap suatu masalah yang sedang menjadi topik pembicaraan.

Tujuan analisis sentimen adalah bagaimana komputer dapat mengenali tingkat emosional positif maupun negatif dari dokumen teks (Mejova, 2009). Secara umum dokumen teks memiliki 2 makna sentimen yaitu sentimen positif dan negatif. Namun, pada sebuah teks memiliki kemungkinan yang berisi pernyataan ganda atau memiliki makna campuran.

2.2.1. Level Analisis Sentimen

Menurut Liu (2012), terdapat beberapa tingkat analisis yang dapat dilakukan dalam analisis sentimen, diantaranya adalah sebagai berikut:

1. *Document Level*

Analisis terhadap sentimen atau opini seseorang dilakukan pada *level* dokumen secara menyeluruh. Apakah dokumen tersebut memiliki sentimen yang bernilai positif atau negatif. Dokumen yang dianalisa dianggap sebagai sebuah entitas tunggal.

2. *Sentence Level*

Analisis sentimen atau opini ini dilakukan pada *level* kalimat. Kalimat-kalimat yang diungkapkan oleh seseorang nantinya akan dianalisis dan diklasifikasikan menjadi kalimat positif, negatif ataupun netral. Netral mempunyai maksud sebagai kalimat yang tidak beropini atau kalimat fakta.

3. *Entity and Aspect Level*

Analisis terhadap sentimen atau opini dilakukan pada *level* yang lebih mendalam. Apabila tidak ditemukan secara pasti kelas dari klasifikasi pada *level* ini maka analisis dilakukan hingga pada *level* struktur bahasa (dokumen, paragraf, kalimat, klausa atau frasa). Pada *level* ini juga dapat mencari ide atau target opini yang membangun suatu opini seseorang.

Analisis sentimen bisa dianggap sebagai kombinasi dari *text mining* dengan *natural language processing*. *Text mining* merupakan cabang ilmu dari *data mining*, dimana *data mining* merupakan proses untuk mencari sebuah pola pada kumpulan data-data yang ada. Pada dasarnya proses kerja dari *text mining* banyak mengadopsi penelitian dari *data mining*, namun yang menjadi perbedaan adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alami yang tidak terstruktur, sedangkan dalam *data mining* pola yang diambil adalah dari data yang terstruktur (Han & Kamber, 2000).

2.3. Media Sosial

Media sosial adalah media yang terdiri atas tiga bagian, yaitu : Infrastruktur informasi dan alat yang digunakan untuk memproduksi dan mendistribusikan isi media, isi media dapat berupa pesan-pesan pribadi, berita, gagasan, dan produk-produk budaya yang berbentuk digital, kemudian yang memproduksi dan mengkonsumsi isi media dalam bentuk digital adalah individu, organisasi, dan industri (Howard & Parks, 2012).

Perkembangan media yang begitu pesat, memunculkan banyak media *online* dari media berita sampai media sosial. Contohnya dari *Facebook, Twitter, Path, Instagram, Google+, Tumblr, LinkedIn* dan sebagainya masih banyak lagi (Rofqoh et al., n.d.). Pada media sosial khususnya *twitter*, kerap dimanfaatkan sebagai tempat untuk mengeluarkan pendapat tentang suatu maupun mengeluarkan keluh kesah atau keresahan yang dihadapi oleh banyak mahasiswa. Tidak jarang juga *twitter* menjadi tempat untuk melakukan diskusi tentang berbagai hal yang menimbulkan pro dan kontra.

2.4. Web Scraping

Web scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, secara umum halaman-halaman web dalam bahasa seperti HTML atau XHTML, menganalisa dokumen tersebut untuk dilakukan pengambilan data tertentu dari halaman tersebut untuk digunakan bagi kepentingan yang lain (Gunawan et al., 2017).

Manfaat dari web scraping agar informasi yang dikeruk lebih terfokus sehingga akan memudahkan dalam melakukan pencarian sesuatu. *Web scraping* juga dikenal sebagai *screen scraping*. *Web scraping* tidak dapat dimasukkan dalam bidang data mining karena *data mining* menyiratkan upaya untuk memahami pola suatu semantik atau tren dalam sejumlah besar data yang sudah diperoleh.

2.5. Machine Learning

Machine learning dapat didefinisikan sebagai metode komputasi dalam meningkatkan performa dalam membuat prediksi yang akurat. Menurut Nielson (1998), *machine learning* mencakup beberapa proses yang sulit didefinisikan untuk mendapatkan pengetahuan, pemahaman, keterampilan dalam belajar berdasarkan pengalaman. Terdapat beberapa proses dalam pembelajaran *machine learning* sebagai berikut :

1. *Supervised Learning*

Memiliki kesamaan dengan klasifikasi, proses *data mining* yang proses pelatihannya menggunakan data latih yang sudah diberi label.

2. *Unsupervised Learning*

Memiliki kesamaan dengan *clustering*, proses *data mining* yang proses pelatihannya menggunakan data latih yang tidak diberi label awal.

3. *Semi-supervised Learning*

Merupakan sebuah teknik pelatihan yang mengkombinasikan data latih yang telah diberi label dengan data latih yang belum diberi label.

4. *Active Learning*

Merupakan pendekatan *machine learning* yang dimana proses pelatihannya melibatkan *user* yang berperan aktif dalam proses pembeajarannya.

2.6. Text Preprocessing

Banyak dokumen teks atau berita tidak punya struktur yang jelas. *Pre-processing* teks merupakan langkah mengganti data yang punya struktur menjadi data yang terstruktur. Data terstruktur ini yang nantinya bisa diolah sesuai dengan kebutuhan. *Pre-processing* sangat penting untuk dokumen teks media sosial yang berisi data yang tidak terstruktur (Mujilawati, 2016). *Cleansing*, *tokenization*, *stopword removal*, dan *stemming* adalah langkah-langkah umum dalam *pre-processing* teks.

2.6.1. Cleansing & Case Folding

Cleansing adalah proses untuk menghilangkan *noise* atau menghilangkan elemen-elemen yang tidak diperlukan dalam sebuah dokumen. Tujuan *cleansing* adalah untuk meningkatkan dan memperbaiki kualitas data. Proses *cleansing* pada komentar dapat berupa menghilangkan tanda baca, menghilangkan angka-angka, menghilangkan *emoticon* dan mengubah karakter menjadi *lowercase*.

2.6.2. Tokenizing (Tokenisasi)

Proses *tokenizing* merupakan proses pemecahan kata-kata yang ada di suatu kalimat. *Tokenizing* dilakukan dengan memisahkan setiap kata dengan spasi. Contoh dari tokenisasi adalah memecah kalimat “saya pergi ke kantor polisi” menjadi kumpulan kata-kata “saya”, “pergi”, “ke”, “kantor”, “polisi”. Setiap kata yang telah dipecah disebut sebagai token dan akan dimanfaatkan untuk menghitung frekuensi dari setiap kata dan akan digunakan dalam proses klasifikasi.

2.6.3. Stopword Removal

Menurut (Feldman & Sanger, 2007), *stopword* merupakan kata-kata umum dan sering digunakan dalam suatu bahasa. Namun kata-kata umum tersebut tidak mempunyai pengaruh dan tidak memiliki nilai pada dokumen teks tersebut. Sehingga kata-kata yang termasuk dalam *stopword* akan dihilangkan, sehingga ukuran pada data juga akan berkurang. Contoh kata-kata yang termasuk dalam *stopword* dalam Bahasa Indonesia adalah “untuk”, “kapan”, “yang”, “di”, “iya”, “itu”, dan lain sebagainya. Kamus *stopwords* yang digunakan berasal dari Nurirwan Saputra yang diunduh di situs <http://hikaruyyuki.lecture.ub.ac.id> (Saputra, 2015).

2.6.4. Stemming

Stemming merupakan proses pemotongan imbuhan pada kata yang memiliki imbuhan, yang dilakukan dengan menggunakan algoritma tertentu untuk mengembalikan suatu kata

kebentuk kata dasarnya (Mauliadi, 2017). Untuk mengubah kata berimbuhan menjadi kata dasar dibutuhkan beberapa aturan yang disebut dalam morfologi. Berikut ini aturan morfologi yang harus diketahui sebelum melakukan proses *stemming* adalah sebagai berikut:

1. Afiks

Afiks merupakan beberapa huruf tertentu yang apabila huruf tersebut ditambahkan pada kata dasar akan mengubah makna dan membentuk suatu kata yang baru.

2. Prefiks

Prefiks merupakan imbuhan yang melekat didepan kata dasar yang membentuk suatu kata yang baru dengan arti yang berbeda.

3. Sufiks

Sufiks merupakan imbuhan yang melekat dibelakang kata dasar yang membentuk suatu kata baru dengan arti yang berbeda.

4. Konfiks

Konfiks merupakan kombinasi dari prefiks dan sufiks. Konfiks merupakan imbuhan yang berada di awal kata dasar dan di akhir kata dasar.

5. Infiks

Infiks merupakan imbuhan yang melekat ditengah kata dasar yang membentuk suatu kata yang baru dengan arti yang berbeda.

Proses inti dari *stemming* adalah untuk membuang kumpulan imbuhan yang berada pada suatu kata dasar. Ada beberapa algoritma yang dapat digunakan untuk melakukan proses *stemming* pada kata berbahasa Indonesia, salah satunya adalah algoritma Nazief-Adriani. Algoritma Nazief-Adriani prosesnya berjalan berdasarkan pada aturan morfologi komprehensif yang mengelompokkan imbuhan yang diperbolehkan dan imbuhan yang tidak diperbolehkan, termasuk morfologi prefiks, sufiks, infiks dan konfiks. Algoritma Nazief-Adriani menggunakan kamus kata dasar yang akan menjadi acuan dan mendukung proses

dari *stemming*, yaitu dengan penyusunan ulang kata-kata yang mengalami proses *stemming*. yaitu penyusunan kembali kata-kata yang mengalami proses *stemming* yang berlebih. Aturan yang digunakan adalah sebagai berikut :

[DP +[DP + [DP+]]] root word [[+DS][+PP][+P]]

1. *Derivation* Prefiks (DP)

Kumpulan awalan yang dapat langsung diberikan pada kata dasar murni atau pada kata dasar yang mendapat penambahan sampai dengan 2 awalan, termasuk didalamnya awalan bermorfologi (“me-”, “be-”, “pe-”, dan ”te-”) dan awalan tidak bermorfologi (“di”, ”ke-” dan “se”).

2. *Root Word* (Kata Dasar)

3. *Derivation Suffixes* (DP)

Kumpulan akhiran asli yang secara langsung ditambahkan pada kata dasar yaitu akhiran “-i”, “-kan”, “-an”.

4. *Possesive Pronoun* (PP)

Kata ganti kepemilikan (“-ku”, “-mu”, “-nya”)

5. *Particle* (P)

Imbuhan partikel, seperti “-lah”, “-kah” dan “-pun”.

Beberapa pondasi aturan tata bahasa yang digunakan dalam penerapan algoritma *stemming* adalah sebagai berikut:

1. Kata yang mengandung kurang 3 karakter tidak mengandung imbuhan, proses *stemming* tidak diterapkan
2. Afiks atau imbuhan tidak pernah berulang, sehingga proses *stemming* hanya menghilangkan satu set imbuhan.
3. Tidak semua kombinasi imbuhan, baik prefiks atau suffiks dapat diperbolehkan. Pada tabel 2.1. dapat dilihat kombinasi imbuhan yang tidak diperbolehkan.

Tabel 2.1. Afiks yang tidak diperbolehkan

Awalan (Prefiks)	Akhiran (Sufiks)
be-	-i
di-	-an
ke-	-i -kan
me-	-an
se-	-i -kan
te-	-an

4. Jika karakter-karakter dikembalikan setelah penghapusan prefiks, akan dilakukan proses *stemming* bila diperlukan.

Menurut Adriani dan Asian (2007), Algoritma *stemming* Bahasa Indonesia yang dikembangkan oleh Nazief & Adriani memiliki tahapan sebagai berikut :

1. Cari kata yang akan diubah menjadi kata dasar dalam kamus. Jika ditemukan maka kata tersebut dianggap sebagai *root word* dan algoritma selesai.
2. Hapus *Infection Suffiex* (“lah”, “-kah”, “-ku”, “-mu”, atau “-nya”). Jika imbuhan yang dihapus berupa *Particles* (“--lah”, “-kah”, dan “-tah”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”).
3. Hapus *Derivation Suffliex* (“-i”, “-an” atau “-kan”). Jika kata tersebut kemudian ditemukan di kamus, maka algoritma berhenti. Jika tidak terdapat dalam kamus, maka lanjut ke langkah 3a.
 - a. Jika telah terjadi penghapusan imbuhan “-an” dan huruf terakhir dari kata tersebut adalah “-k” juga tidak dihapus. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran *Derivation Suffiex* yang telah dihapus di kembalikan dan lanjut ke langkah 4.
4. Hapus *Derivation Suffiex*. Jika pada langkah 3 ada suffiks yang dihapus maka lanjut kelangkah 4a, jika tidak maka lanjut ke langkah 4b.

- a. Ada beberapa hal yang akan menyebabkan algoritma berhenti: Jika prefiks kata membentuk kombinasi prefiks dan sufiks yang tidak diperbolehkan sesuai pada Tabel 2.1, jika imbuhan awal yang ditemukan identik dengan imbuhan awal yang sebelumnya, atau telah dilakukan pemotongan imbuhan sebanyak 3 kali.
 - b. Tentukan tipe awalan. Imbuhan awal terdiri dari 2 jenis, awalnya sederhana (“di-”, “ke-”, atau “se-”) yang dapat dihapus secara langsung dan awalnya kompleks (“te-”, “me-”, “be”, “pe-”) adalah tipe-tipe awalan yang dapat bermorfologi sesuai dengan kata dasar yang mengikutinya.
5. Apabila setelah langkah 4 kata dasar masih belum ditemukan, maka akan dilakukan proses stemming dengan menambahkan karakter stemming di awal kata yang dipenggal.
 6. Jika semua langkah gagal, maka input kata yang diuji pada algoritma ini dianggap sebagai kata dasar.

Kelebihan pada algoritma Nazief-Andriani adalah algoritma ini memperhatikan kemungkinan partikel-partikel yang mungkin mengikuti kata yang berimbuhan, sehingga dapat dilihat pada rumus algoritma ini, yaitu dengan adanya penempatan *possesive pronoun* dan juga partikel yang ada pada suatu kata yang berimbuhan. Akhir dari algoritma ini apabila pemotongan semua imbuhan telah berhasil dilakukan dan hasil pemotongan tersebut ada pada kamus, maka algoritma ini dapat dinyatakan berhasil dalam menentukan kata dasar. Dan sebaliknya bahwa pada algoritma Nazief-Andriani ini setelah dilakukan pemotongan pada kata dan tidak terdapat pada kamus yang ada maka kata berimbuhan tersebut telah mengalami pemotongan dan dikembalikan pada keadaan semula.

2.7. Seleksi Fitur

Seleksi fitur merupakan salah satu teknik yang cukup sering digunakan dalam preprocessing teks. Teknik berperan dalam mengurangi jumlah fitur yang terlibat dalam menentukan suatu nilai pada kelas, mengurangi fitur irelevan yang berlebihan dan yang

menyebabkan salah pengertian terhadap kelas target. Tujuan utama dari seleksi fitur ini ialah memilih fitur terbaik dari suatu kumpulan fitur data (Maulida et al., 2016).

2.8. Information Gain

Information gain adalah teknik seleksi fitur yang menggunakan metode *scoring* dalam menentukan nominal ataupun pembobotan atribut kontinu yang didiskritkan menggunakan nilai *entropy*. *Entropy* digunakan dalam mendefinisikan nilai *information gain*. *Entropy* menggambarkan banyaknya informasi yang dibutuhkan dalam melakukan pengkodean pada suatu kelas. *Information gain* dari suatu *term* diukur dengan menghitung jumlah *bit* informasi yang diambil dari prediksi kategori dengan ada atau tidaknya *term* dalam suatu dokumen (Maulida et al., 2016).

Algoritma pada metode ini menggunakan konsep dari *entropy*. Konsep *entropy* yang digunakan untuk mengukur seberapa informatifnya sebuah *node*. Dari nilai *entropy* tersebut kemudian dihitung nilai *information gain* pada masing-masing atribut. Perhitungan nilai *entropy* menggunakan rumus persamaan sebagai berikut :

$$\text{Entropy (S)} = \sum_{i=1}^k (P_i) \log_2(P_i) \quad (2.1)$$

Rumus tersebut merupakan perhitungan *entropy* yang digunakan dalam menentukan seberapa informatif atribut tersebut.

Keterangannya :

1. s : Himpunan kasus
2. k : Jumlah partisi
3. p_i : Jumlah kasus pada partisi ke- i

Pada rumus 2.2 Nilai v merupakan semua nilai yang mungkin dari atribut A . S_v merupakan nilai *subset* dari S dimana atribut A bernilai V .

$$\text{Entropy (S)} = \sum_{i=1}^k \left(\frac{S_v}{v} \times \text{Entropy (S}_v) \right) \quad (2.2)$$

Pada rumus 2.3 merupakan rumus yang digunakan untuk melakukan perhitungan pada *information gain* setelah melakukan perhitungan pada *entropy*.

$$\text{GAIN}(S, A) = \text{Entropy}(S) - \text{Entropy}(S, A) \quad (2.3)$$

Keterangannya :

1. s : Himpunan kasus
2. n : Jumlah partisi atribut A
3. $|S_i|$: Jumlah kasus pada partisi ke- i
4. $|S|$: Jumlah kasus dalam S

Dengan nilai *gain* (S, A) merupakan nilai *information gain*. *Entropy* (S) adalah nilai *entropy* sebelum dipisah. *Entropy* (S, A) merupakan nilai *entropy* setelah dilakukan pemisahan. Besarnya nilai *information gain* menunjukkan seberapa besar pengaruh suatu atribut terhadap pengklasifikasian data (Rasywir & Purwarianti, 2015).

2.9. Maximum Entropy

Maximum entropy adalah salah satu metode klasifikasi dokumen dengan menggunakan nilai probabilitas dalam melakukan proses klasifikasinya. Prinsip pada *maximum entropy* menyatakan bahwa probabilitas yang paling mewakili keadaan merupakan nilai *entropy* terbesar. *Maximum entropy* merupakan teknik umum untuk mengestimasi probabilitas distribusi ketika tidak ada yang diketahui distribusinya diuniformkan yaitu memiliki *maximum entropy*, dalam mengklasifikasikan teks metode ini mengestimasi distribusi label dalam dokumen (Susilo, 2013).

Metode *maximum entropy* bertujuan dalam memaksimalkan *entropy* dalam sistem dengan cara melakukan prediksi distribusi kondisi dari label pada setiap kelas. Metode *maximum entropy* dapat memodelkan sesuatu dengan informasi yang terbatas dan dengan prinsipnya yang hanya memodelkan sesuatu dari informasi yang diketahui saja.

Nilai *entropy* yang digunakan untuk mendapatkan nilai dari *maximum entropy*. Rumus pada metode *maximum entropy* pada persamaan adalah sebagai berikut :

$$Entropy(X) = - \sum_{i=1}^n P(X_i) \log_2(PX_i) \quad (2.4)$$

Keterangan :

1. $Entropy(X)$ = Himpunan informasi dari suatu kejadian x
2. $P(X)$ = Probabilitas dari kemunculan kejadian x

Pada rumus 2.5 merupakan rumus untuk mencari nilai kepangkatan pada *term*.

$$Z(d) = \sum \exp(\sum \lambda_i f_i(d, c)) \quad (2.5)$$

Keterangan :

1. $Z(d)$ = Nilai kepangkatan pada term d
2. λ = Parameter
3. $f_i(d, c)$ = Probabilitas kemunculan term d di kelas c

Proses klasifikasi pada metode *maximum entropy* hanya menggunakan informasi kemunculan dari suatu fitur dalam sebuah dokumen (Anggraeni, 2008). Secara garis besar, metode *maximum entropy* akan mencari probabilitas yang paling sama dengan menggunakan asumsi minimal. Pada kasus klasifikasi teks *maximum entropy* menggunakan rumus pada persamaan sebagai berikut :

$$P(c|d) = \frac{1}{Z(d)} \exp(\sum \lambda_i f_i(d, c)) \quad (2.6)$$

Keterangan :

1. $P(c|d)$ = Probabilitas kemunculan term d di kelas c
2. $Z(d)$ = Derajat kepangkatan term d
3. λ = Parameter
4. $f_i(d, c)$ = Probabilitas kemunculan term d di kelas c

2.10. Pengujian Klasifikasi

Teknik yang digunakan dalam mencari nilai akurasi disebut sebagai *confusion matrix multiclass*. *Confusion matrix* menggunakan tabel dalam menjelaskan jumlah data uji yang benar dan yang salah. Tabel 2.2 menampilkan *confusion matrix*

Tabel 2.2. *Confusion Matrix*

Prediksi	AKTUAL		
	Positif	Negatif	Netral
Positif	TPos	FPosNeg	FPosNet
Negatif	FNegPos	TNet	FNegNet
Netral	FNetPos	FNetNeg	TNet

Keterangan:

1. TP (*true positive*) dan TN (*true negative*) : Klasifikasi kelas yang benar
2. FP (*false positive*) : Nilai yang di prediksi ke dalam kelas positif namun seharusnya berada pada kelas negatif

Precision merupakan jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks yang terpilih oleh sistem. *Precision* menunjukkan nilai benar positif dibagi seluruh nilai yang diperkirakan positif (Powers & Ailab, 2011). Rumus *precision* adalah sebagai berikut :

$$Precision = \frac{TPos}{TPos + FPosNeg + FPosNet} \quad (2.7)$$

Recall menunjukkan perbandingan antara nilai benar positif dengan seluruh data yang sebenarnya (Powers & Ailab, 2011). Rumus *recall* adalah sebagai berikut :

$$Recall = \frac{TPos}{TPos + FNegPos + FNetPos} \quad (2.8)$$

Nilai prediksi benar dibandingkan dengan seluruh nilai data akan menghasilkan nilai akurasi (Powers & Ailab, 2011). Rumus akurasi adalah sebagai berikut :

$$Accuracy = \frac{TPos + TNeg + TNet}{TPos + FPosNeg + FPosNet + FNegPos + TNeg + FNegNet + FNetPos + FNetNeg + TNet} \quad (2.9)$$

Keterangan:

1. TPos = *True Positif*
2. FPosNeg = *False Positif Negatif*
3. FPosNet = *False Positif Netral*
4. FNegPos = *False Negatif Positif*
5. TNeg = *True Negatif*
6. FNetNeg = *False Netral Negatif*
7. FNetPos = *False Netral Positif*
8. FNetNeg = *False Netral Negatif*
9. TNet = *True Netral*

2.11. Penelitian Sebelumnya

Adapun beberapa penelitian yang telah dilakukan dan dijadikan referensi dalam penelitian ini dapat dilihat pada tabel berikut:

1. Penelitian dengan judul “Klasifikasi Hoaks Menggunakan Metode *Maximum Entropy* Dengan Seleksi Fitur *Information Gain*” (Alroy et al., 2019). Data pada penelitian ini mengambil data yang sudah pernah digunakan pada penelitian yang dilakukan oleh Rahutomo (2017). Terdapat 600 artikel yang terdiri dari 372 artikel berlabel fakta dan 228 artikel yang berlabel hoaks. Pada penelitian ini menghasilkan akurasi tertinggi sebesar 0,8 dengan seleksi fitur *information gain (threshold = 50%)*, *precision* 1, *recall* 0,8, dan *f-measure* 0,8. Adapun perbedaan dengan penelitian ini ialah data belum pernah digunakan, jumlah data yang berbeda dan terdapat label sentimen positif, negatif dan netral.
2. Penelitian dengan judul “Analisis Sentimen Pemilihan Presiden 2019 pada Twitter Menggunakan Metode *Maximum Entropy*” (Sabily et al., 2019). Pada penelitian ini, peneliti melakukan pengambilan data opini masyarakat tentang pemilihan presiden pada

tahun 2019 melalui setiap *tweet* terkait yang ada pada aplikasi *twitter* dan kemudian akan dilakukan analisis terhadap pemilihan presiden tahun 2019, dengan melakukan pengambilan data teks pada komentar yang ada pada aplikasi *twitter*. Data latih yang digunakan adalah sebanyak 300 data yang mana data tersebut memiliki jumlah 150 *tweet* positif dan 150 *tweet* negatif. Untuk pengujian dilakukan menggunakan data uji sebanyak 120 data. Hasil evaluasi pada penelitian ini menghasilkan nilai akurasi *macro* sebesar 89,16% dengan nilai *precision* dan *recall* sebesar 100% dan juga nilai *f-measure* sebesar 94,27%. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan penambahan seleksi fitur *information gain*.

3. Penelitian dengan judul “Klasifikasi Teks Pengaduan Suara Warga Kabupaten Pasuruan Menggunakan Metode *Maximum Entropy*” (Rini et al., 2019). Data yang digunakan dalam penelitian ini adalah data pada pengaduan warga Kabupaten Pasuruan dengan total jumlah 200 data pengaduan. Data terbagi dari Dinas Pekerjaan Umum dan Bina Marga (DPUBM) sebanyak 100 data, Dinas Kependudukan dan Catatan Sipil (Dispendukcapil) sebanyak 400 data, Dinas Komunikasi dan Informatika (Diskominfo) sebanyak 32 data, dan Badan Kepegawaian Daerah sebanyak 28 data. Data yang ada kemudian dibagi menjadi data latih yang terdiri dari 150 pengaduan dan data uji yang terdiri dari 50 pengaduan. Hasil evaluasi pada penelitian ini memiliki data latih yang seimbang dan menghasilkan data hasil lebih baik daripada hasil evaluasi dengan data latih yang tidak seimbang dengan akurasi sebesar 89,27%, *precision* 92,49%, *recall* 89,27%, dan *f-measure* 89,44%. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan penambahan seleksi fitur *information gain*.

4. Penelitian yang berjudul “Analisis Sentimen Pada Data Ulasan Produk Toko *Online* Dengan Menggunakan Metode *Maximum Entropy*” (Syah et al., 2017). Data yang digunakan pada penelitian ini yaitu *review* produk dari Amazon yang berfokus pada penjualan *cell phones* dan *accessories*. Pada penelitian ini bertujuan untuk membuat suatu sistem yang akan melakukan klasifikasi *review* produk tersebut yang bersifat *review* positif atau *review* negatif. Hasil evaluasi penelitian ini mendapatkan percobaan terbaik dengan akurasi 83% dan *f-measure* 90.074% pada iterasi ke 1000. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, media pengambilan data, penambahan sentimen netral dan penambahan seleksi fitur *information gain*.
5. Penelitian yang berjudul “Penerapan Analisis Sentimen Untuk Menilai Suatu Produk Pada *Twitter* Berbahasa Indonesia Dengan Metode *Naive Bayes Classifier* dan *Information Gain*” (Attabi et al., 2018) . Pengujian yang dilakukan pada penelitian ini memiliki jumlah data sebanyak 200 data, yang terbagi 100 dokumen positif dan 100 dokumen negatif. Hasil yang diperoleh terdapat peningkatan akurasi dengan selisih 4%, nilai rata-rata akurasi jika tidak menggunakan metode *information gain (threshold 0)* sebesar 70%, sedangkan dengan menggunakan metode *information gain (threshold 0,01)* sebesar 74%. Nilai akurasi yang paling tinggi didapatkan pada penelitian ini adalah K1(*threshold 0,02*), K5 (*threshold 0,01*), K6 (*threshold 0,01*) dan K7 (*threshold 0.02 dan 0.08*) dengan nilai presentase sebesar 85% sedangkan pada bagian k tertentu memperoleh titik terendah sebesar 50%. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan perbedaan pada metode yang digunakan.
6. Pada penelitian yang berjudul “Analisis Sentimen Maskapai Penerbangan Menggunakan Metode *Naive Bayes* Dan Seleksi Fitur *Information Gain*” (Negara et al., 2020). Jumlah data yang digunakan pada penelitian ini adalah sebanyak 500 kalimat

positif dan 500 kalimat negatif. Data yang dikumpulkan sebanyak 100 ulasan dari masing-masing maskapai. Hasil pengujian pada penelitian menunjukkan bahwa nilai rata-rata akurasi, *precision*, *recall* setelah melakukan penambahan fitur *information gain* menunjukkan hasil yang lebih baik sebesar 0,865 jika dibandingkan sebelum dilakukan penambahan fitur *information gain* yakni sebesar 0,81. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan perbedaan pada metode yang digunakan.

7. Pada penelitian yang berjudul “Analisis Sentimen Pada Media Sosial Twitter Menggunakan *Naive Bayes Classifier* Terhadap Kata Kunci Kurikulum 2013” (Pamungkas et al., 2015). Pada penelitian ini, peneliti melakukan pengambilan data terhadap komentar-komentar yang ada pada *twitter* mengenai kurikulum 2013, data komentar tersebut akan diklasifikasi menjadi opini yang bersifat positif maupun opini negatif. Penelitian ini menggunakan metode *naive bayes classifier* dalam pengklasifikasian opini-opini tersebut. Data yang digunakan pada penelitian ini sebanyak 3300 data *tweet* tentang yang dilakukan pencarian melalui kata kunci “kurikulum 2013”. Data tersebut diklasifikasi secara manual dan dibagi menjadi 1000 data sentimen positif, negatif dan netral. Untuk proses latih digunakan 3000 data dan 1000 *tweet* pada setiap kategori sentimennya. Hasil pada penelitian ini menunjukkan sebuah sistem yang mengklasifikasi sentimen secara otomatis dengan hasil pengujian 3000 data latih dan 100 *tweet* data uji coba mencapai nilai 91%. Adapun perbedaan dengan penelitian ini ialah pada metode yang digunakan dan adanya penambahan seleksi fitur *information gain*.
8. Penelitian yang berjudul “Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial *Twitter* Menggunakan Metode *K-Nearest Neighbor* dan Pembobotan Jumlah *Retweet*” (Nurjanah et al., 2017). Proses klasifikasi

pada penelitian ini menambahkan fitur pembobotan jumlah *retweet*. Data yang digunakan berupa opini masyarakat terhadap tayangan televisi yang ada pada *twitter* dengan jumlah data sebanyak 400. Dari hasil pengujian akurasi yang diperoleh dengan menggunakan pembobotan tekstual diperoleh sebesar 82,50%, pembobotan non-tekstual sebesar 60% dan penggabungan dari keduanya mendapatkan hasil sebesar 83,33%. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan perbedaan pada metode yang digunakan serta penambahan seleksi fitur *information gain*.

9. Pada penelitian yang berjudul “Studi Komparatif Metode Ekstraksi Fitur Pada Analisis Sentimen Maskapai Penerbangan Menggunakan *Support Vector Machine* dan *Maximum Entropy*” (Cindo et al., 2019). Penelitian menggunakan 171 data bersifat positif dan 172 data bersifat negatif, sistem pada penelitian berhasil memprediksi 162 data opini positif dan 156 opini negatif pada metode *Maximum Entropy*, sedangkan pada metode *Support Vector Machine* berhasil memprediksi 156 opini positif dan 150 opini negatif. Hasil terbaik didapat pada metode *Maximum Entropy* menggunakan semua ekstraksi fitur dengan akurasi sebesar 92,7% dan pada metode *Support Vector Machine* mendapatkan hasil sebesar 89,2%. Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan terdapat penambahan seleksi fitur *information gain*.
10. Penelitian yang berjudul “*Sentiment Analys* Pada Media Sosial Dengan Metode *Naive Bayes Classifier* (Studi Kasus : *Twitter Komisi Pemilihan Umum*)” (Hidayat, 2019). Penelitian melakukan pengambilan data sebanyak 600 data dengan mengambil opini masyarakat melalui akun *twitter* Komisi Pemilihan Umum. Pengujian pada penelitian ini mengambil 70% dari 600 data sebagai data latih dan 30% sebagai data uji. Hasil pada penelitian ini mendapatkan akurasi sebesar 85,5%, *recall* 81%, dan presisi sebesar 76%.

Adapun perbedaan dengan penelitian ini ialah jumlah data yang berbeda, penambahan sentimen netral pada *tweet* dan perbedaan pada metode yang digunakan serta terdapat penambahan seleksi fitur *information gain*.

Tabel 2.2. Penelitian Sebelumnya

No	Penulis	Judul	Intisari
1	Albert Bill Alroy, Putra Pandu Adikara, Sigit Adinugroho (2019)	Klasifikasi Hoaks Menggunakan Metode Maximum Entropy Dengan Seleksi Fitur Information Gain	Data pada penelitian ini sebelumnya sudah pernah digunakan pada penelitian yang dilakukan oleh Rahutomo (2017). Terdapat 600 artikel yang terdiri dari 372 artikel berlabel fakta dan 228 artikel yang berlabel hoaks. Pada penelitian ini menghasilkan akurasi tertinggi sebesar 0,8 dengan seleksi fitur <i>information gain</i> (threshold = 50%), precision 1, recall 0,8, dan <i>f-measure</i> 0,8.
2	Alvandi Fadhil Sabily, Putra Pandu Adikara, Mochammad Ali Fauzi (2019)	Analisis Sentimen Pemilihan Presiden 2019 pada Twitter Menggunakan Metode <i>Maximum Entropy</i>	Pada penelitian ini, peneliti melakukan pengambilan data opini masyarakat tentang pemilihan presiden pada tahun 2019 melalui setiap <i>tweet</i> terkait yang ada pada aplikasi <i>twitter</i> . Data latih yang digunakan adalah sebanyak 300 data yang mana data tersebut memiliki jumlah 150 <i>tweet</i> positif dan 150 <i>tweet</i> negatif. Untuk pengujian dilakukan menggunakan data uji sebanyak 120 data. Hasil evaluasi pada penelitian ini menghasilkan nilai akurasi <i>macro</i> sebesar 89,16% dengan nilai <i>precision</i> dan <i>recall</i> sebesar 100% dan juga nilai <i>f-measure</i> sebesar 94,27%.
3	Mayang Panca Rini, Putra Pandu Adikara, Sigit Adinugroho (2019)	Klasifikasi Teks Pengaduan Suara Warga Kabupaten Pasuruan	Data yang digunakan dalam penelitian ini adalah data pada pengaduan warga Kabupaten Pasuruan dengan total jumlah 200 data pengaduan. Data terbagi dari Dinas Pekerjaan Umum dan Bina Marga

Tabel 2.3. Lanjutan Penelitian Sebelumnya

No	Penulis	Judul	Intisari
		Menggunakan Metode <i>Maximum Entropy</i>	(DPUBM) sebanyak 100 data, Dinas Kependudukan dan Catatan Sipil (Dispendukcapil) sebanyak 400 data, Dinas Komunikasi dan Informatika (Diskominfo) sebanyak 32 data, dan Badan Kepegawaian Daerah sebanyak 28 data. Data yang ada kemudian dibagi menjadi data latih yang terdiri dari 150 pengaduan dan data uji yang terdiri dari 50 pengaduan. Hasil evaluasi pada penelitian ini memiliki data latih yang seimbang dan menghasilkan data hasil lebih baik daripada hasil evaluasi dengan data latih yang tidak seimbang dengan akurasi sebesar 89,27%, <i>precision</i> 92,49%, <i>recall</i> 89,27%, dan <i>f-measure</i> 89,44%.
4	Alvi Pranandha Syah, Adiwijaya, Said Al Faraby (2017)	Analisis Sentimen Pada Data Ulasan Produk Toko <i>Online</i> Dengan Menggunakan Metode <i>Maximum Entropy</i>	Data yang digunakan pada penelitian ini yaitu <i>review</i> produk dari Amazon yang berfokus pada penjualan <i>cell phones</i> dan <i>accessories</i> . Pada penelitian ini bertujuan untuk membuat suatu sistem yang akan melakukan klasifikasi <i>review</i> produk tersebut yang bersifat <i>review</i> positif atau <i>review</i> negatif. Hasil evaluasi penelitian ini mendapatkan percobaan terbaik dengan akurasi 83% dan <i>f-measure</i> 90.074% pada iterasi ke 1000.
5	Ahmad Wildan Attabi Lailil Muflikhah Mochammad Ali Fauzi (2018)	Penerapan Analisis Sentimen Untuk Menilai Suatu Produk Pada <i>Twitter</i> Berbahasa Indonesia Dengan Metode <i>Naive Bayes Classifier</i> dan <i>Information Gain</i>	Pengujian yang dilakukan pada penelitian ini memiliki jumlah data sebanyak 200 data. Hasil yang diperoleh terdapat peningkatan akurasi dengan selisih 4%, nilai rata-rata akurasi jika tidak menggunakan metode <i>Information Gain</i> (<i>threshold</i> 0) sebesar 70%, sedangkan

Tabel 2.4. Lanjutan Penelitian Sebelumnya

No	Penulis	Judul	Intisari
		<i>Bayes Classifier</i> dan <i>Information Gain</i>	dengan <i>Gain</i> (<i>threshold</i> 0,01) sebesar 74%. Nilai akurasi yang paling tinggi didapatkan pada penelitian ini adalah K1(<i>threshold</i> 0,02), K5 (<i>threshold</i> 0,01), K6 (<i>threshold</i> 0,01) dan K7 (<i>threshold</i> 0.02 dan 0.08) dengan nilai presentase sebesar 85% sedangkan pada bagian k tertentu memperoleh titik terendah sebesar 50%.
6	Arif Bijaksana Putra Negara, Hafiz Muhardi Indira, Melinda Putri (2020)	Analisis Sentimen Maskapai Penerbangan Menggunakan Metode <i>Naive Bayes</i> Dan Seleksi Fitur <i>Information Gain</i>	Jumlah data yang digunakan pada penelitian ini adalah sebanyak 500 kalimat positif dan 500 kalimat negatif. Data yang dikumpulkan sebanyak 100 ulasan dari masing-masing maskapai. Hasil pengujian pada penelitian menunjukkan bahwa nilai rata-rata akurasi, <i>precision</i> , <i>recall</i> setelah melakukan penambahan fitur <i>information gain</i> menunjukkan hasil yang lebih baik sebesar 0,865 jika dibandingkan sebelum dilakukan penambahan fitur <i>information gain</i> yakni sebesar 0,81
7.	Dyarsa Singgih Pamungkas, Noor Ageng Setiyanto, Erlin Dolphina (2015)	Analisis Sentimen Pada Media Sosial Twitter Menggunakan <i>Naive Bayes Classifier</i> Terhadap Kata Kunci Kurikulum 2013	Data yang digunakan pada penelitian ini sebanyak 3300 data <i>tweet</i> tentang yang dilakukan pencarian melalui kata kunci “ diklasifikasi secara manual dan dibagi menjadi 1000 data sentimen positif, negatif dan netral. Untuk proses latih digunakan 3000 data dan 1000 <i>tweet</i> pada setiap kategori sentimennya. Hasil pada penelitian ini menunjukkan sebuah sistem yang mengklasifikasi sentimen secara otomatis dengan hasil pengujian 3000 data latih dan 100 <i>tweet</i> data uji coba mencapai nilai 91%.

Tabel 2.5. Lanjutan Penelitian Sebelumnya

No	Penulis	Judul	Intisari
8	Winda Estu Nurjanah, Rizal Setya Perdana, Mochammad Ali Fauzi (2017)	Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial <i>Twitter</i> Menggunakan Metode <i>K-Nearest Neighbor</i> dan Pembobotan Jumlah <i>Retweet</i>	Proses klasifikasi pada penelitian ini menambahkan fitur pembobotan jumlah <i>retweet</i> . Data yang digunakan berupa opini masyarakat terhadap tayangan televisi yang ada pada <i>twitter</i> dengan jumlah data sebanyak 400. Dari hasil pengujian akurasi yang diperoleh dengan menggunakan pembobotan tekstual diperoleh sebesar 82,50%, pembobotan non-tekstual sebesar 60% dan penggabungan dari keduanya mendapatkan hasil sebesar 83,33%.
9	Mona Cindo, Dian Palupi Rini, Ermatita (2019)	Studi Komparatif Metode Ekstraksi Fitur Pada Analisis Sentimen Maskapai Penerbangan Menggunakan <i>Support Vector Machine</i> dan <i>Maximum Entropy</i>	Penelitian menggunakan 171 data bersifat positif dan 172 data bersifat negatif, sistem pada penelitian berhasil memprediksi 162 data opini positif dan 156 opini negatif pada metode <i>Maximum Entropy</i> , sedangkan pada metode <i>Support Vector Machine</i> berhasil memprediksi 156 opini positif dan 150 opini negatif. Hasil terbaik didapat pada metode <i>Maximum Entropy</i> menggunakan semua ekstraksi fitur dengan akurasi sebesar 92,7% dan pada metode <i>Support Vector Machine</i> mendapatkan hasil sebesar 89,2%.
10	Randa Hidayat (2019)	<i>Sentiment Analys</i> Pada Media Sosial Dengan Metode <i>Naive Bayes</i>	Penelitian melakukan pengambilan data sebanyak 600 data dengan mengambil opini masyarakat melalui akun <i>twitter</i> Komisi Pemilihan Umum. Pengujian pada penelitian ini mengambil 70% dari 600 data sebagai data latih dan 30% sebagai data uji. Hasil pada penelitian ini mendapatkan akurasi sebesar 85,5%, <i>recall</i> 81%, dan presisi sebesar 76%.

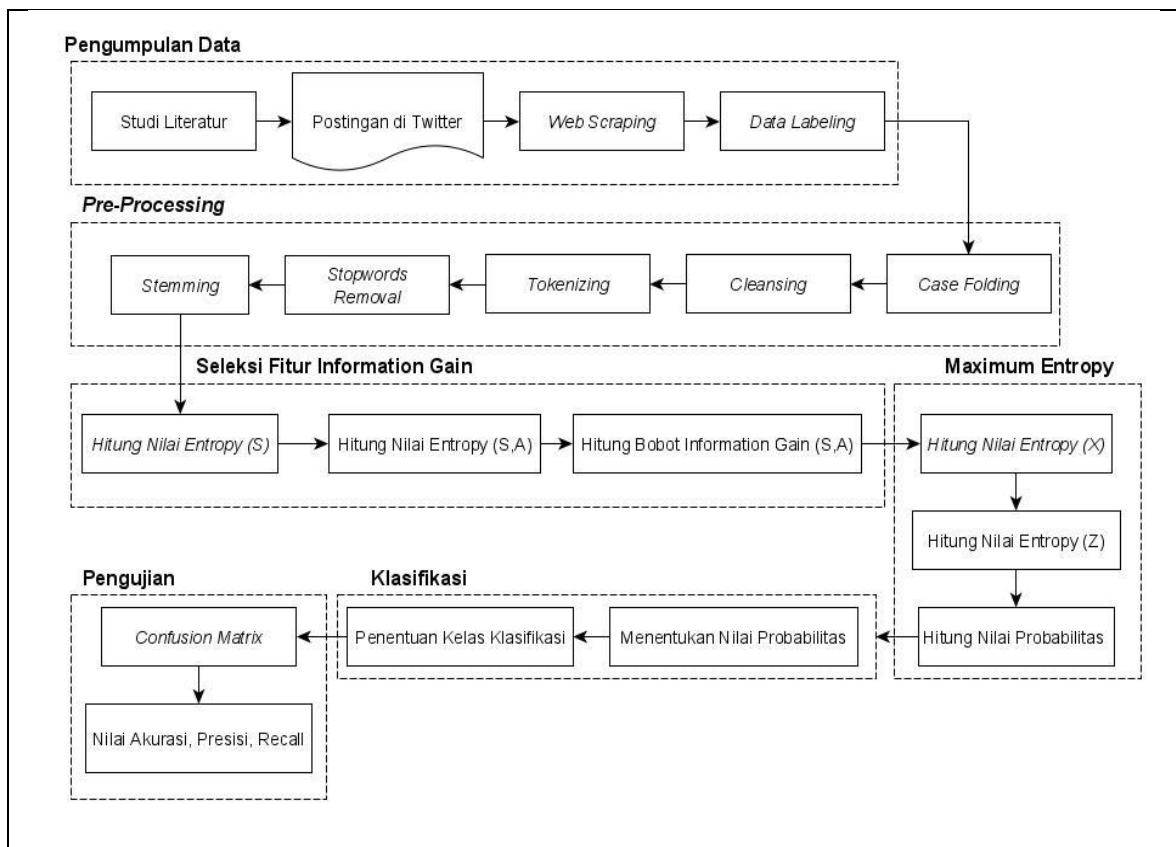
BAB III

METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM

3.1. Metodologi Penelitian dan Pengembangan Sistem

Penelitian ini bersifat kuantitatif, yaitu sebuah proses dalam menemukan suatu ilmu pengetahuan melalui data berupa angka sebagai alat data untuk di analisis keterangannya mengenai yang akan diketahui (Kasiram (2008: 149) dalam bukunya Metodologi Penelitian Kualitatif dan Kuantitatif).

Data yang digunakan memiliki beberapa aspek seperti studi literatur hingga pengambilan data. Setelah permasalahan ditemukan, maka akan dikaji dengan studi literatur yang sesuai dengan permasalahan yang ada. Selanjutnya akan dilakukan pengambilan data sekunder dari media sosial twitter berupa *tweet* mengenai kuliah online untuk dilakukan pengujian.



Gambar 3.1. Metodologi Penelitian

3.1.1. Studi Literatur

Studi literatur merupakan sebuah teknik pengumpulan data dan informasi dari penelitian sebelumnya. Sumber data ini dapat berupa buku, jurnal, maupun artikel yang memuat sejumlah teori, metode, atau objek yang memiliki kesamaan dengan penelitian penulis, dalam hal ini, literatur yang digunakan adalah yang berhubungan dengan analisis sentimen dengan menggunakan metode *maximum entropy* dengan seleksi fitur *information gain*.

3.1.2. Pengumpulan Data Sekunder

Pengumpulan data dalam penelitian ini diperoleh dengan menggunakan metode *web scraping* pada media sosial *Twitter*. *Web scraping* dapat didefinisikan sebagai proses pengambilan data dari sebuah *website*. Secara umum, ada dua cara yang digunakan untuk melakukan *scaping*, yang pertama melakukan secara manual dan otomatis dengan menggunakan koding, aplikasi atau *extension browser*. Data yang diambil merupakan komentar dari pengguna media sosial *twitter* yang berkaitan dengan perkuliahan *online*.

3.1.3. Analisis Masalah

Tujuan analisis sentimen adalah untuk mengetahui sifat dari opini atau pesan yang disampaikan, apakah opini atau pesan tersebut bersifat positif, negatif atau bahkan netral. Pada penelitian ini dibuat sebuah sistem yang akan menentukan sentimen terhadap opini atau pesan yang disampaikan oleh para pengguna media sosial melalui komentar pada *twitter*. Permasalahan pada analisis sentimen adalah bagaimana melakukan analisis sentimen pada media sosial *twitter* berdasarkan kata kunci atau *keywords* yang terkait dengan opini dari mahasiswa/i tentang pembelajaran *daring* menggunakan metode *maximum entropy* dengan seleksi fitur *information gain*.

3.1.4. Metode Pengembangan Sistem

Metode pengembangan sistem yang akan digunakan dalam penelitian ini yaitu menggunakan metode waterfall. Yang mana metode pengembangan sistem ini menyediakan

dalam pendekatan alur hidup perangkat lunak secara terurut atau linear, dengan tahapan-tahapannya sebagai berikut (Pressman, 2012) :

a. Requirements Analysis

Tahap ini merupakan proses analisa atau pengumpulan data-data yang berkaitan dengan sistem yang akan dibuat.

b. Design

Tahap ini merupakan proses pembangunan struktur data serta perancangan *interface* dari sistem yang akan dibuat.

c. Coding & Testing

Pada tahap ini akan dilakukan penggabungan dari modul-modul yang telah dibuat dan dilakukam pengujian untuk mengetahui apakah ada kesalahan yang terjadi atau tidak.

d. Implementation

Tahap ini merupakan proses dari modul-modul yang sudah diintegrasikan mejadi sistem yang lengkap.

e. Maintenance

Pada tahap ini akan dilakukan pemeliharaan sistem yang telah dibuat dan telah diterapkan, maka tahap ini merupakan tahapan akhir dari metode *waterfall*.

3.2. Analisis Kebutuhan

Analisis kebutuhan bertujuan dalam mendeskripsikan apa yang harus dilakukan oleh sistem. Dalam menyelesaikan beberapa permasalahan yang akan dihadapi, perlu adanya beberapa analisis yang dibutuhkan.

3.2.1. Kebutuhan Data Input

Data *input* merupakan data-data yang akan digunakan sebagai masukan pada sistem yang akan dibangun. Data tersebut akan disimpan dalam database sistem. Data *input* yang akan digunakan dalam sistem ini adalah sebagai berikut :

1. Data Admin

Data *admin* akan digunakan untuk menyimpan data dari *admin* yang berhak dalam mengakses sistem ini. Data *admin* yang diperlukan dalam sistem ini adalah *id user*, *username* dan *password*.

2. Data Komentar

Data Komentar diperoleh melalui pencarian dari *twitter*. Data komentar diperoleh secara manual dengan melakukan pencarian dengan menggunakan *keywords* terkait dalam pembelajaran *daring* seperti belajar *daring* atau kuliah *online*. Data komentar akan dibagi menjadi dua bagian yaitu, data latih dan data uji dengan pembagian data latih 80% dan data uji 20%.

3. Data Stopword

Stopword merupakan kata-kata umum yang sering digunakan pada suatu bahasa, namun kata-kata umum tersebut tidak mempunyai pengaruh dalam klasifikasi teks, seperti : “untuk”, “kapan”, “yang”, “di”, “iya”, “itu”, dan lain sebagainya. Pada sistem ini, kamus stopwords yang akan digunakan berasal dari Nurirwan Saputra yang diunduh di situs <http://hikaruyuuki.lecture.ub.ac.id> (Saputra, 2015)

4. Data Kata Dasar

Data kata dasar didapat melalui kamus bahasa Indonesia *online*. Data kata dasar meliputi *id* kata dan kata.

5. Data Kata Baku

Data kata baku merupakan data yang berisikan kata non baku dan kata yang baku.

6. Data Uji

Data uji merupakan data yang akan digunakan dalam melakukan proses analisis sentimen pada sistem.

3.2.2. Kebutuhan Proses Sistem

1. *Text Preprocessing*

Pre-processing teks merupakan langkah untuk mengganti data yang memiliki struktur menjadi data yang terstruktur. Data terstruktur ini nantinya akan diolah sesuai dengan kebutuhan. Proses *text preprocessing* bertujuan agar data yang telah tersimpan dimengerti oleh sistem pengolahan *text mining* dengan baik. Tujuan utama proses ini adalah mendapatkan data yang siap diproses oleh sistem.

2. Ekstraksi Fitur

Ekstraksi fitur merupakan proses yang bertujuan untuk melakukan perhitungan pada kemunculan suatu *term* atau kata dalam sebuah dokumen.

3. Data Latih dan Data Uji

Proses data latih merupakan proses mencari *keyword* serta probabilitas yang nantinya digunakan pada proses pengujian. Sedangkan proses data uji merupakan proses pengklasifikasian terhadap data yang belum diketahui kategorinya.

3.2.3. Kebutuhan Proses Output

Data *output* merupakan data keluaran dari hasil pada proses yang dilakukan oleh sistem.

Data *output* dari sistem tersebut adalah sebagai berikut :

1. Data pesan atau komentar bersih dari hasil *preprocessing* yang siap untuk dilakukan pengklasifikasian dan diolah untuk menentukan sentimen.
2. Nilai probabilitas positif, probabilitas negatif dan nilai probabilitas netral.

3.2.4. Kebutuhan Perangkat Lunak

Analisis kebutuhan perangkat lunak pada sistem ini merupakan kebutuhan perangkat lunak yang digunakan dalam proses pembangunan perangkat lunak ini. adapun perangkat lunak yang digunakan adalah sebagai berikut :

1. Sistem Operasi Windows 10
2. *Web Server XAMPP*
3. *Code Editor*
4. MySQL
5. *Web Browser*

3.3. Perancangan Sistem

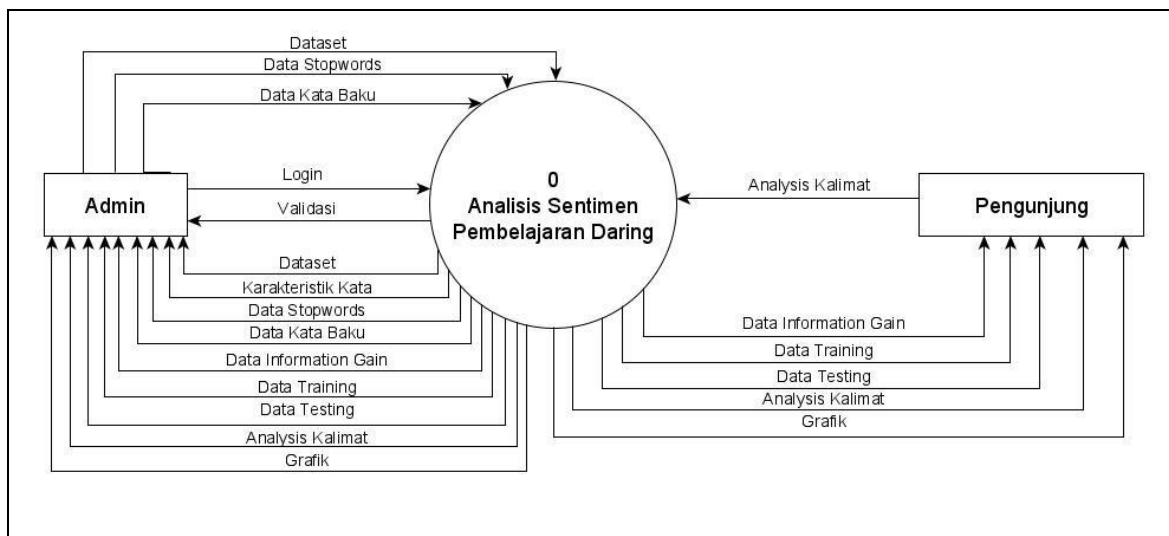
Dalam perancangan sistem ini akan terdiri dari beberapa perancangan, yaitu gambaran umum sistem, *Data Flow Diagram (DFD)*, *Flowchart*, *Entity Relationship Diagram (ERD)*, Relasi Antar Tabel (RAT) serta perancangan struktur menu dan *user interface*. Perancangan sistem ini bertujuan untuk memberikan gambaran tentang aplikasi yang akan dibuat.

3.3.1. Data Flow Diagram (DFD)

Pada perancangan *Data Flow Diagram* yang dilakukan dari DFD level 0 dan DFD level 1 sebagai berikut :

A. DFD Level 0

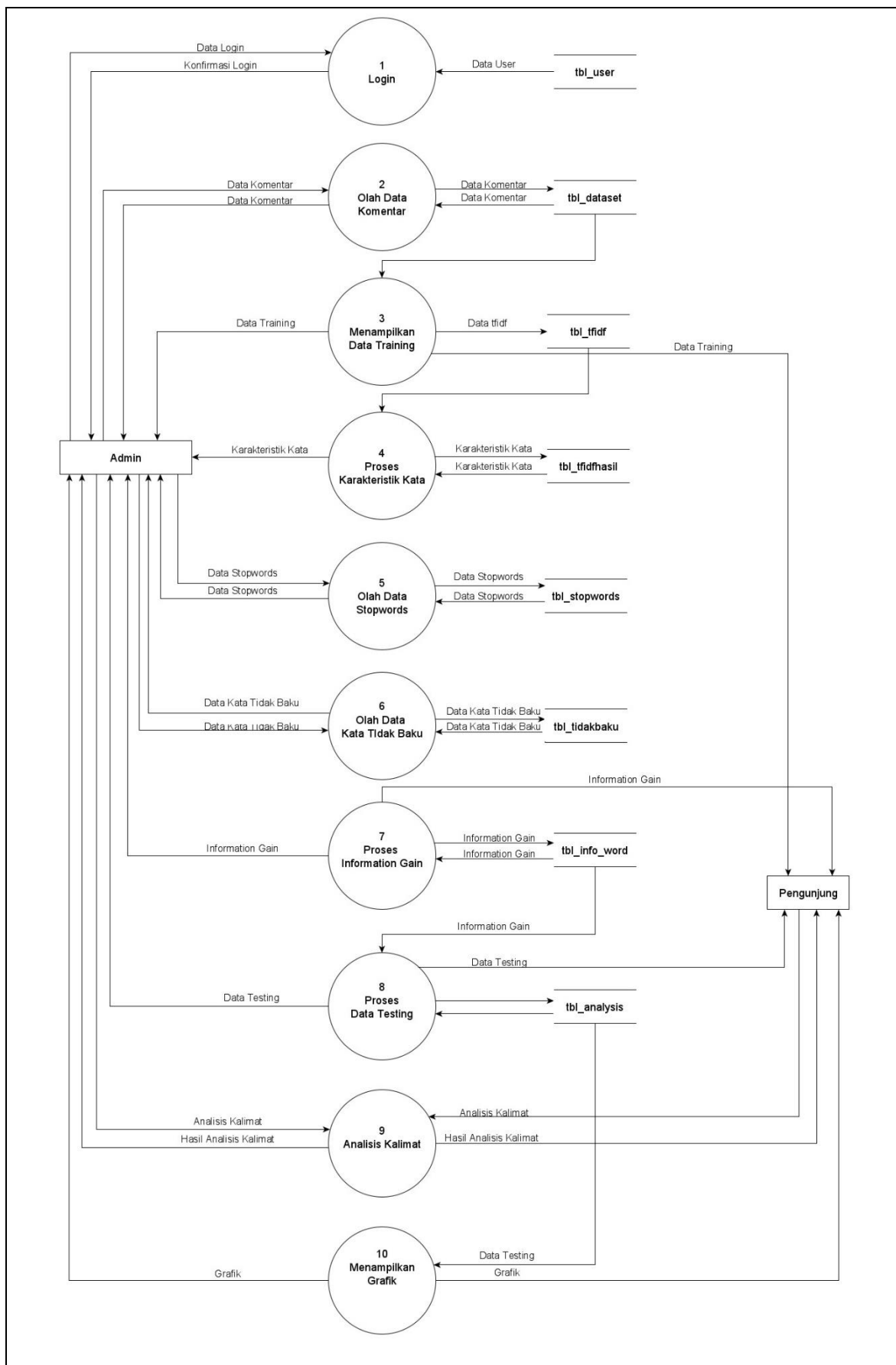
DFD level 0 merupakan gambaran garis besar dari proses sistem yang akan dibangun. Pada DFD level 0, terdapat 2 entitas eksternal yaitu entitas admin dan pengunjung, serta terdapat sebuah proses yaitu analisis sentimen pembelajaran daring. DFD level 0 dapat dilihat pada gambar 3.2 berikut :



Gambar 3.2. DFD Level 0

B. DFD Level 1

DFD level 1 merupakan turunan dari proses DFD level 0. Pada proses DFD level 1 dibagi menjadi 5 proses yaitu, proses dataset, training, information gain, testing dan grafik. pada proses analisis sentimen menggunakan metode maximum entropy dengan seleksi fitur information gain akan di golongankan secara rinci pada flowchart metode pada gambar 3.4, gambar 3.5 dan gambar 3.6. DFD level 1 dapat dilihat pada gambar 3.3 berikut :



Gambar 3.3. DFD Level 1

3.3.2. Flowchart

Pada penelitian ini terdapat beberapa *flowchart* untuk menggambarkan alur proses ataupun tahapan-tahapan yang akan digunakan pada sistem sebagai berikut :

1. *Flowchart Preprocessing*

Preprocessing merupakan proses untuk membersihkan dan mempersiapkan data sebelum dilakukannya proses klasifikasi. Tujuan dari *preprocessing* adalah untuk dan melakukan penyeragaman kata, sehingga kata yang sudah diseragamkan tersebut siap untuk di ekstraksi ke tahap selanjutnya. Secara umum tahapan *preprocessing* sebagai berikut :

a. *Cleansing & Case Folding*

Case Folding merupakan proses untuk menghilangkan *noise* atau menghilangkan elemen-elemen yang tidak diperlukan dalam sebuah dokumen. Tujuan proses adalah untuk meningkatkan dan memperbaiki kualitas data. Proses *cleansing* pada komentar dapat berupa menghilangkan tanda baca, menghilangkan angka-angka, menghilangkan *emoticon* dan mengubah karakter menjadi *lowercase*. Contoh *case folding* dapat dilihat pada tabel 3.1 berikut :

Tabel 3.1. Contoh *Case Folding*

Kalimat	Cleansing & Case Folding
Selamat hari senin, semangat kuliah online nya	selamat hari senin semangat kuliah online nya
Have a nice day-!! jangan lupa sarapan. semangat sekolah/kuliah online nya	have a nice day jangan lupa sarapan semangat sekolahkuliah online nya

b. *Tokenizing*

Proses *tokenizing* merupakan proses pemecahan kata-kata yang ada di suatu kalimat. *Tokenizing* dilakukan dengan memisahkan setiap kata dengan spasi. Contoh dari tokenisasi adalah memecah kalimat “saya pergi ke kantor polisi” menjadi kumpulan kata-kata “saya”, “pergi”, “ke”, “kantor”, “polisi”. Setiap kata yang telah

dipecah disebut sebagai token dan akan dimanfaatkan untuk menghitung frekuensi dari setiap kata dan akan digunakan dalam proses klasifikasi. Contoh dari *tokenizing* dapat dilihat pada tabel 3.2 berikut :

Tabel 3.2. Contoh *Tokenizing*

Kalimat	Tokenizing
selamat hari senin, semangat kuliah online nya	selamat, hari, senin, semangat, kuliah, online, nya
have a nice day-!! jangan lupa sarapan. semangat sekolah/kuliah online nya	have, a, nice, day, jangan, lupa, sarapan, semangat, sekolahkuliah, online, nya

c. *Stopwords Removal*

Stopword merupakan kata-kata umum dan sering digunakan dalam suatu bahasa. Namun kata-kata umum tersebut tidak mempunyai pengaruh dan tidak memiliki nilai pada dokumen teks tersebut. Sehingga kata-kata yang termasuk dalam *stopword* akan dihilangkan, sehingga ukuran pada data juga akan berkurang. Contoh kata-kata yang termasuk dalam *stopword* dalam Bahasa Indonesia adalah “untuk”, “kapan”, “yang”, “di”, “iya”, “itu”, dan lain sebagainya. Kamus *stopwords* yang digunakan berasal dari Nurirwan Saputra yang diunduh di situs <http://hikaruyuuki.lecture.ub.ac.id> (Saputra, 2015). Contoh dari *stopwords removal* dapat dilihat pada tabel 3.3 berikut :

Tabel 3.3. Contoh *Stopwords Removal*

Kalimat	Stopwords Removal
selamat hari senin, semangat kuliah online nya	selamat senin semangat kuliah online nya
have a nice day-!! jangan lupa sarapan. semangat sekolah/kuliah online nya	have a nice day lupa sarapan semangat sekolahkuliah online nya

d. *Stemming*

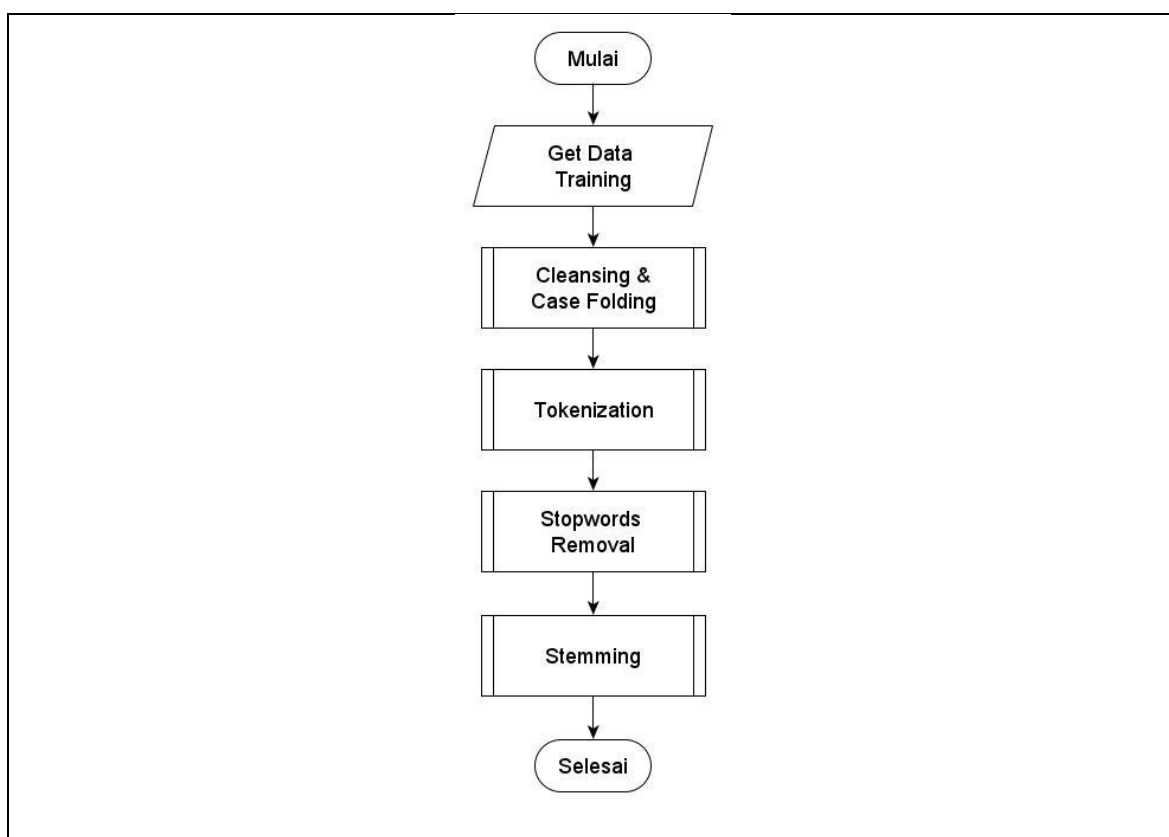
Stemming merupakan proses pemotongan imbuhan pada kata yang memiliki imbuhan, yang dilakukan dengan menggunakan algoritma tertentu untuk

mengembalikan suatu kata kebentuk kata dasarnya (Mauliadi, 2017). Untuk mengubah kata berimbuhan menjadi kata dasar dibutuhkan beberapa aturan yang disebut dalam morfologi.

Tabel 3.4. Contoh *Stemming*

Kalimat	Stemming
selamat hari senin, semangat kuliah online nya	selamat senin semangat kuliah online
have a nice day-!! jangan lupa sarapan. semangat sekolah/kuliah online nya	Have a nice day lupa sarapan semangat sekolahkuliah online

Pada gambar 3.4 berikut terdapat gambar *flowchart* yang merupakan proses atau tahapan dari *preprocessing*



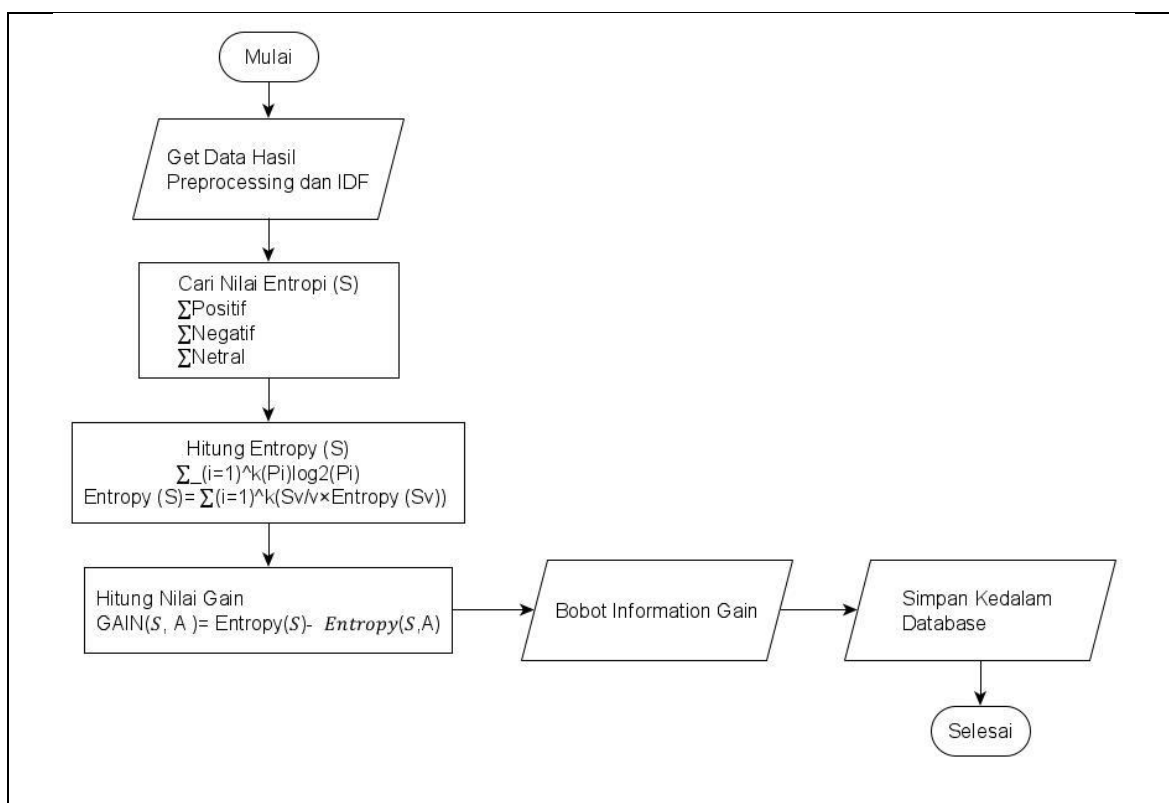
Gambar 3.4. *Flowchart Preprocessing*

2. *Flowchart Seleksi Fitur Information Gain*

Information gain adalah teknik seleksi fitur yang menggunakan metode *scoring* dalam menentukan nominal ataupun pembobotan atribut kontinu yang didiskritkan

menggunakan nilai *entropy*. *Entropy* digunakan dalam mendefinisikan nilai *information gain*. *Entropy* menggambarkan banyaknya informasi yang dibutuhkan dalam melakukan pengkodean pada suatu kelas. *Information gain* dari suatu *term* diukur dengan menghitung jumlah *bit* informasi yang diambil dari prediksi kategori dengan ada atau tidaknya *term* dalam suatu dokumen (Maulida et al., 2016)

Algoritma pada metode ini menggunakan konsep dari *entropy*. Konsep *entropy* yang digunakan untuk mengukur seberapa informatifnya sebuah *node*. Dari nilai *entropy* tersebut kemudian dihitung nilai *information gain* pada masing-masing atribut. Pada gambar 3.5 merupakan flowchart untuk menghitung nilai gain dari sistem ini.



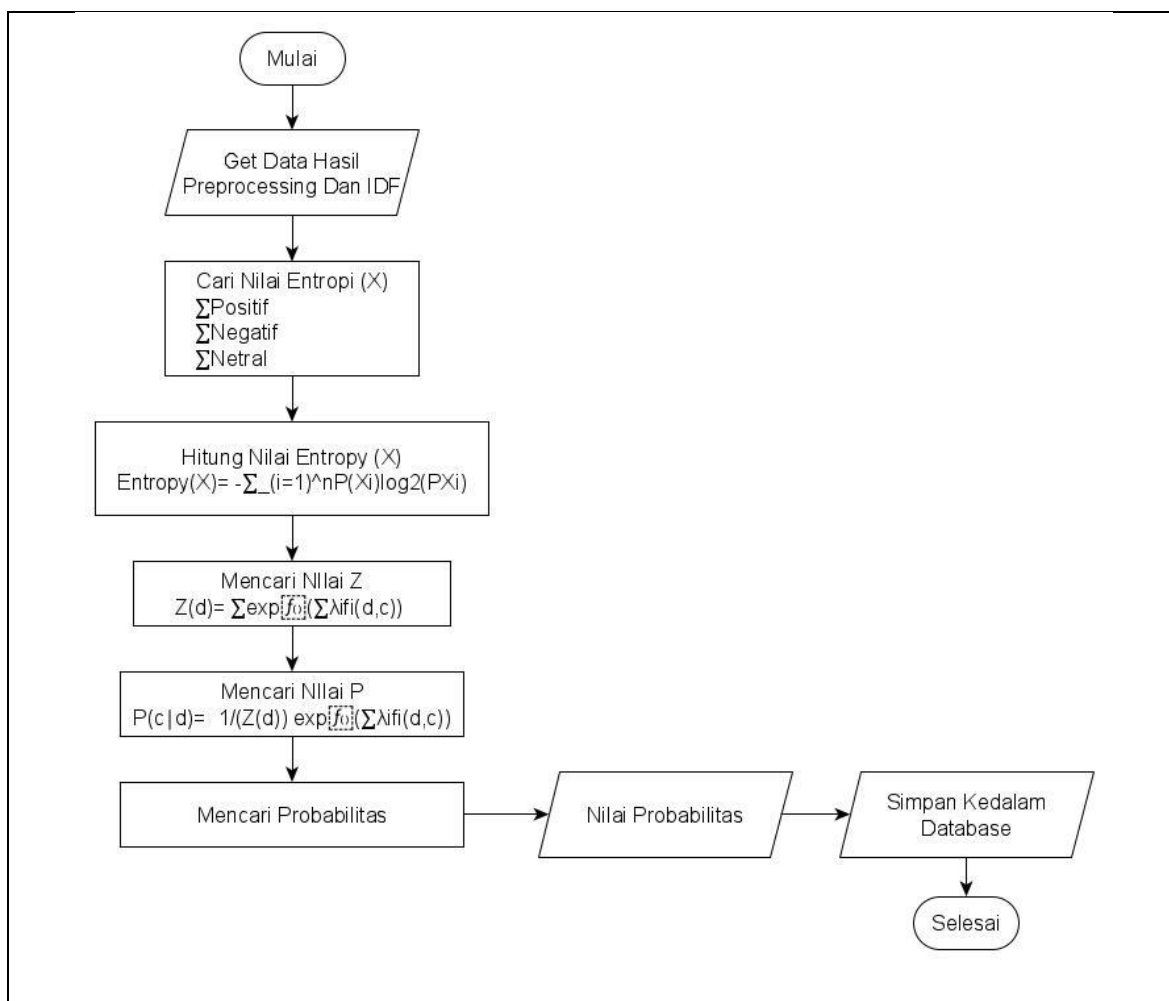
Gambar 3.5. Flowchart Information Gain

3. Flowchart Metode Maximum Entropy

Maximum entropy merupakan teknik umum untuk mengestimasi probabilitas distribusi ketika tidak ada yang diketahui distribusinya di *uniform* kan yaitu memiliki

maximum entropy, dalam mengklasifikasikan teks metode ini mengestimasi distribusi label dalam dokumen (Susilo, 2013).

Metode *maximum entropy* dapat memodelkan sesuatu dengan informasi yang terbatas dan dengan prinsipnya yang hanya memodelkan sesuatu dari informasi yang diketahui saja. Pada gambar 3.6 merupakan flowchart untuk menghitung nilai probabilitas dalam menentukan klasifikasi pada data.



Gambar 3.6. Flowchart Maximum Entropy

3.4. Seleksi Fitur Information Gain

Pada pembahasan ini merupakan contoh perhitungan *information gain* secara manual. Dengan cara menghitung bobot sesuai rumus 2.3 pada bab 2. Sebelum masuk ke contoh perhitungan dengan fitur *Information gain*, pada tabel 3.5 terdapat total *dataset*, kemudian

pada tabel 3.6 merupakan contoh data komentar dengan sentimen manual, dan tabel 3.7 merupakan data kata dari komentar yang sudah melalui tahap *text preprocessing*.

Tabel 3.5. Total Dataset

Sentimen	Total Sentimen
Positif	4
Negatif	3
Netral	3
Total Data	10

Tabel 3.6. Data Komentar

No	Komentar	Label
1	kuliah online adu strategi adu kecerdasan	Positif
2	jangan lupa sarapan. semangat sekolah/kuliah online nya	Positif
3	selamat hari senin, semangat kuliah online nya	Positif
4	Buat yg kerja,daring or kuliah semangatt yaa!	Positif
5	Kuliah daring bikin mood ambyar	Negatif
6	Kuliah daring sangat tidak recommended buat gua	Negatif
7	Dari kuliah online kita tidak belajar'	Negatif
8	mari kuliah online	Netral
9	Kuliah online kok gini amat yaaaak ?	Netral
10	kuliah offline tapi ujiannya online di rumah	Netral

Maka data komentar akan dipecah menjadi kata untuk dihitung nilai *entropy* dan *gain*

seperti pada tabel 3.7 sebagai berikut :

Tabel 3.7. Data Term

Kata (Term)	Positif	Negatif	Netral
Kuliah	4	3	3
Online	3	1	3
Adu	1	0	0
Strategi	1	0	0
Kecerdasan	1	0	0
Jangan	1	0	0
Lupa	1	0	0
Sarapan	1	0	0
Semangat	3	0	0
Sekolah	1	0	0
Selamat	1	0	0
Hari	1	0	0
Senin	1	0	0
Buat	1	1	0
Yang	1	0	0
Kerja	1	0	0

Tabel 3.8. Lanjutan Data *Term*

Kata (<i>Term</i>)	Positif	Negatif	Netral
Daring	1	2	0
Or	1	0	0
Yaa	1	0	0
Bikin	0	1	0
Mood	0	1	0
Ambyar	0	1	0
Sangat	0	1	0
Tidak	0	2	0
Recomended	0	1	0
Gua	0	1	0
Dari	0	1	0
Kita	0	1	0
Tidak	0	1	0
Belajar	0	1	0
Mari	0	0	1
Kok	0	0	1
Gini	0	0	1
Amaat	0	0	1
Yaak	0	0	1
Offline	0	0	1
Ujian	0	0	1
Di	0	0	1
Rumah	0	0	1

Pada kasus perhitungan bobot *information gain* akan diambil kata kuliah dengan menggunakan rumus persamaan 2.1 sebagai berikut :

$$\text{Entropy (Set)} = -\left(\frac{4}{10}\right) \log_2 \left(\frac{4}{10}\right) + \left(\frac{3}{10}\right) \log_2 \left(\frac{3}{10}\right) + \left(\frac{3}{10}\right) \log_2 \left(\frac{3}{10}\right) = 1,570950594$$

Tahap selanjutnya ambil contoh pada kata kuliah yang memiliki 3 kelas sentimen yaitu positif, negatif dan netral, sehingga dapat dilakukan perhitungan untuk mencari *entropy* dari kata kuliah, dengan menggunakan rumus persamaan 2.1 sebagai berikut :

$$\text{Entropy (Positif)} = -\left(\frac{4}{10}\right) \times \log_2 \left(\frac{4}{10}\right) = 0,528771238$$

$$\text{Entropy (Negatif)} = -\left(\frac{3}{10}\right) \times \log_2 \left(\frac{3}{10}\right) = 0,521089678$$

$$\text{Entropy (Netral)} = -\left(\frac{3}{10}\right) \times \log_2 \left(\frac{3}{10}\right) = 0,521089678$$

Selanjutnya akan dilakukan perhitungan untuk menemukan nilai *entropy* dan *gain* dari kata kuliah dengan rumus persamaan 2.2 sebagai berikut :

$$\text{Entropy (Positif)} = \left(\frac{3}{10}\right) \times 0,528771238 = 0,211508495$$

$$\text{Entropy (Negatif)} = \left(\frac{14}{80}\right) \times 0,521089678 = 0,156326903$$

$$\text{Entropy (Netral)} = \left(\frac{43}{80}\right) \times 0,521089678 = 0,156326903$$

$$\text{Entropy (S,A)} = 0,211508495 + 0,156326903 + 0,156326903 = 0,524162302$$

Langkah terakhir untuk mencari nilai dari *information gain* dari kata kuliah dengan menggunakan rumus 2.3 sebagai berikut :

$$\text{Gain (S,A)} = 1,570950594 - 0,524162302 = 1,046788292$$

Dengan bobot *information gain* tersebut pada setiap kata akan dilakukan perankingan dan hasilnya akan didapatkan fitur terbaik. Dalam keperluan klasifikasi data *testing* yang nantinya akan diambil fitur dengan akurasi tertinggi.

3.5. Metode Maximum Entropy

Pada pembahasan ini merupakan contoh perhitungan *maximum entropy* secara manual. Sebagai contoh perhitungan akan menggunakan kata kuliah sebagai contoh yang dapat dilihat dari tabel 3.7 yang sudah berisikan data *term* dengan menggunakan rumus persamaan 2.4. Kata kuliah yang memiliki 3 kelas sentimen yaitu positif, negatif dan netral, dengan rumus persamaan 2.4 akan dilakukan perhitungan untuk mencari nilai *entropy* setiap kelas dari kata kuliah sebagai berikut :

$$\text{Entropy (Positif)} = -\left(\frac{4}{10}\right) \times \log_2\left(\frac{4}{10}\right) = 0,528771238$$

$$\text{Entropy (Negatif)} = -\left(\frac{3}{10}\right) \times \log_2\left(\frac{3}{10}\right) = 0,521089678$$

$$\text{Entropy (Netral)} = -\left(\frac{3}{10}\right) \times \log_2\left(\frac{3}{10}\right) = 0,521089678$$

Gambaran dari contoh kasus ini dapat dilihat pada tabel 3.8 yang berisikan hasil nilai *entropy* dari setiap kelas sentimen sebagai berikut :

Tabel 3.9. Hasil Nilai *Entropy* Kelas Sentimen

Kata (Term)	Entropy Positif	Entropy Negatif	Entropy Netral
Kuliah	0,528771238	0,521089678	0,521089678
Online	0,521089678	0,332192809	0,521089678
Adu	0,332192809	0	0
Strategi	0,332192809	0	0
Kecerdasan	0,332192809	0	0
Jangan	0,332192809	0	0
Lupa	0,332192809	0	0
Sarapan	0,332192809	0	0
Semangat	0,332192809	0	0
Sekolah	0,332192809	0	0
Selamat	0,332192809	0	0
Hari	0,332192809	0	0
Senin	0,332192809	0	0
Buat	0,332192809	0	0
Yang	0,332192809	0	0
Kerja	0,332192809	0	0
Daring	0,332192809	0,464385619	0
Or	0,332192809	0	0
Yaa	0,332192809	0	0
Bikin	0	0,332192809	0
Mood	0	0,332192809	0
Ambyar	0	0,332192809	0
Sangat	0	0,332192809	0
Tidak	0	0,464385619	0
Recomended	0	0,332192809	0
Gua	0	0,332192809	0
Dari	0	0,332192809	0
Kita	0	0,332192809	0
Tidak	0	0,332192809	0
Belajar	0	0,332192809	0
Mari	0	0	0,332192809
Kok	0	0	0,332192809
Gini	0	0	0,332192809
Amaat	0	0	0,332192809
Yaak	0	0	0,332192809

Tabel 3.10. Lanjutan Hasil Nilai *Entropy* Kelas Sentimen

Kata (<i>Term</i>)	Entropy Positif	Entropy Negatif	Entropy Netral
Offline	0	0	0,332192809
Ujian	0	0	0,332192809
Di	0	0	0,332192809
Rumah	0	0	0,332192809

Kemudian setelah mendapatkan nilai entropy dari setiap kelas sentimen, akan dilakukan perhitungan untuk mendapatkan nilai z , dimana nilai z merupakan nilai kepangkatan dari setiap *term* pada kelas sentimen, dimana kata kuliah sebagai contoh perhitungan untuk mencari nilai z tersebut dengan menggunakan rumus 2.5 sebagai berikut :

$$z \text{ Positif (kuliah)} = \exp(0,528771238) = 1,696846007$$

$$z \text{ Negatif (kuliah)} = \exp(0,521089678) = 1,683861518$$

$$z \text{ Netral (kuliah)} = \exp(0,521089678) = 1,683861518$$

Gambaran dari contoh kasus ini dapat dilihat pada tabel 3.9 yang berisikan hasil nilai entropy dari setiap kelas sentimen sebagai berikut :

Tabel 3.11. Hasil Nilai Z

Kata	Nilai Z Positif	Nilai Z Negatif	Nilai Z Netral
Kuliah	1,696846007	1,683861518	1,683861518
Online	1,683861518	1,394021603	1,683861518
Adu	1,394021603	1	1
Strategi	1,394021603	1	1
Kecerdasan	1,394021603	1	1
Jangan	1,394021603	1	1
Lupa	1,394021603	1	1
Sarapan	1,394021603	1	1
Semangat	1,683861518	1	1
Sekolah	1,394021603	1	1
Selamat	1,394021603	1	1
Hari	1,394021603	1	1
Senin	1,394021603	1	1
Buat	1,394021603	1,394021603	1
Yang	1,394021603	1	1
Kerja	1,394021603	1	1

Tabel 3.12. Lanjutan Hasil Nilai Z

Kata	Nilai Z Positif	Nilai Z Negatif	Nilai Z Netral
Daring	1,394021603	1,591036386	1
Or	1,394021603	1	1
Yaa	1,394021603	1	1
Bikin	1	1,394021603	1
Mood	1	1,394021603	1
Ambyar	1	1,394021603	1
Sangat	1	1,394021603	1
Tidak	1	1,591036386	1
Recomended	1	1,394021603	1
Gua	1	1,394021603	1
Dari	1	1,394021603	1
Kita	1	1,394021603	1
Tidak	1	1,394021603	1
Belajar	1	1,394021603	1
Mari	1	1	1,394021603
Kok	1	1	1,394021603
Gini	1	1	1,394021603
Amaat	1	1	1,394021603
Yaak	1	1	1,394021603
Offline	1	1	1,394021603
Ujian	1	1	1,394021603
Di	1	1	1,394021603
Rumah	1	1	1,394021603

Selanjutnya akan dilakukan perhitungan untuk mencari nilai probabilitas pada *term* dari setiap kelas sentimen dengan menggunakan rumus 2.6 sebagai berikut :

$$P(\text{Positif|Kuliah}) = \left(\frac{1}{1,696846007} \right) \times (0,528771238) = 0,31162005$$

$$P(\text{Negatif|Kuliah}) = \left(\frac{1}{1,683861518} \right) \times (0,521089678) = 0,309461124$$

$$P(\text{Netral|Kuliah}) = \left(\frac{1}{1,683861518} \right) \times (0,521089678) = 0,309461124$$

Gambaran dari contoh kasus ini dapat dilihat pada tabel 3.11 yang berisikan hasil nilai probabilitas dari setiap *term* yang ada pada masing-masing kelas sentimen sebagai berikut :

Tabel 3.13. Hasil Nilai Probabilitas

Kata	Prob. Positif	Prob. Negatif	Prob. Netral
Kuliah	0,31162005	0,309461124	0,309461124
Online	0,309461124	0,238298179	0,309461124
Adu	0,238298179	0	0
Strategi	0,238298179	0	0
Kecerdasan	0,238298179	0	0
Jangan	0,238298179	0	0
Lupa	0,238298179	0	0
Sarapan	0,238298179	0	0
Semangat	0,309461124	0	0
Sekolah	0,238298179	0	0
Selamat	0,238298179	0	0
Hari	0,238298179	0	0
Senin	0,238298179	0	0
Buat	0,238298179	0,238298179	0
Yang	0,238298179	0	0
Kerja	0,238298179	0	0
Daring	0,238298179	0,291876178	0
Or	0,238298179	0	0
Yaa	0,238298179	0	0
Bikin	0	0,238298179	0
Mood	0	0,238298179	0
Ambyar	0	0,238298179	0
Sangat	0	0,238298179	0
Tidak	0	0,238298179	0
Recomended	0	0,238298179	0
Gua	0	0,238298179	0
Dari	0	0,238298179	0
Kita	0	0,238298179	0
Tidak	0	0,238298179	0
Belajar	0	0,238298179	0
Mari	0	0	0,238298179
Kok	0	0	0,238298179
Gini	0	0	0,238298179
Amaat	0	0	0,238298179
Yaak	0	0	0,238298179
Offline	0	0	0,238298179
Ujian	0	0	0,238298179
Di	0	0	0,238298179
Rumah	0	0	0,238298179

Langkah terakhir dalam penentuan kelas sentimen adalah dengan melakukan perhitungan pada setiap *term* yang ada pada komentar 1 yang ada pada tabel 3.6 yang berisikan contoh data komentar, dengan menggunakan nilai probabilitas yang sudah didapatkan sebagai berikut :

1. Probabilitas untuk sentimen positif

$$\begin{aligned}
 & P(\text{"Positif"}|\text{"Kuliah"}) \times P(\text{"Positif"}|\text{"Online"}) \times P(\text{"Positif"}|\text{"Adu"}) \times P \\
 & (\text{"Positif"}|\text{"Strategi"}) \times P(\text{"Positif"}|\text{"Adu"}) \times P(\text{"Positif"}|\text{"Kecerdasan"}) \\
 & = 0,31162005 \times 0,309461124 \times 0,238298179 \times 0,238298179 \times 0,238298179 \times \\
 & 0,238298179 \\
 & = 0,000310967
 \end{aligned}$$

2. Probabilitas untuk sentimen negatif

$$\begin{aligned}
 & P(\text{"Negatif"}|\text{"Kuliah"}) \times P(\text{"Negatif"}|\text{"Online"}) \times P(\text{"Negatif"}|\text{"Adu"}) \times P \\
 & (\text{"Negatif"}|\text{"Strategi"}) \times P(\text{"Negatif"}|\text{"Adu"}) \times P(\text{"Negatif"}|\text{"Kecerdasan"}) \\
 & = 0,309461124 \times 0,238298179 \times 0 \times 0 \times 0 \times 0 \\
 & = 0
 \end{aligned}$$

3. Probabilitas untuk sentimen netral

$$\begin{aligned}
 & P(\text{"Netral"}|\text{"Kuliah"}) \times P(\text{"Netral"}|\text{"Online"}) \times P(\text{"Netral"}|\text{"Adu"}) \times P \\
 & (\text{"Netral"}|\text{"Strategi"}) \times P(\text{"Netral"}|\text{"Adu"}) \times P(\text{"Netral"}|\text{"Kecerdasan"}) \\
 & = 0,309461124 \times 0,309461124 \times 0 \times 0 \times 0 \times 0 \\
 & = 0
 \end{aligned}$$

Selanjutnya penentuan kelas sentimen pada komentar 5 yang ada pada tabel 3.6 sebagai berikut :

1. Probabilitas untuk sentimen positif

$$\begin{aligned}
 & P(\text{"Positif"}|\text{"Kuliah"}) \times P(\text{"Positif"}|\text{"Daring"}) \times P(\text{"Positif"}|\text{"Bikin"}) \times P \\
 & (\text{"Positif"}|\text{"Mood"}) \times P(\text{"Positif"}|\text{"Ambyar"})
 \end{aligned}$$

$$= 0,31162005 \times 0,238298179 \times 0 \times 0 \times 0$$

$$= 0$$

2. Probabilitas untuk sentimen negatif

$$P(\text{"Negatif"}|\text{"Kuliah"}) \times P(\text{"Negatif"}|\text{"Daring"}) \times P(\text{"Negatif"}|\text{"Bikin"}) \times P(\text{"Negatif"}|\text{"Mood"}) \times P(\text{"Negatif"}|\text{"Ambyar"})$$

$$= 0,309461124 \times 0,291876178 \times 0,238298179 \times 0,238298179 \times 0,238298179$$

$$= 0,001222269$$

3. Probabilitas untuk sentimen netral

$$P(\text{"Netral"}|\text{"Kuliah"}) \times P(\text{"Netral"}|\text{"Daring"}) \times P(\text{"Netral"}|\text{"Bikin"}) \times P(\text{"Netral"}|\text{"Mood"}) \times P(\text{"Netral"}|\text{"Ambyar"})$$

$$= 0,309461124 \times 0 \times 0 \times 0 \times 0$$

$$= 0$$

Kemudian penentuan kelas pada komentar 8 yang ada pada tabel 3.6 sebagai berikut :

1. Probabilitas untuk sentimen positif

$$P(\text{"Positif"}|\text{"Mari"}) \times P(\text{"Positif"}|\text{"Kuliah"}) \times P(\text{"Positif"}|\text{"Online"})$$

$$= 0 \times 0,31162005 \times 0,309461124$$

$$= 0$$

2. Probabilitas untuk sentimen negatif

$$P(\text{"Negatif"}|\text{"Mari"}) \times P(\text{"Negatif"}|\text{"Kuliah"}) \times P(\text{"Negatif"}|\text{"Online"})$$

$$= 0 \times 0,309461124 \times 0,238298179$$

$$= 0$$

3. Probabilitas untuk sentimen netral

$$P(\text{"Netral"}|\text{"Mari"}) \times P(\text{"Netral"}|\text{"Kuliah"}) \times P(\text{"Netral"}|\text{"Online"})$$

$$= 0,238298179 \times 0,309461124 \times 0,309461124$$

$$= 0,000308813$$

Gambaran dari contoh kasus perhitungan untuk penentuan kelas sentimen ini dapat dilihat pada tabel 3.12 yang berisikan hasil nilai probabilitas sentimen dari setiap komentar sebagai berikut :

Tabel 3.14. Hasil Nilai Probabilitas Sentimen

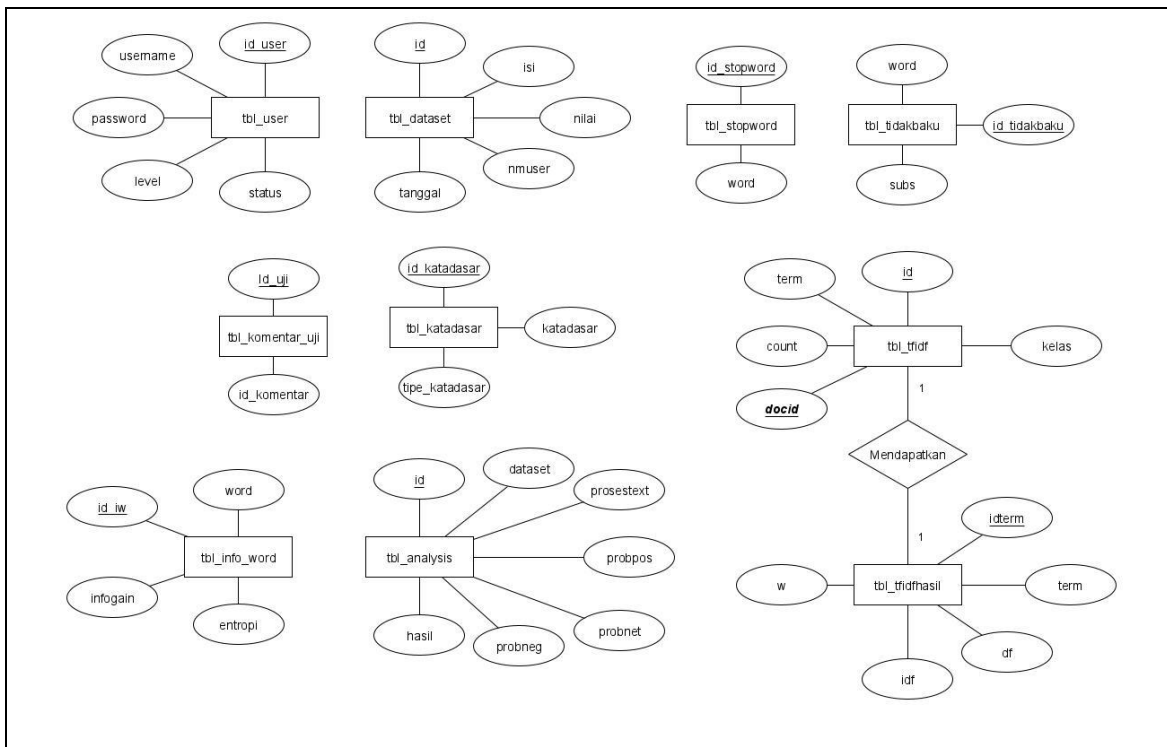
No	Komentar	Prob. Positif	Prob. Negatif	Prob. Netral
1	kuliah online adu strategi adu kecerdasan	0,000310967	0	0
2	jangan lupa sarapan. semangat sekolah/kuliah online nya	0,0000962322	0	0
3	selamat hari senin, semangat kuliah online nya	0,000403831	0	0
4	Buat yg kerja,daring or kuliah semangat yaa!	0,0000176586	0	0
5	Kuliah daring bikin mood ambyar	0	0,001222269	0
6	Kuliah daring sangat tidak recommended buat gua	0	0,0000850132	0
7	Dari kuliah online kita tidak belajar'	0	0,000291265	0
8	mari kuliah online	0		0,022821
9	Kuliah online kok gini amat yaaaak ?	0		0,000309
10	kuliah offline tapi ujiannya online di rumah	0		0,000309

3.6. Perancangan Basis Data

Perancangan basis data merupakan salah satu komponen penting dalam sistem yang berfungsi sebagai basis pengetahuan untuk mengambil kesimpulan. Berikut ini tahapan yang harus dirancang dalam merancang basis data:

3.6.1. Entity Relations Diagram (ERD)

Entity Relationship Diagram (ERD) merupakan model yang digunakan untuk mendesain *database* dengan tujuan menggambarkan data yang berelasi pada sebuah tabel *database*. Pada ERD ini terdapat entitas, atribut dan relasi. Entitas merupakan suatu objek yang diidentifikasi dalam lingkungan pemakai sistem. Atribut merupakan properti yang dimiliki suatu entitas. Relasi merupakan hubungan yang terjadi antara satu atau lebih entitas. ERD sistem ini dapat dilihat pada Gambar 3.7 dibawah ini



Gambar 3.7. Entity Relations Diagram

3.6.2. Perancangan Struktur Tabel

Dalam memenuhi kebutuhan sistem, maka diperlukan basis data yang dapat menampung seluruh data yang dibutuhkan. Basisdata dalam sistem ini terdiri dari beberapa tabel, sebagai berikut :

1. **tbl_user**

tbl_user merupakan tabel yang berisikan data *user*.

Tabel 3.15. tbl_user

No	Nama	Tipe	Constraint	Keterangan
1	<i>id_user</i>	<i>Int (15)</i>	<i>None, primary key, auto increment</i>	Id Pengguna
2	<i>username</i>	<i>Varchar(20)</i>	<i>None</i>	User pengguna
3	<i>passwords</i>	<i>Varchar(20)</i>	<i>None</i>	Kata sandi
4	<i>level</i>	<i>Int(11)</i>	<i>None</i>	Multi user
5	<i>status</i>	<i>Varchar(10)</i>	<i>None</i>	Status user

2. **tbl_dataset**

tbl_dataset merupakan tabel yang digunakan untuk menyimpan *dataset*.

Tabel 3.16. tbl_dataset

No	Nama	Tipe	Constraint	Keterangan
1	id	<i>Int (255)</i>	<i>None, primary key, auto increment</i>	Id komentar
2	isi	<i>Text</i>	<i>None</i>	Isi komentar
3	nilai	<i>Varchar(10)</i>	<i>None</i>	Nilai komentar
4	nmuser	<i>Varchar(50)</i>	<i>None</i>	Nama user
5	tanggal	<i>Datetime</i>	<i>None</i>	Tanggal

3. tbl_stopword

tbl_stopword merupakan tabel yang berisikan data-data *stopwords*.

Tabel 3.17. tbl_stopword

No	Nama	Tipe	Constraint	Keterangan
1	<i>id_stopwords</i>	<i>Int (15)</i>	<i>None, primary key, auto increment</i>	Id <i>stopwords</i>
2	<i>word</i>	<i>Varchar(20)</i>	<i>None</i>	Kata <i>stopwords</i>

4. tbl_katadasar

tbl_katadasar merupakan tabel yang menyimpan informasi kata dasar.

Tabel 3.18. tbl_katadasar

No	Nama	Tipe	Constraint	Keterangan
1	id_ktdasar	<i>Int (10)</i>	<i>None, primary key, auto increment</i>	Id kata
2	katadasar	<i>Varchar(20)</i>	<i>None</i>	Kata
3	tipe_katadasar	<i>Varchar(20)</i>	<i>None</i>	Tipe kata dasar

5. tbl_tidakbaku

tbl_tidakbaku merupakan tabel yang berisikan kata-kata non baku dan kata baku.

Tabel 3.19. tbl_tidakbaku

No	Nama	Tipe	Constraint	Keterangan
1	Id_tidakbaku	<i>Int(10)</i>	<i>None, primary key, auto increment</i>	Id Kata Tidak Baku
2	Words	<i>Varchhar(100)</i>	<i>None</i>	Kata
3	Subs	<i>Varchar(150)</i>	<i>None</i>	Kata Baku

6. `tbl_info_word`

`tbl_info_word` merupakan tabel yang menyimpan nilai informasi dari setiap kata.

Tabel 3.20. `tbl_info_word`

No	Nama	Tipe	Constraint	Keterangan
1	<code>id_iw</code>	<i>Int (255)</i>	<i>None, primary key, auto increment</i>	Id kata
2	<code>word</code>	<i>Varchar(100)</i>	<i>None</i>	Kata
3	<code>entropi</code>	<i>Varchar(30)</i>	<i>None</i>	Nilai entropi
4	<code>infogain</code>	<i>Varchar(30)</i>	<i>None</i>	Nilai <i>Information Gain</i>

7. `tbl_tfidf`

`tbl_tfidf` merupakan tabel data yang berisikan nilai frekuensi dari dokumen kata.

Tabel 3.21. `tbl_idf`

No	Nama	Tipe	Constraint	Keterangan
1	<code>id</code>	<i>Int (255)</i>	<i>None, primary key, auto increment</i>	Id kata
2	<code>term</code>	<i>Varchar(100)</i>	<i>None</i>	Istilah/Kata
3	<code>count</code>	<i>Int(20)</i>	<i>None</i>	Hitung kata
4	<code>docid</code>	<i>Int(100)</i>	<i>None</i>	Id dokumen kata
5	<code>kelas</code>	<i>Varchar(3)</i>	<i>None</i>	Kelas

8. `tbl_tfidfhasil`

Tabel idf hasil merupakan tabel yang merupakan nilai kemunculan dari setiap kata.

Tabel 3.22. Tabel IDF hasil

No	Nama	Tipe	Constraint	Keterangan
1	<code>idterm</code>	<i>Int (255)</i>	<i>None, primary key, auto increment</i>	Id kata
2	<code>term</code>	<i>Varchar(50)</i>	<i>None</i>	Istilah/Kata
3	<code>df</code>	<i>Int(50)</i>	<i>None</i>	Frekuensi Kemunculan Kata
4	<code>Idf</code>	<i>Varchar(50)</i>	<i>None</i>	Nilai Frekuensi

9. `tbl_komentar_uji`

Tabel komentar uji berisikan tabel informasi dari komentar data uji.

Tabel 3.23. tbl_komentar_uji

No	Nama	Tipe	Constraint	Keterangan
1	<i>Id_uji</i>	<i>Int(250)</i>	<i>None, primary key, auto increment</i>	
2	<i>Id_komentar</i>	<i>Int(100)</i>	<i>none</i>	

10. tbl_analysis

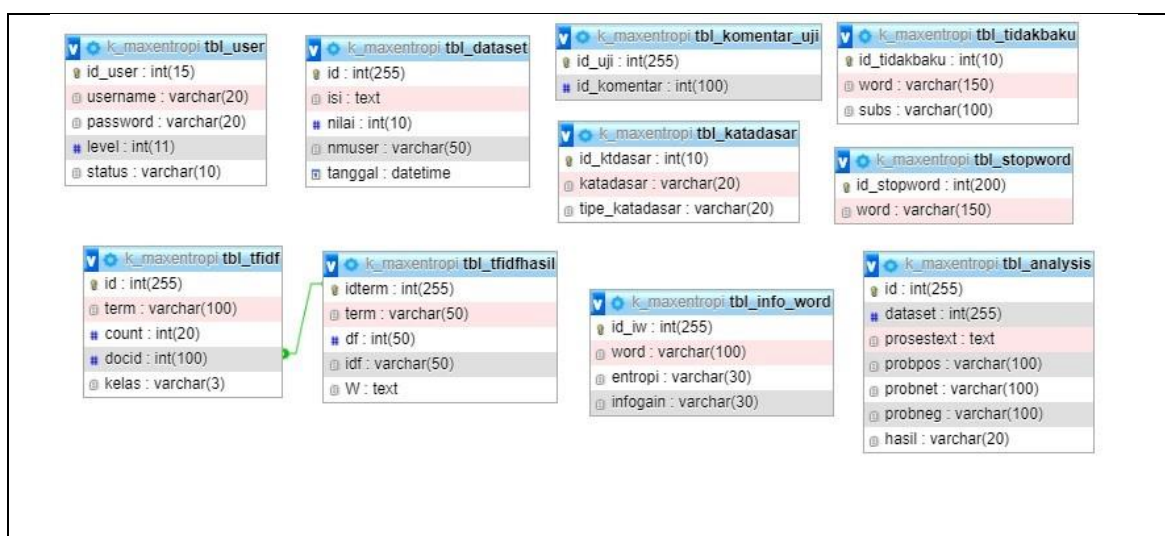
tbl_analysis merupakan tabel data yang berisikan nilai-nilai probabilitas dan hasil yang disimpan kedalam *database*.

Tabel 3.24. tbl_analysis

No	Nama	Tipe	Constraint	Keterangan
1	<i>id</i>	<i>Int(255)</i>	<i>None, primary key, auto increment</i>	Id dataset
2	<i>dataset</i>	<i>Int(255)</i>	<i>None</i>	Dataset
3	<i>prosestext</i>	<i>Text</i>	<i>None</i>	Hasil <i>preprocessing text</i>
4	<i>probpos</i>	<i>Varchar(100)</i>	<i>None</i>	Nilai probabilitas positif
5	<i>probnet</i>	<i>Varchar(100)</i>	<i>None</i>	Nilai probabilitas netral
6	<i>probneg</i>	<i>Varchar(100)</i>	<i>None</i>	Nilai probabilitas negatif
7	hasil	<i>Varchar(20)</i>	<i>None</i>	Nilai sentimen

3.6.3. Relasi Antar Tabel (RAT)

RAT bertujuan untuk mempermudah dan mengetahui tabel dan atribut yang saling berelasi. Relasi antar tabel pada perancangan basis data dapat dilihat pada gambar 3.8 berikut.



Gambar 3.8. Relasi Antar Tabel

3.7. Perancangan Antarmuka Sistem

Rancangan Antarmuka (*User interface*) analisis sentimen pada sosial media *twitter*, terdiri dari beberapa bagian, diantaranya adalah sebagai berikut:

3.7.1 Perancangan Antarmuka Admin

Rancangan Antarmuka (*User interface*) untuk halaman *admin* terdiri dari beberapa bagian, sebagai berikut :

1. Halaman *Login Admin*

Halaman *login* merupakan halaman yang digunakan oleh *admin* untuk masuk kedalam aplikasi. *Admin* mengisikan nama pengguna dan kata sandi dengan benar untuk bisa masuk kedalam aplikasi. Rancangan halaman login dapat dilihat pada gambar 3.9

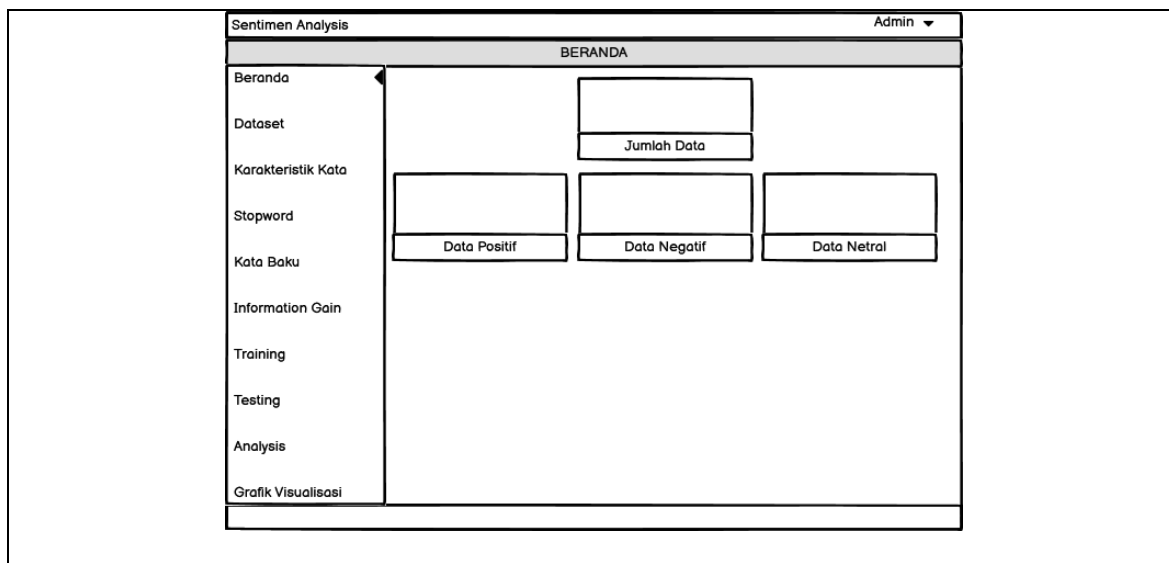


The image shows a wireframe of an admin login page. It features a title bar at the top labeled 'Login'. Below the title bar, there are two input fields: the first is labeled 'Username' and has a user icon on the right; the second is labeled 'Password' and has a lock icon on the right. Below these fields is a blue button with a magnifying glass icon and the text 'Login'.

Gambar 3.9. Rancangan Halaman *Login Admin*

2. Rancangan Halaman Beranda

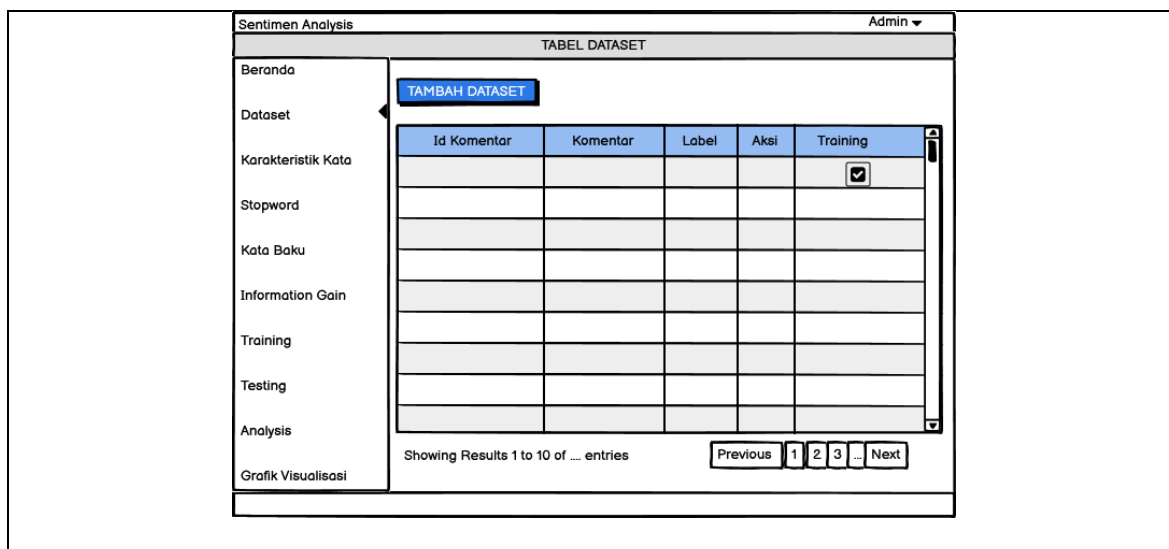
Halaman beranda merupakan halaman yang akan muncul ketika *login* berhasil. Perancangan halaman beranda dapat dilihat pada gambar 3.10



Gambar 3.10. Rancangan Halaman Beranda *Admin*

3. Rancangan Halaman *Dataset*

Halaman *dataset* merupakan halaman untuk mengolah komentar. Melalui data komentar, *admin* dapat melakukan *input* data. *Admin* juga dapat menghapus data maupun menambah komentar dan memilih data *training*. Rancangan Halaman data komentar dapat dilihat pada gambar 3.11.



Gambar 3.11. Rancangan Halaman *Dataset*

4. Rancangan Halaman *Input Dataset*

Pada halaman *input dataset admin* dapat melakukan penambahan dataset dengan format *file* csv. Halaman input dataset dapat dilihat pada gambar 3.12.

Gambar 3.12. Rancangan Halaman *Input Dataset Admin*

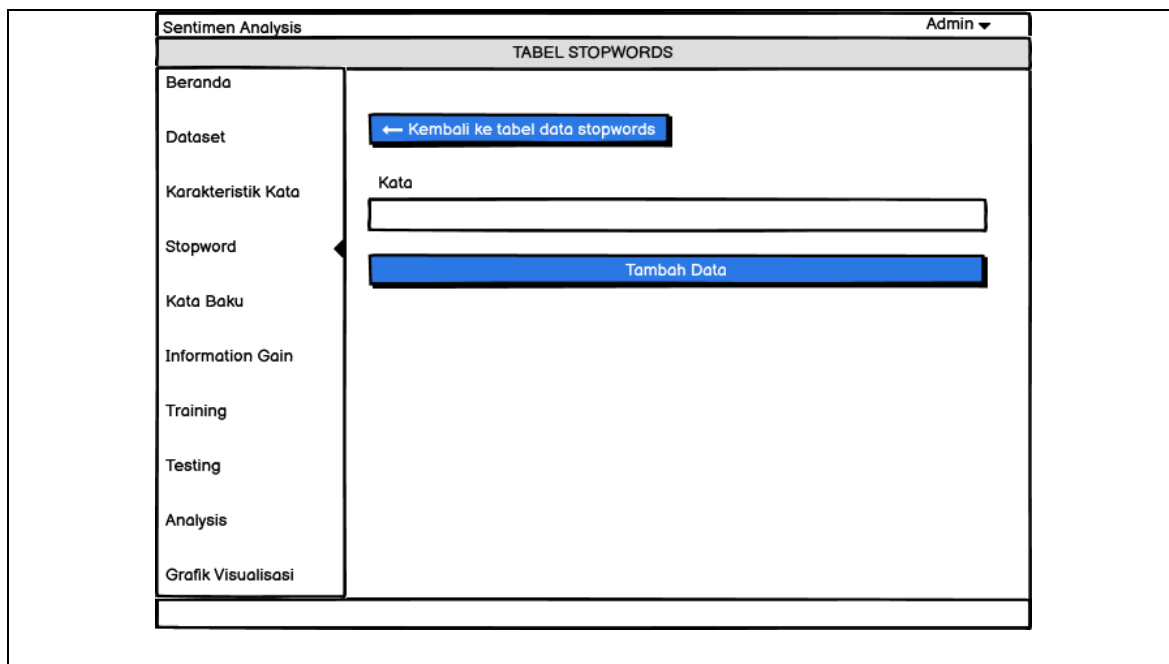
5. Rancangan Halaman *Stopwords*

Pada halaman *stopwords admin* dapat melakukan *update* data untuk perubahan dan penambahan data. Halaman data *stopwords* dapat dilihat pada gambar 3.13.

Gambar 3.13. Rancangan Halaman *Stopwords*

6. Rancangan Halaman *Input Stopwords*

Pada halaman *input dataset admin* dapat melakukan penambahan data *stopwords*. Halaman *input stopwords* dapat dilihat pada gambar 3.14

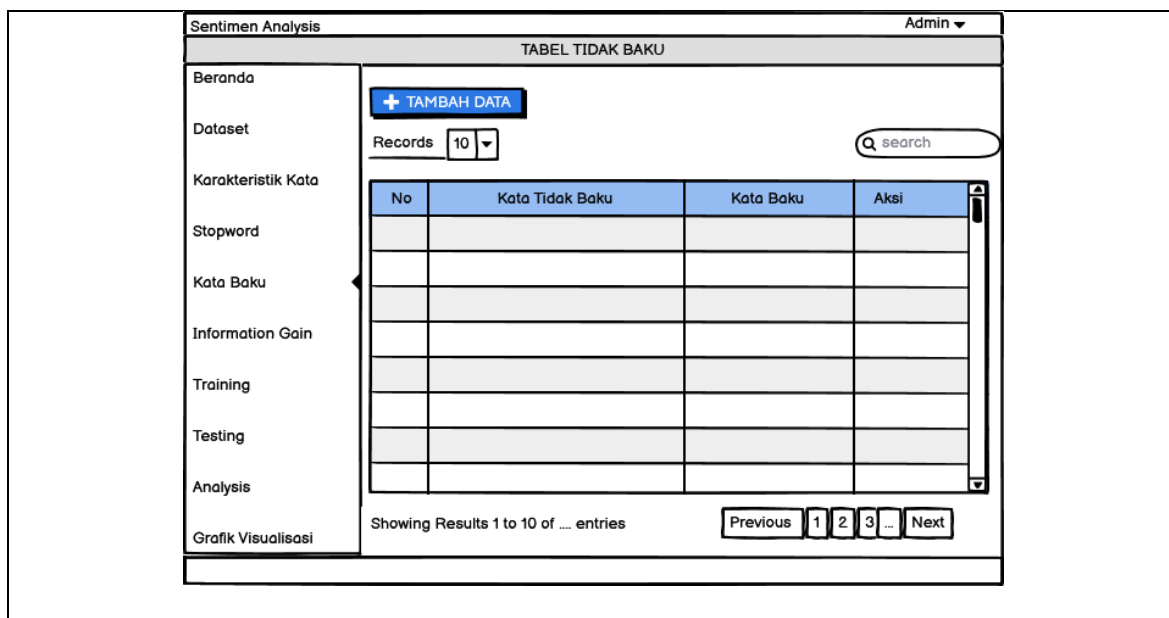


Gambar 3.14. Rancangan Halaman *Input Stopwords Admin*

7. Rancangan Halaman Kata Baku

Pada halaman kata baku berisikan kata-kata yang tidak baku dan baku pada sistem.

Admin dapat melakukan *update*, *edit* dan hapus pada kata baku dan tidak baku.



Gambar 3.15. Rancangan Halaman Kata Baku

8. Rancangan Halaman Input Kata Baku

Halaman *input* kata baku merupakan halaman untuk menambahkan data kata tidak baku dan baku.

Gambar 3.16. Rancangan Halaman *Input* Kata Baku

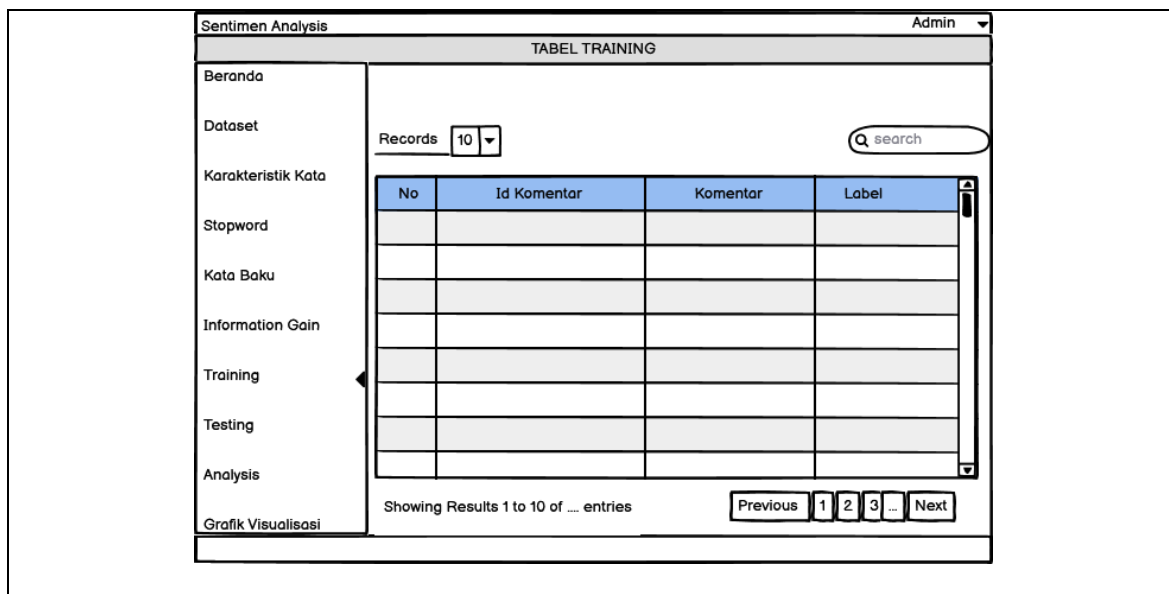
9. Rancangan Halaman *Information Gain*

Pada halaman *information gain* dapat digunakan oleh *user* untuk melihat nilai *entropy* dan *gain* pada tiap kata dan *admin* juga dapat melakukan hitung ulang jika ada data baru yang masuk. Rancangan halaman *information gain* dapat dilihat pada gambar 3.17.

Gambar 3.17. Rancangan Halaman *Information Gain Admin*

10. Rancangan Halaman *Training*

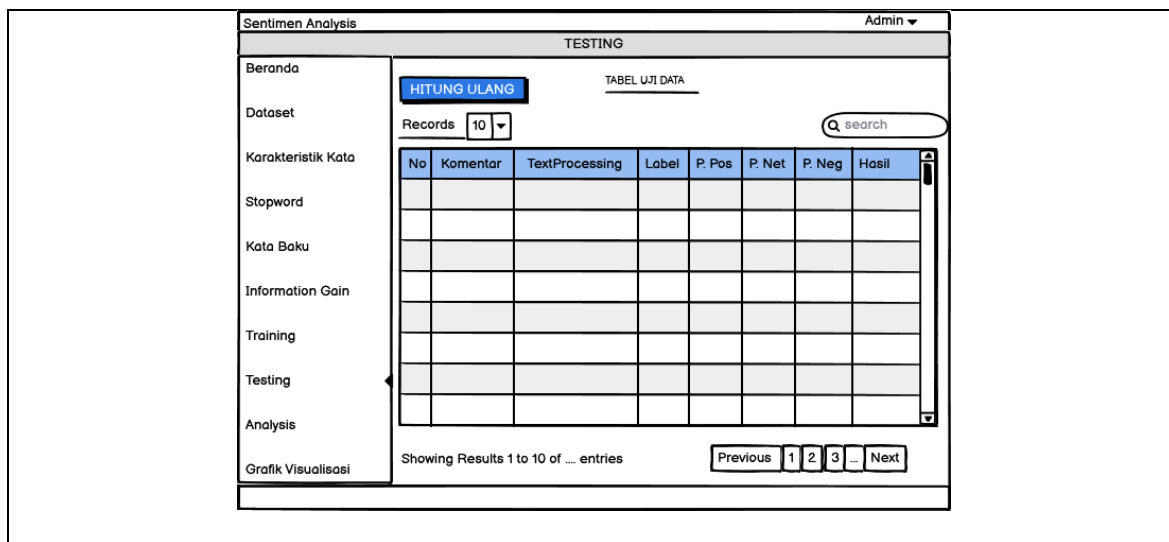
Halaman *training* merupakan halaman yang berisikan tabel data latih komentar dengan sentimen manual. Rancangan halaman training terdapat pada gambar 3.18.



Gambar 3.18. Rancangan Halaman *Training Admin*

11. Rancangan Halaman *Testing*

Halaman *testing* merupakan halaman tabel uji data komentar yang telah mendapatkan hasil pengujian dan perhitungan untuk menentukan nilai probabilitas positif, negatif dan netral. Pada gambar 3.19 merupakan gambar dari rancangan halaman *testing*.

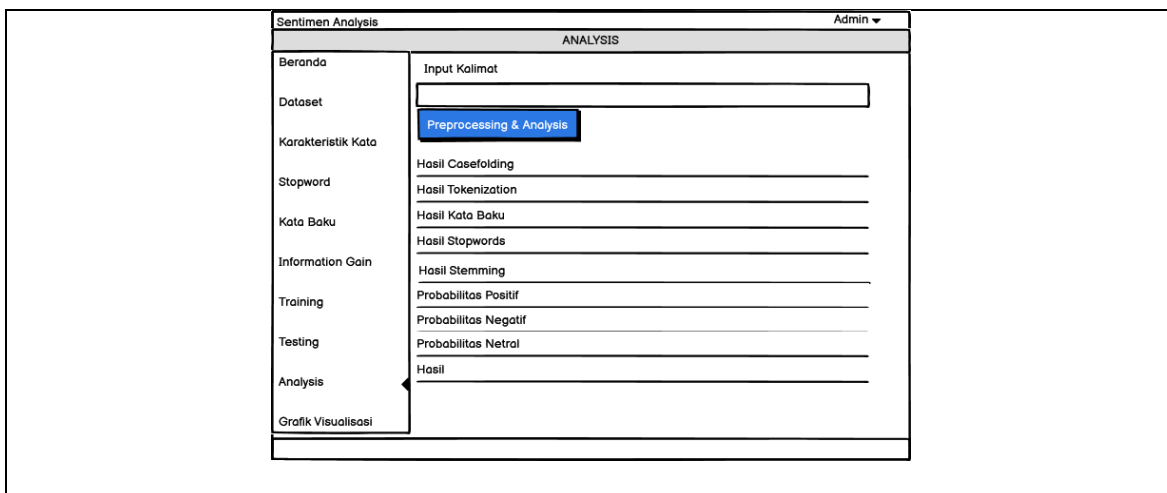


Gambar 3.19. Rancangan Halaman *Testing Admin*

12. Rancangan Halaman *Analysis*

Pada halaman *analysis*, *user* dapat *input* komentar untuk melakukan pengujian hasil sentimen yang ingin diketahui hasilnya. Ketika komentar di *input*, maka akan diproses untuk

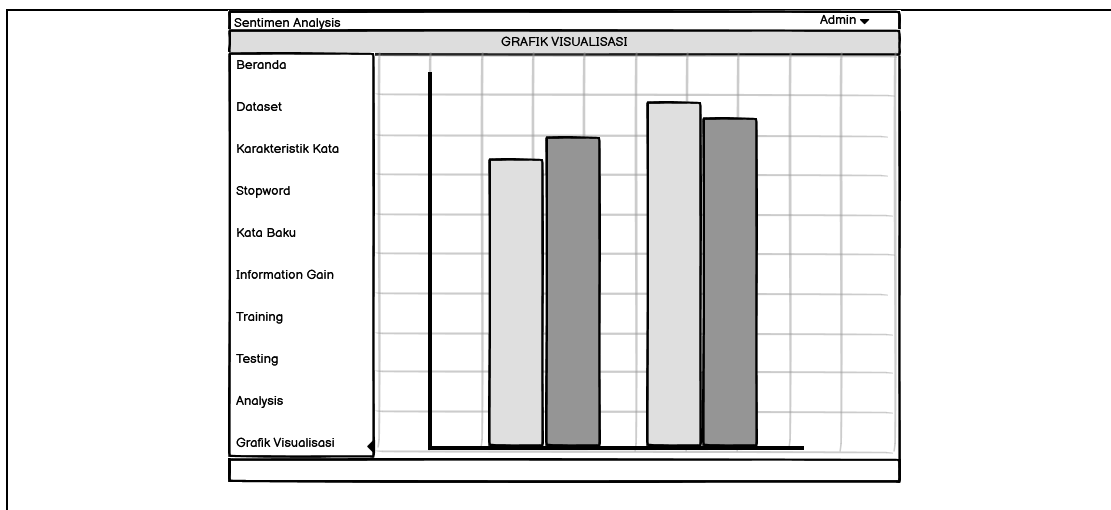
menemukan hasil sentimen dan nilai probabilitasnya dengan menggunakan metode *maximum entropy*. Rancangan halaman *analysis* dapat dilihat pada gambar 3.20.



Gambar 3.20. Rancangan Halaman *Analysis Admin*

13. Rancangan Halaman Grafik

Halaman grafik merupakan halaman yang memuat hasil grafik dari hasil klasifikasi sentimen yang ada pada sistem. Rancangan pada halaman grafik dapat dilihat pada gambar 3.21.



Gambar 3.21. Rancangan Halaman Grafik *Admin*

BAB IV

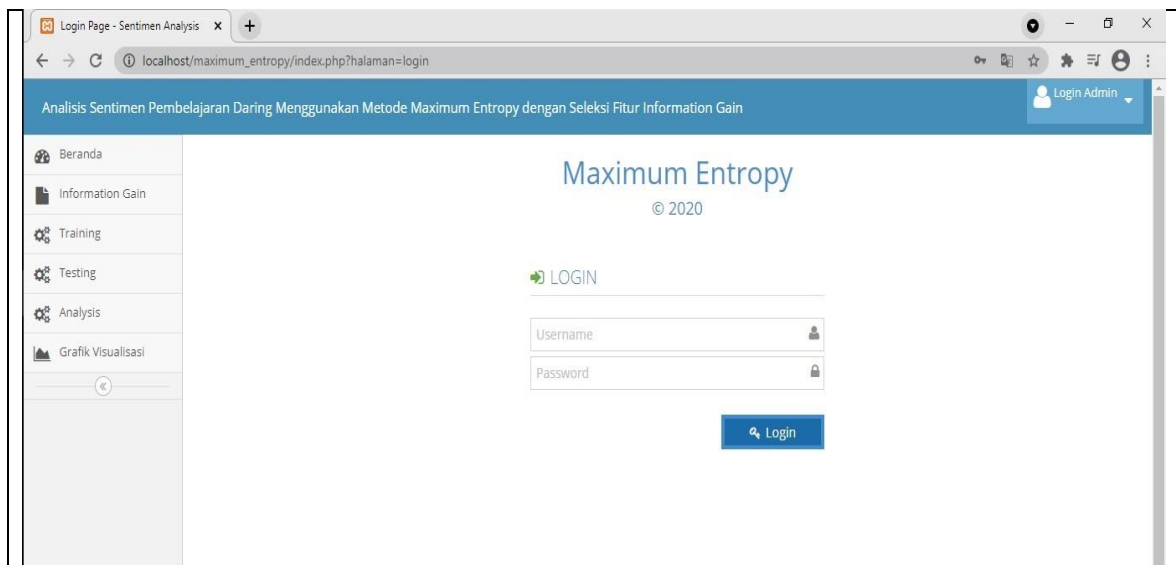
HASIL, PENGUJIAN DAN PEMBAHASAN

4.1 Implementasi Aplikasi

Implementasi aplikasi merupakan tampilan pada program saat aplikasi dioperasikan oleh admin. Berikut ini akan ditampilkan *screenshot* dari *interface* aplikasi yang telah dibangun beserta *source code* dari aplikasi yang sudah dibangun.

4.1.1 Halaman Login

Halaman ini merupakan tahapan *login* bagi admin yang pertama kali muncul saat aplikasi dijalankan. Untuk dapat masuk pada aplikasi yang sudah dijalankan, admin harus terlebih dahulu menginputkan *username* dan *password* agar bisa masuk kehalaman berikutnya. Tampilan halaman *login* dapat dilihat pada gambar 4.1 berikut :



Gambar 4.1. Halaman *Login*

Berikut ini merupakan *source code* pada halaman *login* :

```
<?php
session_start();
ob_start();
include_once "config.database.php";
include_once "module.php";
$koneksi = koneksi();
if(isset($_SESSION['login'])&&isset($_SESSION['id'])&&isset($_SESSION['username'])) {
if(isset($_GET['ajax'])) {
```

Modul Program 4.1. *Source Code* Halaman *Login*

```

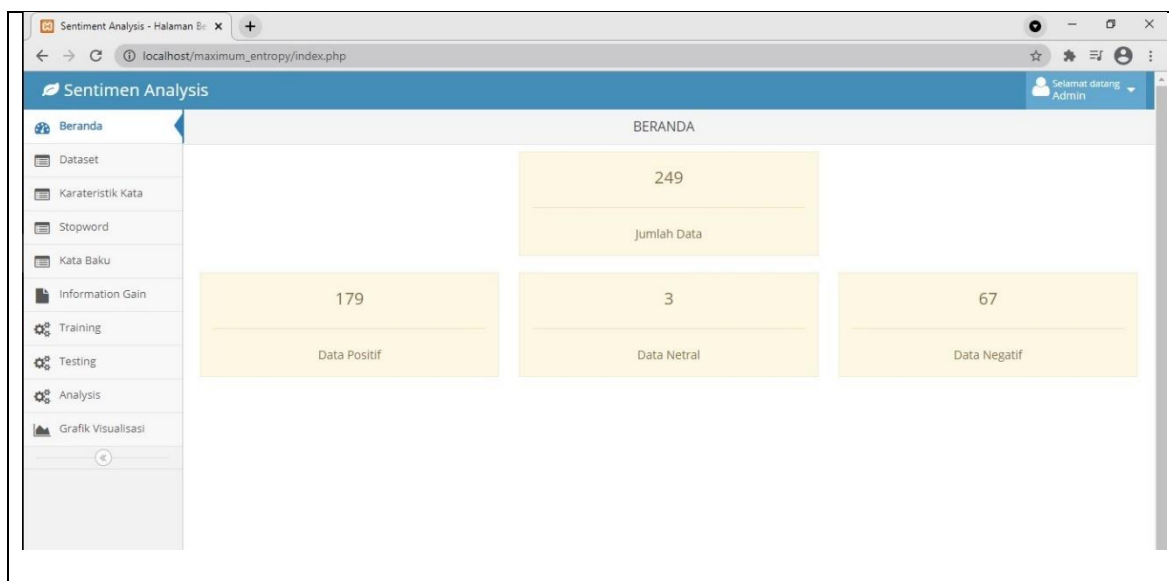
include_once "ajax.".$_SESSION['login'].".$_GET['ajax'].".php";
} else {
include_once "template.".$_SESSION['login'].".php";
}
} else {
    if(isset($_GET['ajax'])) {
        include_once "ajax.guest.".$_GET['ajax'].".php";
    } else {
        include_once "template.guest.php";
    }
}
}

```

Modul Program 4.2. *Source Code* Halaman *Login* (Lanjutan)

4.1.2 Halaman Beranda

Halaman beranda merupakan halaman awal pada aplikasi ketika admin sudah berhasil masuk ke sistem. Pada halaman beranda akan menampilkan jumlah data secara keseluruhan dan data yang sudah klasifikasi ke data positif, data negatif dan data netral. Pada halaman beranda juga terdapat beberapa menu yaitu, dataset, information gain, analysis, dan grafik visualisasi. Tampilan pada halaman beranda admin dapat dilihat pada gambar 4.2 berikut :



Gambar 4.2. Halaman Beranda

Berikut ini merupakan *source code* pada halaman beranda :

```

<?php
$string          = "SELECT * FROM `tbl_analysis`";
$q               = $koneksi->query($string);
$f               = $q->fetchAll(PDO::FETCH_NUM);
$totaldata      = count($f);
$pos             = 0;
$net             = 0;
$neg             = 0;

```

Modul Program 4.3. *Source Code* Halaman Beranda

```

foreach((array)$f as $fd) {
    //echo var_dump($fd[5])."<br />";
    if($fd[6]=="positif") {
        $pos = $pos+1;
    } else if($fd[6]=="negatif") {
        $neg = $neg + 1;
    } else {
        $net = $net + 1;
    }
}
?>

```

Modul Program 4.4. Lanjutan Source Code Halaman Beranda

4.1.3 Halaman Dataset

Halaman dataset merupakan halaman untuk mengelola data komentar, data komentar dibagi menjadi 2 kategori data yaitu data latih dan data uji. Pada halaman dataset admin juga bisa melakukan akses untuk melakukan penambahan, pengeditan maupun menghapus dataset komentar yang terdapat pada aplikasi. Dataset terdiri dari 300 komentar data uji sebanyak 100 komentar. Tampilan dari halaman dataset dapat dilihat pada gambar 4.3 berikut :

ID Komentar	Komentar	Label	Aksi	Training
22073	Kuliah online + ind*mie + Hujann Fellnya dapet banget anjayani	Positif	HAPUS	<input checked="" type="checkbox"/>
22074	Perdana pake aplikasi classroom untuk kuliah online. Udah pake baju rapi, udah bilang ke orang rumah, udah atur posisi dan pencahayaan, udah pake headset. Ternyata sistemnya lain, berasa di prank sumpaaaaaht!	Positif	HAPUS	<input checked="" type="checkbox"/>
22075	gue selalu berusaha buat ngehargain dosen apapun yg beliau ajarkan dan gmupun cara penyampaiannya kalo lg kuliah online, tp make sense gasih lo dianggap alpa krn terlambat absen? absen jam harus 8.00 trs gue keinput jam 8.01 karena sinyal indihome ilangan2, naudzubillah si dosen.	Positif	HAPUS	<input checked="" type="checkbox"/>
22076	astaga aku kangen kuliah offline :(makin ribet kuliah online	Positif	HAPUS	<input checked="" type="checkbox"/>
22077	Ya Allah semoga offline aku mau offline, aku pen kuliah di Unand tuh karena udaranya terus banyak phon eh malah online pas pertama kuliah mau offline pokoknya	Positif	HAPUS	<input checked="" type="checkbox"/>
22079	Enak gak enak kuliah online itu. Enaknya? Bisa belajar dimana aja. Gak enaknya? Kalau kuliahnya kebanyakan praktik ya bingung sendiri akhirnya belajar mandiri.	Positif	HAPUS	<input checked="" type="checkbox"/>
22078	Personally emg lebih enak ikut webinar drpd kuliah online. Kuliah trialu manut sama silabus, lain hal sama webinar yang pembahasannya lebih dinamis dan populer. I hope silabus perkuliahan jg bisa seseru materi2 webinar :)	Positif	HAPUS	<input checked="" type="checkbox"/>
22080	"Cuma mau bilang ""Terimakasih Kuliah Online""Kalo dibilang stress tugas sih ya stress, but karna online dan lagi di rumah aku nerata hener hener henvukur haneet.Karena ketika kita bosan haneet sama tugas, kita bisa lansung danet hiburan."	Positif	HAPUS	<input checked="" type="checkbox"/>

Gambar 4.3. Halaman Dataset

Berikut ini merupakan *source code* pada halaman dataset


```

<?php
if(isset($_GET['delkom'])) {
    $string = "DELETE FROM `tbl_dataset` WHERE
`id`='{$_GET['delkom']}';
    $q = $koneksi->query($string);

    if($q) {
        $notice = "Berhasil menghapus data";
    } else {
        $error = "Gagal menghapus data";
    }
}
if(isset($_POST['input_pilih'])) {
    $input= isset($_POST['input_centang']);}

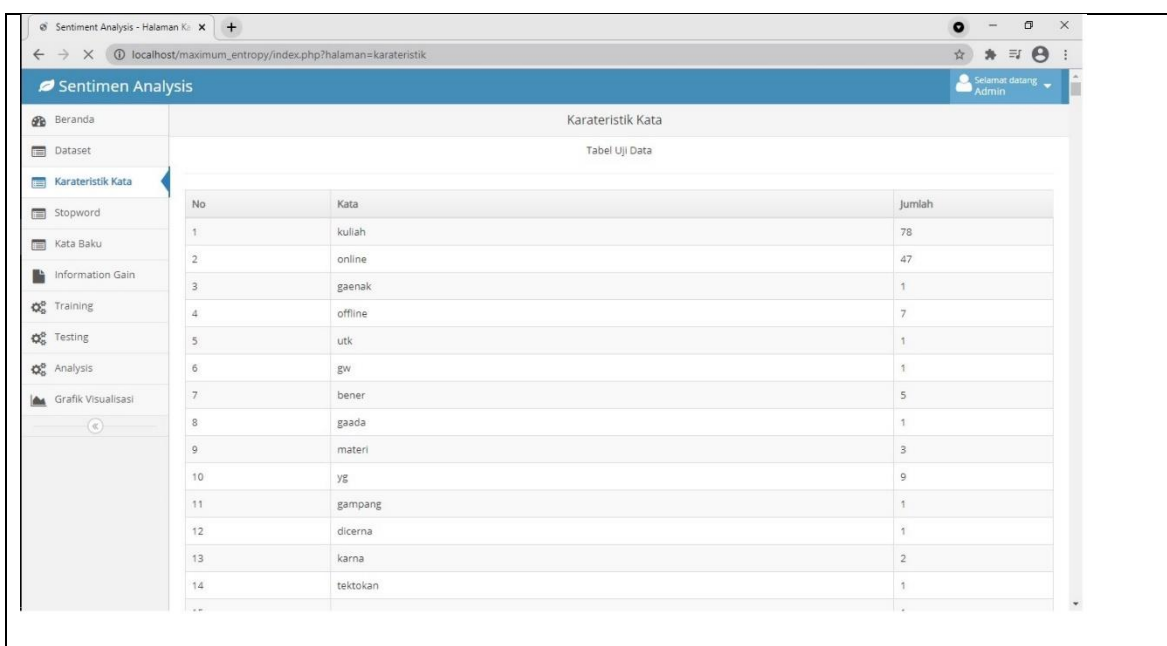
$string = "SELECT a.*,b.`id_uji` FROM tbl_dataset AS a
LEFT JOIN `tbl_komentar_uji` AS b ON a.`id`=b.`id_komentar`";
$q = $koneksi->query($string);
if ($q && $q->rowCount () > 0) {
    $f = $q->fetchAll(PDO::FETCH_NUM);
} else {
    $f = array();
}
?>

```

Modul Program 4.5. *Source Code* Halaman Dataset

4.1.4 Halaman Karakteristik Kata

Halaman karakteristik kata merupakan halaman yang berisikan jumlah kata yang digunakan pada sistem. Pada halaman karakteristik ini pengguna bisa melihat informasi jumlah kata yang paling sering muncul. Tampilan halaman karakteristik kata dapat dilihat pada gambar 4.4 berikut :

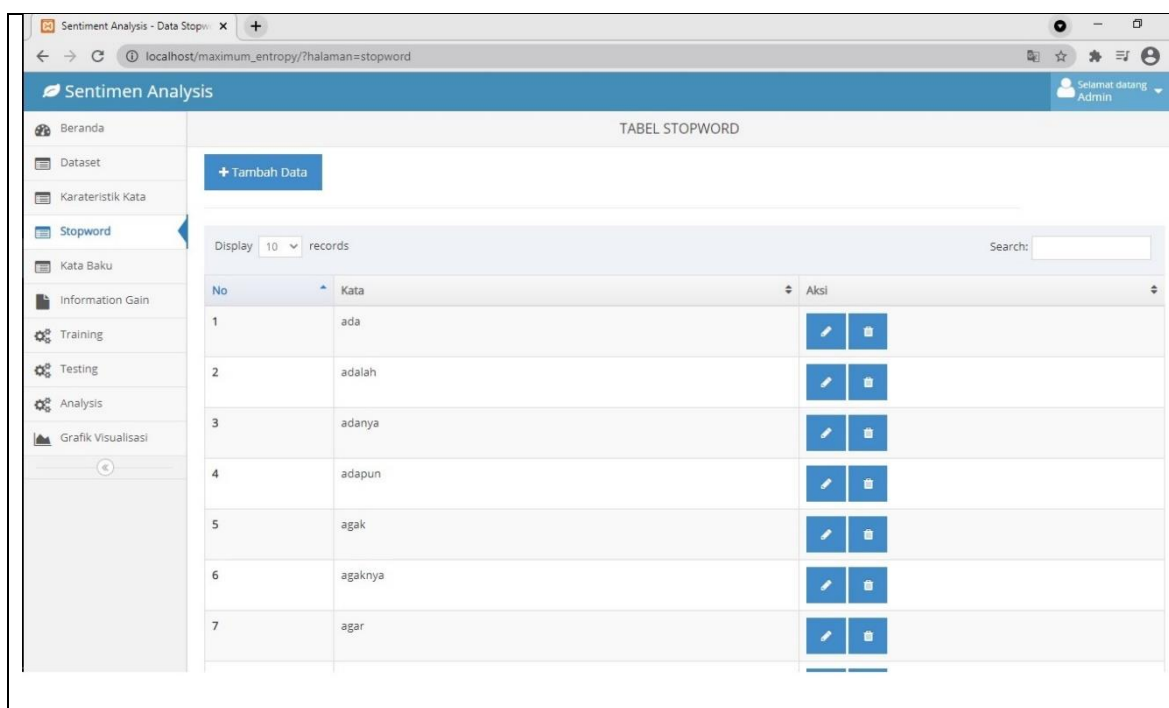


No	Kata	Jumlah
1	kuliah	78
2	online	47
3	gaenak	1
4	offline	7
5	utk	1
6	gw	1
7	bener	5
8	gaada	1
9	materi	3
10	yg	9
11	gampang	1
12	dicerna	1
13	karna	2
14	tektokan	1
..	.	.

Gambar 4.4. Halaman Karakteristik Kata

4.1.5. Halaman Stopword

Halaman *stopword* merupakan halaman yang berisikan kata-kata yang digunakan pada saat *preprocessing*. *Stopword* merupakan kumpulan kata-kata umum dan yang sering digunakan pada dalam suatu bahasa namun, kata-kata tersebut tidak memiliki makna dalam klasifikasi. Data *stopword* dapat ditambah, diedit dan dihapus oleh admin. Tampilan halaman *stopword* dapat dilihat pada gambar 4.5 berikut :



Gambar 4.5. Halaman Stopword

Berikut ini merupakan *source code* pada halaman *stopword* :

```
<?php
switch(getPageSub()) {
case 'tambah' :
    $kata      = isset($_POST['input_kata'])?$_POST['input_kata']:'';
    if(isset($_POST['input_simpan'])) {
    $string     = "INSERT INTO `tbl_stopword` ( `word` ) VALUE ( '$kata' )";
    $q         = $koneksi->query($string);
    if ($q ) {
    header("Location: index.php?halaman=stopword");
    } else {
    $notif     = "Gagal menambahkan data";
    }
    }
    if (isset($notif)) {
    ?>
```

Modul Program 4.6. *Source Code* Halaman Stopword

```

<?php
    break;
    case 'ubah' :
        if(isset($_POST['input_ubah'])) {
            $kata = isset($_POST['input_kata'])?$_POST['input_kata']:'';
            $string = "UPDATE `tbl_stopword` SET `word`='$kata' WHERE
            `id_stopword`='".getID()."'";
            $q = $koneksi->query($string);
            if ($q ) {
                $notif = "Berhasil diperbaharui";
            } else {
                $notif = "Gagal memperbaharui data";
            }
        }
        if (isset($notif)) {
            ?>
            <div class="alert alert-danger alert-dismissible">
            <a href="#" class="close" data-dismiss="alert" aria-label="close">×</a>
            <center><?php echo $notif ?></center>
            </div>
            <?php
                }
            $string = "SELECT * FROM `tbl_stopword` WHERE
            `id_stopword`='".getID()."'";
            $q = $koneksi->query($string);
            if ($q && $q->rowCount () > 0) {
                $all = $q->fetchAll(PDO::FETCH_NUM);
                $d = $all[0];
                ?>
            <?php
            } else {
                echo "Data tidak ditemukan";
            }
            break;
            case 'hapus' :
            if(isset($_POST['input_hapus'])) {
                $id = getID();
                $string = "DELETE FROM `tbl_stopword` WHERE `id_stopword`='$id'";
                $q = $koneksi->query($string);
                if ($q ) {
                    $notif = "Berhasil menghapus data";
                } else {
                    $notif = "Gagal menghapus data";
                }
            }
            if (isset($notif)) {
                ?>
            <?php
            } else {
                echo "Data tidak ditemukan";
            } break;
            default :
                $string = "SELECT * FROM `tbl_stopword`";
                $q = $koneksi->query($string);
                if ($q && $q->rowCount () > 0) {
                    $all = $q->fetchAll(PDO::FETCH_NUM);
                } else {
                    $all = array();
                }
            ?>

```

Modul Program 4.7. Lanjutan *Source Code* Halaman Stopword

4.1.6. Halaman Kata Baku

Halaman kata baku merupakan halaman yang berisikan kata tidak baku dan kata baku. Pada halaman ini admin bisa menambah, edit dan menghapus kata baku. Tampilan halaman kata baku dapat dilihat pada gambar 4.6 berikut :

No	Kata Tidak Baku	Kata Baku	Aksi
1	trims	terima kasih	[Edit] [Delete]
2	g	tidak	[Edit] [Delete]
3	dgn	dengan	[Edit] [Delete]
4	cri	cari	[Edit] [Delete]
5	kpd	kepada	[Edit] [Delete]
6	msih	masih	[Edit] [Delete]
7	rmh	rumah	[Edit] [Delete]

Gambar 4.6. Halaman Kata Baku

Berikut ini merupakan *source code* halaman kata baku :

```
<?php
switch(getPageSub()) {
case 'tambah' :
    $kata = isset($_POST['input_kata'])?$_POST['input_kata']:'';
    $subs = isset($_POST['input_subs'])?$_POST['input_subs']:'';
    if(isset($_POST['input_simpan'])) {
        $string = "INSERT INTO `tbl_tidakbaku` ( `word`,`subs` ) VALUE (
        '$kata','$subs')";
        $q = $koneksi->query($string);
        if ($q) {
            header("Location: index.php?halaman=tidakbaku");
        } else { $notif = "Gagal menambahkan data";
        } } if (isset($notif)) {
    ?>
<?php
break;
case 'ubah' :
    if(isset($_POST['input_ubah'])) {
        $kata = isset($_POST['input_kata'])?$_POST['input_kata']:'';
        $string = "UPDATE `tbl_tidakbaku` SET `word`='$kata', `subs`='$subs'
        WHERE `id_tidakbaku`='".getID()."'";
```

Modul Program 4.8. Source Code Halaman Kata Baku

```

$qq          = $koneksi->query($string);
if ($qq ) {
$notif      = "Berhasil diperbaharui";
}
}
if (isset($notif)) {
?>
<?php
}
$string      = "SELECT * FROM `tbl_tidakbaku` WHERE
`id_tidakbaku`='".getID()."'";
$qq          = $koneksi->query($string);
if ($qq && $qq->rowCount () > 0) {
$all        = $qq->fetchAll(PDO::FETCH_NUM);
$d          = $all[0];
?>
<?php
} else {
echo "Data tidak ditemukan";
} break;
'hapus' :
if(isset($_POST['input_hapus'])) {
$id         = getID();
$string     = "DELETE FROM `tbl_tidakbaku` WHERE `id_tidakbaku`='$id'";
$qq        = $koneksi->query($string);
if ($qq ) {
$notif     = "Berhasil menghapus data";
} else {
$notif     = "Gagal menghapus data";
}
}
if (isset($notif)) {
?>
<?php
}

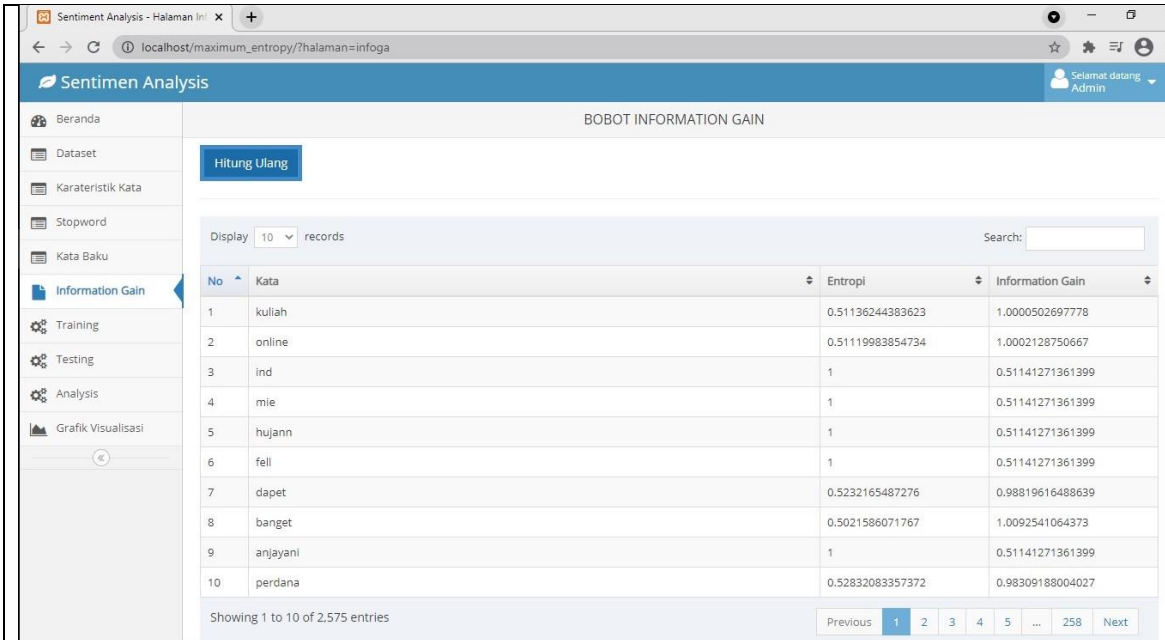
$string      = "SELECT * FROM `tbl_tidakbaku` WHERE
`id_tidakbaku`='".getID()."'";
$qq          = $koneksi->query($string);
if ($qq && $qq->rowCount () > 0) {
$all        = $qq->fetchAll(PDO::FETCH_NUM);
$d          = $all[0];
?>
<?php
} else {
echo "Data tidak ditemukan";
}
break;
default :
$string     = "SELECT * FROM `tbl_tidakbaku`";
$qq        = $koneksi->query($string);
if ($qq && $qq->rowCount () > 0) {
$all       = $qq->fetchAll(PDO::FETCH_NUM);
} else {
$all       = array();
}
?>

```

Modul Program 4.9. Lanjutan *Source Code* Halaman Kata Baku

4.1.7. Halaman Information Gain

Halaman *information gain* berisikan informasi bobot atau nilai *entropy* dan nilai *gain* pada setiap kata. *Information gain* merupakan metode *scoring* dalam menentukan atribut kontunu yang didiskritkan menggunakan nilai *entropy*. Tampilan halaman *information gain* dapat dilihat pada gambar 4.7 berikut :



No	Kata	Entropi	Information Gain
1	kuliah	0.51136244383623	1.0000502697778
2	online	0.51119983854734	1.0002128750667
3	ind	1	0.51141271361399
4	mie	1	0.51141271361399
5	hujann	1	0.51141271361399
6	fell	1	0.51141271361399
7	dapet	0.5232165487276	0.98819616488639
8	banget	0.5021586071767	1.0092541064373
9	anjayani	1	0.51141271361399
10	perdana	0.52832083357372	0.98309188004027

Gambar 4.7. Halaman *Information Gain*

Berikut ini merupakan *source code* halaman *information gain* :

```
<?php
$kamus_master          = generateKataDasar();
include_once "stemming/enhanced_CS.php";
include_once "Encoding.php";
use \ForceUTF8\Encoding;
$next                  = isset($_GET['next'])?$_GET['next']:'';
$current                = isset($_GET['current'])?$_GET['current']:'';
$step                  = isset($_GET['step'])?$_GET['step']:0;
$totalstep              = isset($_GET['totalstep'])?$_GET['totalstep']:0;
$status                 = loadStatus("training");
$setCurrent            = $status['current'];
if($totalstep!=$status['totalstep']) {
    $totalstep          = $status['totalstep'];
} else {
    $totalstep          = $totalstep;
}
if($setCurrent==$next && $status['status']==false) {
    $proses              = $setCurrent;
} else {
    $proses              = $next;
}
}
```

Modul Program 4.10. *Source Code* Halaman *Information Gain*

```

switch($proses) {
    case 'load' :
        $data      = loadDataset();
        //echo var_dump($data);
        $string    = "UPDATE `tbl_info_word` SET entropi='0',
infogain='0'";
        $q        = $koneksi->query($string);
        $string = "TRUNCATE TABLE `tbl_tfidf`";
        $q        = $koneksi->query($string);
        $string = "TRUNCATE TABLE `tbl_info_word`";
        $q        = $koneksi->query($string);
        $total    = count($data);
        $straining = round(0.8*$total);
        $content  = array();
        $num      = 0;
        foreach((array)$data as $dkf) {
            if($num<$straining) {
                $content[$dkf[0]]['asli']          = utf8size($dkf[1]);
                $content[$dkf[0]]['asli']          =
Encoding::toUTF8($content[$dkf[0]]['asli']);
                $content[$dkf[0]]['asli']          =
Encoding::fixUTF8($content[$dkf[0]]['asli']);
                //echo var_dump($content[$dkf[0]]['asli']);
                $content[$dkf[0]]['sentimen'] = $dkf[2];
                $content[$dkf[0]]['kategori'] = $dkf[3];
                $content[$dkf[0]]['hasil']      =
strtolower($content[$dkf[0]]['asli']);
                $content[$dkf[0]]['hasil']      =
remove_emoji($content[$dkf[0]]['hasil']);
                $content[$dkf[0]]['hasil']      = preg_replace('/@([\w-]+)/i', '',
$content[$dkf[0]]['hasil'] );
                $content[$dkf[0]]['hasil']      = preg_replace('/[0-9]/i', '',
$content[$dkf[0]]['hasil'] );
                $pattern = "/(?:)\b(?:https?://|www\d{0,3}[.]|[a-z0-9.-]+[.][a-
z]{2,4}\b)(?:[^\s()<>+|\\((([^\s()<>+|\\((([^\s()<>+|\\)))*\\))+(?:\\((([^\s(
)<>+|\\((([^\s()<>+|\\)))*\\)|[^\s!()\[\]{};:\\".,<?>&*)])?)/";
                $content[$dkf[0]]['hasil']      = preg_replace($pattern,"",
$content[$dkf[0]]['hasil'] );
                $content[$dkf[0]]['hasil']      =
preg_replace('/[\^\$\%\^&*()]{@\#~?><>\.!\|\\=\\-\\_\\+\\-\\`\\}\/', " ",
$content[$dkf[0]]['hasil']);
            } else {
                break; }
            $num++;
        }
        $json      = json_encode($content);
        //echo var_dump($json);
        $simpan    = saveToFile("training-load", $json);
        if($simpan) {
            $return = array("status"=>"true","notifikasi"=>"Proses Case Folding
berhasil, melanjutkan ke langkah berikutnya","halaman"=>"infoga",
"next"=>"tokenization", "step"=>'0', "totalstep"=>$straining);
            insertStatus("training", 'load', 'tokenization', '0', $straining, true);
        } else {
            $return= array("status"=>false,"notifikasi"=>"Proses Case Folding
gagal");
            insertStatus("training", 'load', 'load', '0', $total, false);
        }
        echo json_encode($return);
        break;
}

```

```

case 'tokenization' :
$training_start = loadFile("training-load");
$data           = json_decode($training_start,true);
$save_json      = loadFile("training-token");
$saved          = json_decode($save_json,true);
$num           = 0;
if(empty($saved) || $step==0) {
$saved         = array();
}
foreach((array)$data as $dtk => $dt) {
if($step==$num) {
$content      = $dt['hasil'];
$dumpwords    = explode(" ", $content);
foreach((array)$dumpwords as $dw) {
if(!empty($dw)) {
$saved[$dtk][] = $dw;
}}}
$num++;
}
$simpan       = saveToFile("training-token", json_encode($saved));
$pcs          = $step+1;
if($step==($totalstep-1)) {
    $nxt = "stopword";
    $stp = 0;
} else {
    $nxt = "tokenization";
    $stp = $step+1;
}
//echo var_dump($totalstep);
if($simpan) {
$return = array("status"=>"true","notifikasi"=>"Melakukan tokenizing
$pcs/$totalstep ", "halaman"=>"infoga", "next"=>$nxt, "step"=>$stp,
"totalstep"=>$totalstep);
insertStatus("training", 'tokenization', $nxt, $stp, $totalstep, true);
} else {
$return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("training", 'tokenization', 'tokenization', $stp,
$totalstep,
false); }
echo json_encode($return);
break;
case 'stopword' :
$data_start    = loadFile("training-token");
$data          = json_decode($data_start,true);
$save_json     = loadFile("training-stopword");
$saved         = json_decode($save_json,true);
$stopword      = loadStopword();
$num           = 0;
if(empty($saved) || $step==0) {
    $saved         = array();
}
foreach((array)$data as $dtk => $dt) {
    if($step==$num) {
        foreach((array)$dt as $dm) {
            $stop = false;

```

Modul Program 4.12. Lanjutan *Source Code* Halaman *Information Gain*


```

foreach((array)$stopword as $stw) {
    if($stw==$dm) {
        $stop = true;
    }
    break;
}
if($stop==false) {
    $saved[$dtk][] = $dm;
}
}
$num++;
}
    $simpan = saveToFile("training-stopword",
json_encode($saved));
    $pcs = $step+1;
    if($step==($totalstep-1)) {
        $nxt = "stemming";
        $stp = 0;
    } else {
        $nxt = "stopword";
        $stp = $step+1;
    }
}
if($simpan) {
    $return =
array("status"=>"true","notifikasi"=>"Melakukan stopword
$pcs/$totalstep","halaman"=>"infoga", "next"=>$nxt, "step"=>$stp,
"totalstep"=>$totalstep);
insertStatus("training", 'stopword', $nxt, $stp, $totalstep, true);
} else {
$return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("training", 'stopword', 'stopword', $stp, $totalstep,
false);
} echo json_encode($return);
break;
case 'stemming' :
    $data_stopword = loadFile("training-stopword");
    $data = json_decode($data_stopword,true);
    $save_json = loadFile("training-stemming");
    $saved = json_decode($save_json,true);
    $num = 0;
    if(empty($saved) || $step==0) {
        $saved = array();
    }
    foreach((array)$data as $dtk => $dt) {
        if($step==$num) {
            foreach((array)$dt as $dm) {
                if($dm!=' ' && $dm!='') {
                    $dm = Enhanced_CS($dm);
                    if(preg_match('/^[a-z]+$/',$dm)) {
                        $saved[$dtk][] = $dm;
                    }
                }
            }
        }
    }
    $num++;
}
    $simpan = saveToFile("training-stemming",
json_encode($saved));
    $pcs = $step+1;

```

Modul Program 4.13. Lanjutan Source Code Halaman Information Gain

```

if($step==($totalstep-1)) {
    $nxt = "idf";
    $stp = 0;
} else {
    $nxt = "stemming";
    $stp = $step+1;
}

    if($simpan) {
$return = array("status"=>"true","notifikasi"=>"Melakukan Stemming
$pcs/$totalstep","halaman"=>"infoga", "next"=>$nxt, "step"=>$stp,
"totalstep"=>$totalstep);
insertStatus("training", 'stemming', $nxt, $stp, $totalstep, true);
    } else {
$return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("training", 'stemming', 'stemming', $stp, $totalstep,
false);
    }
echo json_encode($return);
break;
case 'idf' :
    $data_stem = loadFile("training-stemming");
    $data = json_decode($data_stem,true);
    $data_load = loadFile("training-load");
    $dalo = json_decode($data_load,true);
    $save_json = loadFile("training-idf");
    $saved = json_decode($save_json,true)
    if(empty($saved) || $step==0) {
$saved = array();
    }
$num = 0;
    foreach((array)$data as $dtk => $dt) {
    if($step==$num) {
    $docid = $dtk;
    foreach((array)$dt as $dm) {
    if(isset($saved)) {
    if(isset($saved[$dm]['nilai'][$docid])) {
    $count = $saved[$dm]['nilai'][$docid];
    $count += 1;
    } else {
    $count = 1;
    $saved[$dm]["kelas"][$docid] =
isset($dalo[$dtk]['sentimen'])? $dalo[$dtk]['sentimen']:0;
    }
    $saved[$dm]['nilai'][$docid] = $count;
    } else {
    $count = 1;
    $saved[$dm]['nilai'][$docid] = $count;
    $saved[$dm]["kelas"][$docid] =
isset($dalo[$dtk]['sentimen'])? $dalo[$dtk]['sentimen']:0;
    }
    }
    }
    $num++;
    }
    $simpan = saveToFile("training-idf", json_encode($saved));
if($step==($totalstep-1)) {
    $nxt = "save-idf";
    $stp = 0;
}

```

Modul Program 4.14. Lanjutan *Source Code* Halaman *Information Gain*

```

} else {
    $nxt = "idf";
    $stp = $step+1;
}
if($simpan) {
$return = array("status"=>"true","notifikasi"=>"Melakukan
perhitungan idf $stp/$totalstep","halaman"=>"infoqa", "next"=>$nxt,
"step"=>$stp, "totalstep"=>$totalstep);
insertStatus("training", 'idf', $nxt, $stp, $totalstep, true);
} else {
$return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$stp");
insertStatus("training", 'idf', 'idf', $stp, $totalstep, false);
}
echo json_encode($return);
break;
case 'save-idf' :
    $data_index = loadFile("training-idf");
    $data_index = json_decode($data_index,true);
    $indval = array();
    foreach((array)$data_index as $dik => $di) {
        //echo var_dump($di);
        foreach((array)$di['nilai'] as $dock => $doc) {
            $kelas = isset($di['kelas'][$dock])?$di['kelas'][$dock]:'0';
            $indval[] = "('{$dik}','{$doc}','{$dock}','{$kelas}')";
        }
    }
    $string = "INSERT INTO `tbl_tfidf` ( `term`,`count`,`docid`,`kelas`
) VALUES ".implode(", ", $indval);
    $q = $koneksi->query($string);
    $nxt = "tfidf";
    $stp = 0;
    $notif = "Menyimpan nilai IDF";
    if($q) {
        $return =
array("status"=>"true","notifikasi"=>$notif,"halaman"=>"infoqa",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("kamus", 'db-slang', $nxt, $stp, $totalstep, true);
    } else {
$return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("kamus", 'db-slang', 'db-slang', $stp, $totalstep, false);
    }
    echo json_encode($return);
    break;
case 'tfidf' :
    $data_index = loadFile("training-idf");
    $data_index = json_decode($data_index,true);
    if($step==0) {
        $totalstep = count($data_index);
        $tfidf = array();
    } else {
        $tfidf = loadFile("training-tfidf");
        $tfidf = json_decode($tfidf,true);
    }
    $num = 0;
    foreach((array)$data_index as $dtk => $dt) {
        if($step==$num) {

```

Modul Program 4.15. Lanjutan *Source Code* Halaman *Information Gain*

```

$dump      = array();
$termexis  = array();
$df         = array();
    $tempdoc  = "";
    $tempcount = 0;
    $itung    = 1;
$stringdo   = "SELECT a.*, (SELECT SUM(b.count) FROM `tbl_tfidf` AS b
WHERE b.`term`='$dtk') AS n, (SELECT COUNT(*) FROM `tbl_dataset`)*0.8 AS
N FROM `tbl_tfidf` AS a WHERE term='$dtk'"; //echo var_dump($stringdo);
    $qdo      = $koneksi->query($stringdo);
    $datao    = $qdo->fetchAll(PDO::FETCH_NUM);
foreach((array)$datao as $dto) {
    $df[]     = $dto[5];
    $N        = floor($dto[6]);
}
//echo var_dump($document);
$hitidf     = rumusIdf(array_sum($df), $N);
$dump['term'] = $dtk;
$dump['df']   = count($df);
$dump['idf']  = $hitidf;
$stringi     = "INSERT INTO `tbl_tfidfhasil` ( `term`,`df`,`idf` )
VALUES ( '$dtk','".count($df)."', '$hitidf')";
$qi         = $koneksi->query($stringi);
$tfidf[]    = $dump;
}
    $num++;
}
if($step==($totalstep-1)) {
    $nxt = "entropi-total";
    $stp = 0;
    $notif = "Melakukan perhitungan normalisasi TF-IDF seluruh word";
} else {
    $nxt = "tfidf";
    $stp = $step+1;
    $notif = "Melakukan penyusunan matriks TF-IDF seluruh word";
}
$simpan1 = saveToFile("training-tfidf", json_encode($tfidf));
if($simpan1) {
    $return =
array("status"=>"true", "notifikasi"=>$notif, "halaman"=>"infoga",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("training", 'tfidf', $nxt, $stp, $totalstep, true);
} else {
    $return = array("status"=>false, "notifikasi"=>"Proses gagal data :
$stp");
insertStatus("training", 'tfidf', 'tfidf', $stp, $totalstep, false);
}
echo json_encode($return);
break;
case 'entropi-total' :
    $data_index = loadFile("training-idf");
    $data_index = json_decode($data_index, true);
    $dataset = loadFile("training-load");
    $dataset = json_decode($dataset, true);
    if($step==0) {
    $totalstep = count($data_index);
    $entropi = array();
} else {

```

```

} else {
    $entropi          = loadFile("training-entropy");
    $entropi          = json_decode($entropi,true);
}
//melakukan foreach untuk mengetahui entropy S
$spos               = 0;
$snet               = 0;
$sneg              = 0;
$totaldataset       = count($dataset); //80
$num                = 0;
foreach((array)$dataset as $ds) {
//echo var_dump($ds['sentimen'])."<br />";
if($ds['sentimen']==1) {
    $spos += 1;
} else if($ds['sentimen']==-1) {
    $sneg += 1;
} else {
    $snet += 1;
}
}
// menghitung entropy S dataset
$logpos            = -($spos/$totaldataset)*log($spos/$totaldataset,2);
$lognet            =
($snet/$totaldataset)*log($snet/$totaldataset,2);
$logneg            = -($sneg/$totaldataset)*log($sneg/$totaldataset,2);
$entropyS          = $logpos+$lognet+$logneg;
$num               = 0;
foreach((array)$data_index as $dtk => $dt) {
if($step==$num) {
$total             = count($dt['kelas']);
//echo var_dump($total);
    $pos            = 0;
    $net            = 0;
    $neg            = 0;
foreach((array)$dt['kelas'] as $ke) {
    if($ke==1) {
        $pos += 1;
    } else if($ke==-1) {
        $neg += 1;
    } else {
        $net += 1;
    }
}
// menghitung entropy tiap kata
    $klogpos        = 0;
    $klognet         = 0;
    $klogneg         = 0;
if($pos>0) {
    $klogpos         = -($pos/$total)*log($pos/$total,2);
}
if($net>0) {
    $klognet         = -($net/$total)*log($net/$total,2);
}
if($neg>0) {
    $klogneg         = -($neg/$total)*log($neg/$total,2);
}
// menghitung Gain
    $minpos          = 0;
    $minnet           = 0;
    $minneg           = 0;

```

```

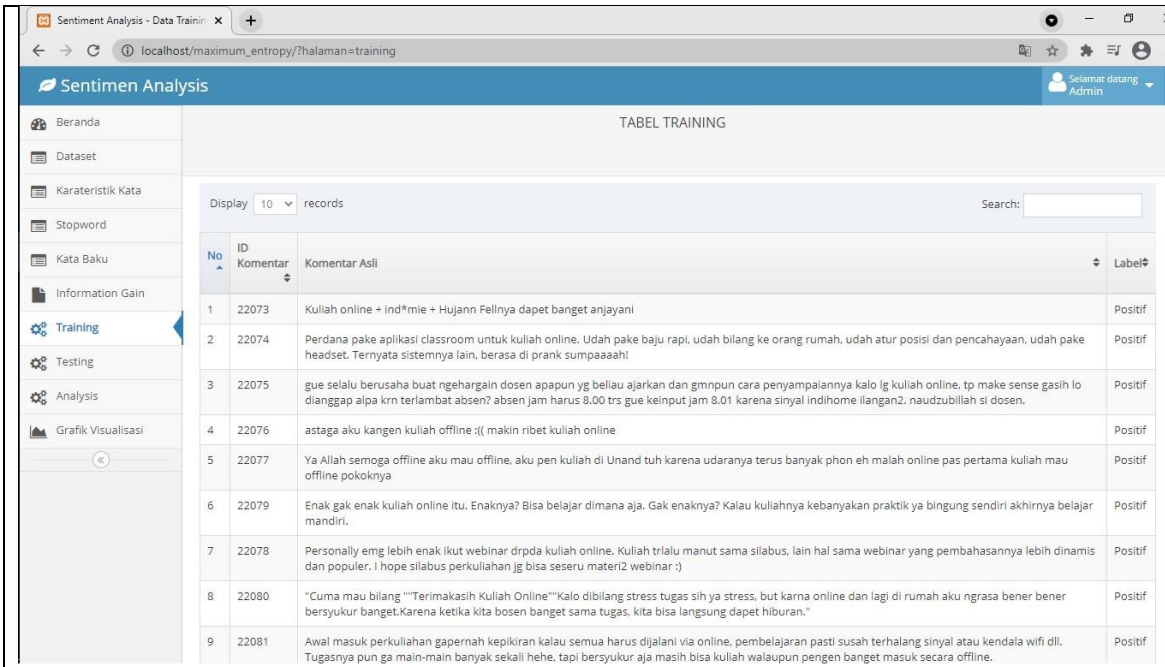
if($pos>0) {
    $minpos      = ($pos/$total)*$klogpos;
}
if($net>0) {
    $minnet      = ($net/$total)*$klognet;
}
if($neg>0) {
    $minneg      = ($neg/$total)*$klogneg;
}
if(count($dt['kelas'])==1) {
    $totmin      = 1;
} else {
    $totmin      = $minpos+$minnet+$minneg;
}
//echo var_dump($totmin);
$gain          = $entropyS-$totmin;
//echo var_dump($gain);
$stringi       = "INSERT INTO `tbl_info_word` (
`word`,`entropi`,`infogain`) VALUES ('$dtk','$totmin','$gain')";
//echo var_dump($stringi);
$q            = $koneksi->query($stringi);
if($qi) {
    $hasil      = true;
} else {
    $hasil      = false;
}}
$num++;}
if($step==($totalstep-1)) {
    $nxt        = "finish";
    $stp        = 0;
    $notif      = "Proses selesai";
} else {
    $nxt        = "entropi-total";
    $stp        = $step+1;
    $notif      = "Melakukan seleksi information gain";
} if($hasil) {
    $return     =
array("status"=>"true","notifikasi"=>$notif,"halaman"=>"infoga",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("training", 'entropi-total', $nxt, $stp, $totalstep, true);
} else {
    $return     = array("status"=>false,"notifikasi"=>"Proses gagal data :
    $stp");
insertStatus("training", 'entropi-total', 'entropi-total', $stp,
$totalstep, false);
} echo json_encode($return);
break;
case 'finish' :
    $nxt        = "finish";
    $stp        = 0;
    $notif      = "Proses telah selesai, halaman akan direfresh";
    $return     =
array("status"=>"true","notifikasi"=>$notif,"halaman"=>"kamus",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("kamus", 'finish', $nxt, $stp, $totalstep, true);
echo json_encode($return);
break;}
?>

```

Modul Program 4.18. Lanjutan *Source Code* Halaman *Information Gain*

4.1.8. Halaman Training

Halaman *training* merupakan halaman tabel data latih yang berisikan komentar yang sudah diberi label. Tampilan pada halaman *training* dapat dilihat pada gambar 4.8 berikut :



The screenshot shows a web browser window with the URL `localhost/maximum_entropy/?halaman=training`. The page title is "Sentimen Analysis" and the user is logged in as "Admin". The main content area is titled "TABEL TRAINING" and displays a table with 9 records. The table has columns for "No", "ID Komentar", "Komentar Asli", and "Label". All labels in the table are "Positif".

No	ID Komentar	Komentar Asli	Label
1	22073	Kuliah online + ind*mie + Hujann Fellnya dapet banget anjayani	Positif
2	22074	Perdana pake aplikasi classroom untuk kuliah online. Udah pake baju rapi, udah bilang ke orang rumah, udah atur posisi dan pencahayaan, udah pake headset. Ternyata sistemnya lain, berasa di prank sumpaaaaah!	Positif
3	22075	gue selalu berusaha buat ngehargain dosen apapun yg beliau ajarkan dan gmnpun cara penyampaiannya kalo lg kuliah online, tp make sense gash lo dianggap alpa krn terlambat absen? absen jam harus 8.00 trs gue keinput jam 8.01 karena sinyal indihome ilangan2. naudzubillah si dosen.	Positif
4	22076	astaga aku kangen kuliah offline :(makin ribet kuliah online	Positif
5	22077	Ya Allah semoga offline aku mau offline, aku pen kuliah di Unand tuh karena udaranya terus banyak phon eh malah online pas pertama kuliah mau offline pokoknya	Positif
6	22079	Enak gak enak kuliah online itu, Enaknya? Bisa belajar dimana aja. Gak enaknya? Kalau kuliahnya kebanyakan praktik ya bingung sendiri akhirnya belajar mandiri.	Positif
7	22078	Personally emg lebih enak ikut webinar drpd kuliah online. Kuliah trialu manut sama silabus, lain hal sama webinar yang pembahasannya lebih dinamis dan populer. I hope silabus perkuliahan jg bisa seseru materi2 webinar :)	Positif
8	22080	"Cuma mau bilang""Terimakasih Kuliah Online""Kalo dibilang stress tugas sih ya stress, but karna online dan lagi di rumah aku ngrasa bener bener bersyukur banget.Karena ketika kita bosan banget sama tugas, kita bisa langsung dapet hiburan."	Positif
9	22081	Awal masuk perkuliahan gapernah kepikiran kalau semua harus dijalani via online, pembelajaran pasti susah terhalang sinyal atau kendala wifi dll. Tugasnya pun ga main-main banyak sekali hehe, tapi bersyukur aja masih bisa kuliah walaupun pengen banget masuk secara offline.	Positif

Gambar 4.8. Halaman Training

Berikut ini merupakan *source code* halaman *training* :

```
<?php
$string = "SELECT a.*,b.`id_uji` FROM tbl_dataset AS a
LEFT JOIN `tbl_komentar_uji` AS b ON a.`id`=b.`id_komentar` WHERE
a.`id`=b.`id_komentar`";
$q = $koneksi->query($string);
$f = $q->fetchAll(PDO::FETCH_NUM);
$training = round(0.8*count($f));
$all = array();
$num = 0;
foreach((array)$f as $d) {
if($num<$training) {
$all[] = $d;
}
$num++;
}
?>
<?php
$z = 1;
foreach((array)$all as $d) {
?>
<tr>
<td><?php echo $z ?></td>
<td><?php echo $d[0] ?></td>
```

Modul Program 4.19. Source Code Halaman Training

```

<td><?php echo $d[1] ?></td>
<td><?php
if ($d[2]==1) {
    echo "Positif";
}
if ($d[2]==-1) {
    echo "Negatif";
}
if ($d[2]==0) {
    echo "Netral";
}
?>

```

Modul Program 4.20. Lanjutan Source Code Halaman Training

4.1.9. Halaman Testing

Halaman *testing* merupakan halaman tabel data uji yang sudah mendapatkan hasil klasifikasi dan perhitungan nilai probabilitas dari komentar dengan menggunakan metode *maximum entropy*. Tampilan halaman *testing* dapat dilihat pada gambar 4.9 berikut :

No	Komentar Asli	Text Processing	Label	Probabilitas Positif	Probabilitas Netral	Probabilitas Negatif	Hasil Sentimen
66	Ketika adek sepupu lagi kuliah online... jd pengen kuliah lagi	adek sepupu kuliah online jd pengen kuliah	POSITIF	0.093818202606676	0.047780425789793	0.080424836442812	POSITIF
228	Buat adek2 yg masih sekolah atau kuliah dan harus belajar daring, semangat. Yuk bisa yuk. Sama2 berjuang di sini. Guru berjuang, siswa juga berjuang. Bisa koq. Semua ini pasti terlewat!	adek yg seko kuliah ajar daring semangat yuk yuk berjuang guru berjuang siswa berjuang koq terlewat!	POSITIF	0.034566806025025	0.015775878215355	0.030086194182546	POSITIF
189	Akuu selama kuliah online kalo tanggal merah gapernah ada kelas pengganti yeayyy..	akuu kuliah online kalo tanggal merah gapernah kelas pengganti yeayyy	POSITIF	0.080502567205012	0.02509758498849	0.067452777047271	POSITIF
207	Andai semua dosen kek gini Peduli akan tanggung jawab, kita tau sama sama kalau kuliah daring kurang efektif, tapi tdk ada solusi lain dengan kondisi seperti ini. Kadang ada dosen yang cmn ngasih tugas selama pertemuan, yg terkesan tdk peduli kepada mah	andai dosen kek gini peduli tanggung tau kuliah daring efektif tdk solusi kondisi kadang dosen cmn ngasih tugas pertemuan yg terkesan tdk peduli mah	POSITIF	0.024907609745488	0.012504185638704	0.022065510351861	POSITIF

Gambar 4.9. Halaman Testing

Berikut ini merupakan *source code* halaman *testing* :

```

<?php
$kamus_master = generateKataDasar();
include_once "stemming/enhanced_CS.php";
include_once "Encoding.php";
use \ForceUTF8\Encoding;
$next = isset($_GET['next'])?$_GET['next']:'';
$current = isset($_GET['current'])?$_GET['current']:'';
$step = isset($_GET['step'])?$_GET['step']:0;
$totalstep = isset($_GET['totalstep'])?$_GET['totalstep']:0;
$status = loadStatus("testing");

```

Modul Program 4.21 Source Code Halaman Testing


```

$setCurrent = $status['current'];
if($totalstep!=$status['totalstep']) {
$totalstep = $status['totalstep'];
} else {
$totalstep = $totalstep;
}
if($setCurrent==$next && $status['status']==false) {
$proses = $setCurrent;
} else {
$proses = $next;
}
switch($proses) {
case 'load' :
$data = loadDataset();
$stringdel = "DELETE FROM `tbl_analysis` WHERE `dataset`>0";
$qdel = $koneksi->query($stringdel);
$stringal = "ALTER TABLE tbl_analysis AUTO_INCREMENT=0";
$qal = $koneksi->query($stringal);
$total = count($data);
$training = round(0.2*$total);
$testing = $total-$training;
$content = array();
$num = 0;
foreach((array)$data as $dkf) {
if($num>=$testing) {
$content[$dkf[0]]['asli'] = utf8size($dkf[1]);
$content[$dkf[0]]['asli'] =
Encoding::toUTF8($content[$dkf[0]]['asli']);
$content[$dkf[0]]['asli'] =
Encoding::fixUTF8($content[$dkf[0]]['asli']);
//echo var_dump($content[$dkf[0]]['asli']);
$content[$dkf[0]]['sentimen'] = $dkf[2];
$content[$dkf[0]]['kategori'] = $dkf[3];
$content[$dkf[0]]['hasil'] = strtolower($content[$dkf[0]]['asli']);
$content[$dkf[0]]['hasil'] =
remove_emoji($content[$dkf[0]]['hasil']);
$content[$dkf[0]]['hasil'] = preg_replace('/@([\w-]+)/i', '',
$content[$dkf[0]]['hasil'] );
$content[$dkf[0]]['hasil'] = preg_replace('/[0-9]/i', '',
$content[$dkf[0]]['hasil'] );
$pattern = "/(?:)\b(?:https?:\/\//|www\d{0,3}[.]| [a-z0-9.-]+[.] [a-z]{2,4}\.\/)(?:[^\s()<>+|\\([^\s()<>+|\\([^\s()<>+\\)])*)\\)+(?:\\([^\s()<>+|\\([^\s()<>+\\)])*)| [^\s`!\\[\]{};:'\".,<>?*\"])/";
$content[$dkf[0]]['hasil'] = preg_replace($pattern,"",
$content[$dkf[0]]['hasil'] );
$content[$dkf[0]]['hasil'] =
preg_replace('/[^\\"£$%^&* ()]{@:\#~?><.\,!;\|\\=\-_\+~\`\\}\/', " ",
$content[$dkf[0]]['hasil']);
}
$num++;
}
$json = json_encode($content);
//echo var_dump($json);
$simpan = saveToFile("testing-load", $json);
if($simpan) {
$return = array("status"=>"true", "notifikasi"=>"Proses Case Folding
berhasil, melanjutkan ke langkah
berikutnya", "halaman"=>"testing", "next"=>"tokenization", "step"=>'0',
"totalstep"=>$testing);
}

```

```

insertStatus("testing",'load', 'tokenization', '0', $testing, true);
} else {
$return      = array("status"=>false,"notifikasi"=>"Proses Case Folding
gagal");
insertStatus("testing",'load', 'load', '0', $testing, false);
}
echo json_encode($return);
break;
case 'tokenization' :
    $straining_start  = loadFile("testing-load");
    $data              = json_decode($straining_start,true);
    $save_json         = loadFile("testing-token");
    $saved              = json_decode($save_json,true);
    $num               = 0;
if(empty($saved) || $step==0) {
    $saved              = array();
}
foreach((array)$data as $dtk => $dt) {
if($step==$num) {
    $content           = $dt['hasil'];
    $dumpwords         = explode(" ", $content);
foreach((array)$dumpwords as $dw) {
if(!empty($dw)) {
$saved[$dtk][ ]       = $dw;
}
}
}
}
$num++;
}
$simpan              = saveToFile("testing-token", json_encode($saved));
$pcs                  = $step+1;
if($step==($totalstep-1)) {
    $nxt               = "stopword";
    $stp               = 0;
} else {
    $nxt               = "tokenization";
    $stp               = $step+1;
}
//echo var_dump($totalstep);
if($simpan) {
$return              = array("status"=>"true","notifikasi"=>"Melakukan
tokenizing $pcs/$totalstep", "halaman"=>"testing","next"=>$nxt,
"step"=>$stp, "totalstep"=>$totalstep);
insertStatus("testing",'tokenization', $nxt, $stp, $totalstep, true);
} else {
$return              = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("testing",'tokenization', 'tokenization', $stp, $totalstep,
false);
}
echo json_encode($return);
break;
case 'stopword' :
    $data_start        = loadFile("testing-token");
    $data               = json_decode($data_start,true);
    $save_json          = loadFile("testing-stopword");
    $saved               = json_decode($save_json,true);
    $stopword           = loadStopword();

```

Modul Program 4.23. Lanjutan *Source Code* Halaman *Testing*

```

$num = 0;
if(empty($saved) || $step==0) {
    $saved = array();
}
foreach((array)$data as $dtk => $dt) {
    if($step==$num) {
        foreach((array)$dt as $dm) {
            $stop = false;
            foreach((array)$stopword as $stw) {if($stw==$dm) {
                $stop = true;
                break;
            }
        }
        if($stop==false) {
            $saved[$dtk][] = $dm;
        }
    }
    $num++;
}
$simpan = saveToFile("testing-stopword", json_encode($saved));
$pcs = $step+1;
if($step==($totalstep-1)) {
    $nxt = "stemming";
    $stp = 0;
} else {
    $nxt = "stopword";
    $stp = $step+1;
}
if($simpan) {
    $return = array("status"=>"true","notifikasi"=>"Melakukan stopword
    $pcs/$totalstep","halaman"=>"testing","next"=>$nxt,"step"=>$stp,
    "totalstep"=>$totalstep);
    insertStatus("testing",'stopword', $nxt, $stp, $totalstep, true);
} else {
    $return = array("status"=>false,"notifikasi"=>"Proses gagal data :
    $step");
    insertStatus("testing",'stopword', 'stopword', $stp, $totalstep, false);
}
echo json_encode($return);
break;
case 'stemming' :
    $data_stopword = loadFile("testing-stopword");
    $data = json_decode($data_stopword,true);
    $save_json = loadFile("testing-stemming");
    $saved = json_decode($save_json,true)
$num = 0;
if(empty($saved) || $step==0) {
    $saved = array();
}
foreach((array)$data as $dtk => $dt) {
    if($step==$num) {
        foreach((array)$dt as $dm) {
            if($dm!=' ' && $dm!='') {
                $dm = Enhanced_CS($dm);
                if(preg_match('/^[a-z]+$/',$dm)) {
                    $saved[$dtk][] = $dm;
                }
            }
        }
    }
    $num++;
}

```

Modul Program 4.24. Lanjutan *Source Code* Halaman *Testing*

```

$simpan      = saveToFile("testing-stemming", json_encode($saved));
$pcs        = $step+1;
if($step==($totalstep-1)) {
    $nxt     = "entropi";
    $stp     = 0;
} else {
    $nxt     = "stemming";
    $stp     = $step+1;
}
if($simpan) {
$return      = array("status"=>"true","notifikasi"=>"Melakukan Stemming
$pcs/$totalstep","halaman"=>"testing","next"=>$nxt,          "step"=>$stp,
"totalstep"=>$totalstep);
insertStatus("testing",'stemming', $nxt, $stp, $totalstep, true);
} else {
$return      = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("testing",'stemming', 'stemming', $stp, $totalstep, false);
}
echo json_encode($return);
break;
case 'entropi' :
    $data_stemming      = loadFile("testing-stemming");
    $data_stemming      = json_decode($data_stemming,true);
    $save_json          = loadFile("testing-entropi");
    $saved              = json_decode($save_json,true);
if(empty($saved) || $step==0) {
$save        = array();
}
$num = 0;
    $stemming_dt        = array();
foreach((array)$data_stemming as $dtk => $dt) {
    $stemming_dt        = implode(" ", $dt);
if($step==$num) {
    $string              = "SELECT * FROM `tbl_tfidf`";
$q              = $koneksi->query($string);
$f              = $q->fetchAll(PDO::FETCH_NUM);
// mencari entropinya
    $himpos           = array();
    $himnet           = array();
    $himneg           = array();
    $total            = count($f);
foreach((array)$dt as $dm) {
//echo "<br /><br />".var_dump($dm);
$string          = "SELECT * FROM `tbl_tfidf` WHERE `term`='".$dm'";
$q              = $koneksi->query($string);
$f              = $q->fetchAll(PDO::FETCH_NUM);
if(empty($f)) {
$himnet[$dm][ ] = 0;
}
foreach((array)$f as $df) {
//echo var_dump($df[4])."<br />";
if($df[4]==1) {
$himpos[$dm][ ] = $df[0];
//$total          =+ 1;
} else if($df[4]==-1) {
    $himneg[$dm][ ] = $df[0];
//$total          =+ 1;
}
}
}
}

```

Modul Program 4.25. Lanjutan Source Code Halaman Testing

```

}
}
//echo "<br />=====<br /><br />";
}
    $entpos      = array();
    $entnet      = array();
    $entneg      = array();
foreach((array)$himpos as $hp) {
    $entpos[]    = -(count($hp)/$total)*log(count($hp)/$total,2);
}
foreach((array)$himnet as $ht) {
    $entnet[]    = -(count($ht)/$total)*log(count($ht)/$total,2);
}
foreach((array)$himneg as $hg) {
    $entneg[]    = -(count($hg)/$total)*log(count($hg)/$total,2); }
// mencari nilai Z dan P
    $Zpos = array();
    $Zneg = array();
    $Znet = array();
    $epPos = array();
    $epNet = array();
    $epNeg = array();
foreach((array)$entpos as $ep) {
    $Zpos[]      = exp($ep);
}
foreach((array)$entnet as $ep) {
    $Znet[]      = exp($ep);
}
foreach((array)$entneg as $ep) {
    $Zneg[]      = exp($ep);
}
// mencari nilai P
foreach((array)$entpos as $ep) {
    $Ppos[] = (1/array_sum($Zpos))*exp($ep);
}
foreach((array)$entnet as $ep) {
    $Pnet[] = (1/array_sum($Znet))*exp($ep);
}
foreach((array)$entneg as $ep) {
    $Pneg[] = (1/array_sum($Zneg))*exp($ep);
}
// mencari propabilitas
    if(array_sum($Zpos)>0) {
        $probpos      = (1/array_sum($Zpos))*array_sum($entpos);
    } else {
        $probpos      = 0;
    }
    if(array_sum($Znet)>0) {
        $probnet      = (1/array_sum($Znet))*array_sum($entnet);
    } else {
        $probnet      = 0;
    }
    if(array_sum($Zneg)>0) {
        $probneg      =
(1/array_sum($Zneg))*array_sum($entneg);
    } else {
        $probneg      = 0;
    }
}

```

```

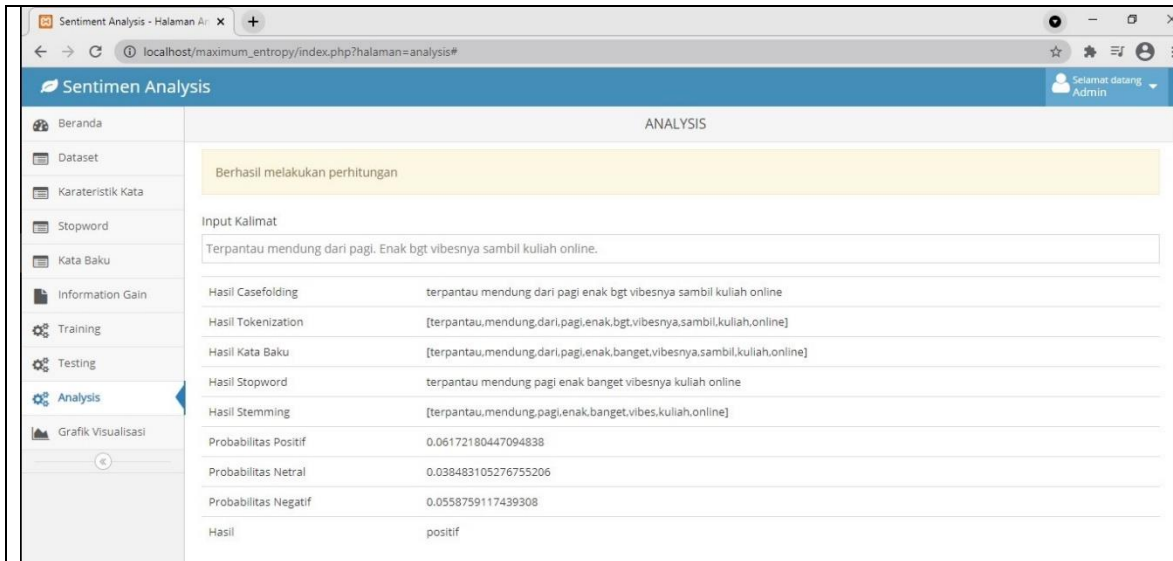
$saved[$dtk]['pos'] = $probpos;
    $saved[$dtk]['net'] = $probnet;
    $saved[$dtk]['neg'] = $probneg;
if($probpos>$probnet && $probpos>$probneg) {
    $hasil = "positif";
} else if($probneg>$probnet && $probneg>$probpos) {
    $hasil = "negatif";
} else {
    $hasil = "netral";
}
$stringi = "INSERT INTO `tbl_analysis` (
`dataset`,`prosestext`,`probpos`,`probnet`,`probneg`,`hasil` ) VALUES
('$dtk','$stemming_dt','$probpos','$probnet','$probneg','$hasil')";
$qi = $koneksi->query($stringi);
}
$num++;
}
$simpan = saveToFile("testing-entropi", json_encode($saved));
if($step==($totalstep-1)) {
    $nxt = "finish";
    $stp = 0;
    $notifikasi = "Berhasil melakukan perhitungan";
} else {
    $nxt = "entropi";
    $stp = $step+1;
    $notifikasi = "Melakukan perhitungan $stp/$totalstep";
}
if($simpan) {
    $return =
array("status"=>"true", "notifikasi"=>$notifikasi, "halaman"=>"testing", "
next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("testing",'idf', $nxt, $stp, $totalstep, true);
} else {
    $return = array("status"=>false, "notifikasi"=>"Proses gagal data :
    $stp");
insertStatus("testing",'idf', 'idf', $stp, $totalstep, false);
}
echo json_encode($return);
break;
}

```

Modul Program 4.27. Lanjutan *Source Code* Halaman *Testing*

4.1.10. Halaman Analysis

Halaman *analysis* merupakan halaman untuk melakukan pengujian serta melakukan perhitungan hasil sentimen dengan melakukan input kalimat, kalimat tersebut akan diproses oleh sistem. kemudian sistem akan menampilkan hasil dari kalimat yang sudah di *input* tersebut. Tampilan pada halaman *analysis* dapat dilihat pada gambar 4.10 berikut :



Gambar 4.10. Halaman *Analysis*

Berikut ini merupakan *source code* halaman *analysis* :

```

$totalstep = isset($_GET['totalstep'])?$_GET['totalstep']:0;
$status     = loadStatus("analisis");
$setCurrent = $status['current'];
if($totalstep!=$status['totalstep']) {
    $totalstep = $status['totalstep'];
} else {
    $totalstep = $totalstep;
}
if($setCurrent==$next && $status['status']==false) {
    $proses     = $setCurrent;
} else {
    $proses     = $next;
}
switch($proses) {
    case 'case-folding' :
        $content['asli'] = $word;
        $content['hasil'] = strtolower($content['asli']);
        $content['hasil'] = remove_emoji($content['hasil']);
        $content['hasil'] = preg_replace('/@([\w-]+)/i', '', $content['hasil']
);
        $content['hasil'] = preg_replace('/[0-9]/i', '', $content['hasil'] );
        $pattern     =  "/(?:)\b(?:https?:\/\|www\d{0,3}[.][a-z0-9.-]+\.[a-
z]{2,4}\|)(?:[^\s()<>+|\\(\([^\s()<>+|\\(\([^\s()<>+|\\(\([^\s(
)<>+|\\(\([^\s()<>+|\\)\)\)\)\)\)]*\\)|[^\s`!()\[\]{};:'\".,<?&quot;
)"/";
        $content['hasil'] = preg_replace($pattern,"", $content['hasil'] );
        $content['hasil'] = preg_replace("/[^\p{L}\p{N}]+/", "",
$content['hasil']);
        $content['id'] = 0;
        $simpan=saveToFile("analisis-casefolding", json_encode($content));
        $nxt = "tokenisasi";
        $stp = 0;
        if($simpan) {
            $return = array("status"=>"true","notifikasi"=>"Proses sedang
berlangsung, melanjutkan ke langkah berikutnya","halaman"=>"analisis",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep, "casefolding" =>
$content['hasil']);

```

Modul Program 4.27. *Source Code* Halaman *Analysis*

```

}
    echo json_encode($return);
break;
case 'tokenisasi' :
// tokenisasi yaitu memecah kalimat menjadi susunan kata
    $data = loadFile("analisis-casefolding");
    $data = json_decode($data,true);
    $content = $data['hasil'];
    $dumpwords = explode(" ", $content);
    $simpan = saveToFile("analisis-tokenisasi",
json_encode($dumpwords));
    $nxt = "sinonim";
    $stp = 0;
    if($simpan) {
    $return = array("status"=>"true","notifikasi"=>"Proses sedang
berlangsung, melanjutkan ke langkah berikutnya","halaman"=>"analisis",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep);
insertStatus("analisis", 'tokenisasi', $nxt, $stp, $totalstep, true);
} else {
    $return = array("status"=>false,"notifikasi"=>"Proses gagal data :
$step");
insertStatus("analisis", 'tokenisasi', 'tokenisasi', $stp, $totalstep,
false);
}
    echo json_encode($return);
break;
case 'sinonim' :
    $data_token = loadFile("analisis-tokenisasi");
    $data_token = json_decode($data_token,true);
    $saved = loadFile("analisis-sinonim");
    $saved = json_decode($saved,true);
    $tdkbaku = loadTidakBaku();
    $num = 0;
if(empty($saved) || $step==0) {
    $totalstep = count($data_token);
    $saved = array();
}
foreach((array)$data_token as $dtk) {
//if($step==$num) {
foreach($tdkbaku as $dtk) {
    if($dtk[1]==$dtk) {
        $dtk = $dtk[2];
    }}
    $saved[] = $dtk;
}

    $stepk = $step+1;
    $simpan = saveToFile("analisis-sinonim", json_encode($saved));
    $nxt = "stopword";
    $stp = 0;
if($simpan) {
    $return = array("status"=>"true","notifikasi"=>"Melakukan Sinonim
$stepk/$totalstep","halaman"=>"analisis", "next"=>$nxt, "step"=>$stp,
"totalstep"=>$totalstep);
insertStatus("analisis", 'slang', $nxt, $stp, $totalstep, true);
} else {
    $return = array("status"=>false,"notifikasi"=>"Proses gagal
data : $step");
insertStatus("analisis", 'slang', 'slang', $stp, $totalstep, false);}

```



```

echo json_encode($return);
break;
case 'stopword' :
    $data_start = loadFile("analisis-sinonim");
    $data       = json_decode($data_start,true);
    $stopword   = loadStopword();
    foreach((array)$data as $dm) {
    $stop = false;
    foreach((array)$stopword as $stw) {
    if($stw==$dm) {
        $stop = true;
    }
    }
    if($stop==false) {
    $saved[]    = $dm;
    }
    }
    $simpan     = saveToFile("analisis-stopword", json_encode($saved));
    $pcs        = $step+1;
    $nxt        = "stemming";
    $stp        = $step+1;
    if($simpan) {
    $return     = array("status"=>"true","notifikasi"=>"Melakukan stopword
    $pcs/$totalstep","halaman"=>"analisis","next"=>$nxt,        "step"=>$stp,
    "totalstep"=>$totalstep);
    insertStatus("analisis",'stopword', $nxt, $stp, $totalstep, true);
    } else {
    $return     = array("status"=>false,"notifikasi"=>"Proses gagal
    data : $step");
    insertStatus("analisis",'stopword', 'stopword', $stp, $totalstep,
    false);
    }
    echo json_encode($return);
    break;
case 'stemming' :
    $data_stopword = loadFile("analisis-stopword");
    $data          = json_decode($data_stopword,true);
    $saved         = array();
    foreach((array)$data as $dm) {
    if($dm!=' ' && $dm!='') {
    $dm           = Enhanced_CS($dm);
    if(preg_match('/^[a-z]+$/',$dm)) {
    $saved[]     = $dm;
    }}}
    $simpan        = saveToFile("analisis-stemming", json_encode($saved));
    $nxt           = "entropi";
    $stp           = $step+1;
    if($simpan) {
    $return        = array("status"=>"true","notifikasi"=>"Melakukan
    Stemming","halaman"=>"analisis","next"=>$nxt,        "step"=>$stp,
    "totalstep"=>$totalstep);
    insertStatus("analisis",'stemming', $nxt, $stp,
    $totalstep, true);
    } else {
    $return        =
    array("status"=>false,"notifikasi"=>"Proses gagal data : $step");
    insertStatus("analisis",'stemming', 'stemming', $stp,
    $totalstep, false);}

```

Modul Program 4.29. Lanjutan *Source Code* Halaman Analysis

```

echo json_encode($return);
break;
case 'entropi' :
    $data_stemming      = loadFile("analisis-stemming");
    $data_stemming      = json_decode($data_stemming,true);
    $data_cf            = loadFile("analisis-casefolding");
    $data_cf            = json_decode($data_cf,true);
    $string             = "SELECT * FROM `tbl_tfidf`";
    $q                  = $koneksi->query($string);
    $f                  = $q->fetchAll(PDO::FETCH_NUM);
    // mencari entropinya
    $himpos            = array();
    $himnet            = array();
    $himneg            = array();
    $total             = count($f);
    foreach((array)$data_stemming as $dm) {
    //echo "<br /><br />".var_dump($dm);
    $string            = "SELECT * FROM `tbl_tfidf` WHERE `term`='$dm'";
    $q                  = $koneksi->query($string);
    $f                  = $q->fetchAll(PDO::FETCH_NUM);
    if(empty($f)) {
        $himnet[$dm][] = 0;
    }
    //echo "-----<br />";
    foreach((array)$f as $df) {
    //echo var_dump($df[4])."<br />";
    if($df[4]==1) {
        $himpos[$dm][] = $df[0];
        //$total
        ==+ 1;
    } else if($df[4]==-1) {
        $himneg[$dm][] = $df[0];
        //$total
        ==+ 1;
    } else {
        $himnet[$dm][] = $df[0];
        //$total
        ==+ 1;
    }
    }
    //echo "<br />=====<br /><br />";
    }
    $entpos            = array();
    $entnet            = array();
    $entneg            = array();
    foreach((array)$himpos as $hp) {
        $entpos[]      =
        (count($hp)/$total)*log(count($hp)/$total,2);
    }
    foreach((array)$himnet as $ht) {
        $entnet[]      =
        (count($ht)/$total)*log(count($ht)/$total,2);
    }
    foreach((array)$himneg as $hg) {
        $entneg[]      =
        (count($hg)/$total)*log(count($hg)/$total,2);
    }
    // mencari nilai Z dan P
    $Zpos              = array();
    $Zneg              = array();
    $Znet              = array();
    $epPos             = array();
    $epNet             = array();
    $epNeg             = array();

```

```

foreach((array)$sentpos as $sep) {
    $Zpos[] = exp($sep);
}
foreach((array)$sentnet as $sep) {
    $Znet[] = exp($sep);
}
foreach((array)$sentneg as $sep) {
    $Zneg[] = exp($sep);
}
// mencari nilai P
foreach((array)$sentpos as $sep) {
    $Ppos[] = (1/array_sum($Zpos))*exp($sep);
}
foreach((array)$sentnet as $sep) {
    $Pnet[] = (1/array_sum($Znet))*exp($sep);
}
foreach((array)$sentneg as $sep) {
    $Pneg[] = (1/array_sum($Zneg))*exp($sep);
}
// mencari propabilitas
if(array_sum($Zpos)>0) {
    $probpos = (1/array_sum($Zpos))*array_sum($sentpos);
} else {
    $probpos = 0;
}
if(array_sum($Znet)>0) {
    $probnet = (1/array_sum($Znet))*array_sum($sentnet);
} else {
    $probnet = 0;
}
if(array_sum($Zneg)>0) {
    $probneg = (1/array_sum($Zneg))*array_sum($sentneg);
} else {
    $probneg = 0;
}
    $saved['pos']= $probpos;
    $saved['net']= $probnet;
    $saved['neg']= $probneg;
if($probpos>$probnet && $probpos>$probneg) {
    $hasil= "positif";
} else if($probneg>$probnet && $probneg>$probpos) {
    $hasil= "negatif";
} else {
    $hasil= "netral";}
$asli = $data_cf['asli'];
$stringi = "INSERT INTO `tbl_analysis` (
`dataset`,`probpos`,`probnet`,`probneg`,`hasil`,`text` ) VALUES
('0','$probpos','$probnet','$probneg','$hasil','$asli')";
//echo var_dump($stringi);
$qj = $koneksi->query($stringi);
$simpan = saveToFile("analisis-entropi", json_encode($saved));
    $nxt = "finish";
    $stp = 0;
    $notifikasi = "Berhasil melakukan perhitungan";
$stopword = implode(" ", json_decode(loadFile("analisis-
stopword"),true));
$tokenization = "[".implode(" ", json_decode(loadFile("analisis-
tokenisasi"),true))."]";

```

```

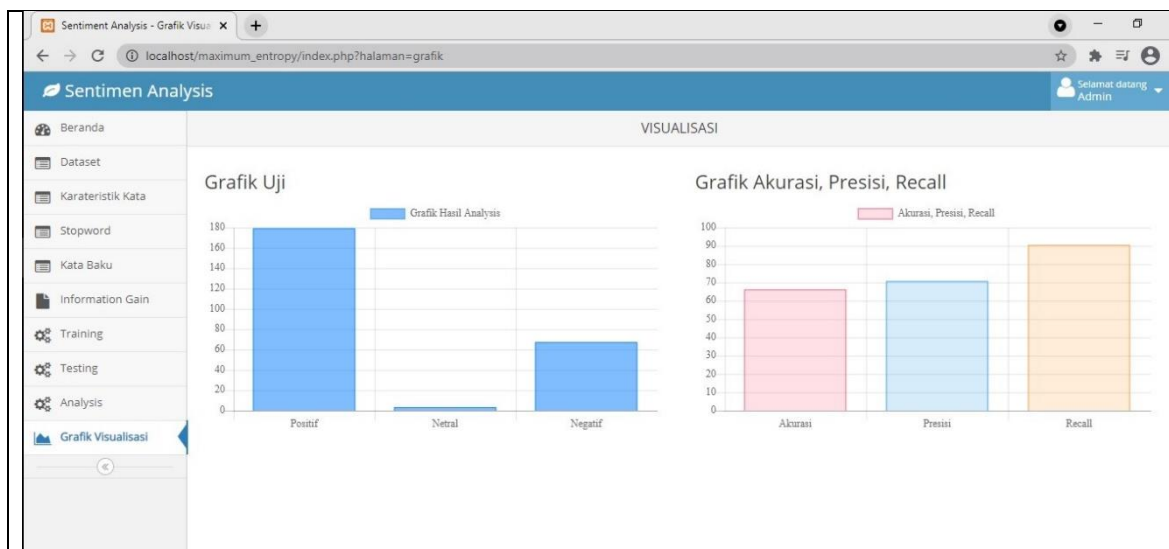
$sinonim = "[".implode(", ", json_decode(loadFile("analisis-sinonim"), true))."];
$stemming = "[".implode(", ", $data_stemming)."];
$casefoldeng= $data_cf['hasil'];
if($q) {
$return =
array("status"=>"true", "notifikasi"=>$notifikasi, "halaman"=>"analisis",
"next"=>$nxt, "step"=>$stp, "totalstep"=>$totalstep,
"casefolding"=>$casefoldeng, "stopword"=>$stopword,
"stemming"=>$stemming, "token"=>$tokenization, "sinonim"=>$sinonim,
"hasil" => $hasil, "pos"=> $probpos, "net"=> $probnet, "neg"=>$probneg);
insertStatus("analisis", 'entropi', $nxt, $stp, $totalstep, true);
} else {
$return = array("status"=>false, "notifikasi"=>"Proses gagal data :
$step");
insertStatus("analisis", 'entropi', 'entropi', $stp, $totalstep, false);
}
echo json_encode($return);
}
?>

```

Modul Program 4.32. Lanjutan *Source Code* Halaman *Analysis*

4.1.11. Halaman Grafik Visualisasi

Halaman grafik visualisasi merupakan halaman yang berisikan informasi hasil dari 1 klasifikasi sentimen yang ada pada sistem dalam bentuk grafik. Tampilan halaman grafik dapat dilihat pada gambar 4.11 berikut :



Gambar 4.11. Halaman Grafik Visualisasi

Berikut ini merupakan *source code* halaman grafik visualisasi :

```

<?php
$string = "SELECT a.*,b.nilai FROM tbl_analysis As a
LEFT JOIN tbl_dataset As b ON a.dataset = b.id";
$q = $koneksi->query($string);
if ($q && $q->rowCount () > 0) {

```

Modul Program 4.33. *Source Code* Halaman Grafik Visualisasi

```

$pos = array();
$net = array();
$neg = array();
$totaltp2 = 0;
$totaltn2 = 0;
$totalfp2 = 0;
$totalfn2 = 0;
foreach((array)$f as $dt) {
if($dt[6]=='positif') {
    $pos[] = 1;
} else if($dt[6]=='negatif') {
    $neg[] = 1;
} else if($dt[6]=='netral'){
    $net[] = 1;
}
}
$aktual = $dt[7];
$hs = $dt[6];
if($dt[6] == 'positif') {
    $mx = 1;
}
if($dt[6] == 'netral') {
    $mx = 0;
}
if($dt[6] == 'negatif') {
    $mx = -1;
}
}
if ($aktual == 1 && $mx == 1 ) {
    $totaltp2 = $totaltp2 + 1;
}
if ($aktual == -1 && $mx == -1 ) {
    $totaltn2 = $totaltn2 + 1;
}
if (($aktual == 1 || $aktual == 0) && $mx == -1 ) {
    $totalfp2 = $totalfp2 + 1;
}
if ($aktual == 1 && $mx == 0 ) {
    $totalfp2 = $totalfp2 + 1;
}
if (($aktual == -1 || $aktual == 0) && $mx == 1 ) {
    $totalfn2 = $totalfn2 + 1;
}
if ($aktual == -1 && $mx == 0 ) {
    $totalfn2 = $totalfn2 + 1;
}
}
}
$totak2 = 0;
$totpre2 = 0;
$totreca2 = 0;

$totak2 = ($totaltp2 + $totaltn2)/($totaltp2 + $totalfp2 + $totalfn2
+ $totaltn2);
$totpre2 = ($totaltp2!=0)?($totaltp2)/($totaltp2 + $totalfp2):0;
$totreca2 = ($totaltp2!=0)?($totaltp2)/($totaltp2 + $totalfn2):0;
    $akurasi_ent = $totak2 * 100;
    $presisi_ent= $totpre2 * 100;
    $recall_ent = $totreca2 * 100;
} else {

```

Modul Program 4.34. Lanjutan *Source Code* Halaman Grafik Visualisasi

```

$lpos = array();
  $lnet = array();
  $lneg = array();

  $akurasi_ent      = 0;
  $presisi_ent      = 0;
  $recall_ent      = 0;
}
$graph              = array(count($pos), count($net), count($neg));
$graph13            = array($akurasi_ent, $presisi_ent, $recall_ent);
?>

```

Modul Program 4.35. Lanjutan *Source Code* Halaman Grafik Visualisasi

4.2 Pengujian

Tahap pengujian bertujuan untuk mengetahui tingkat keberhasilan pada sistem yang telah dibangun. Implementasi pada pengujian sistem penelitian ini menggunakan pengujian *confusion matrix*. Melalui *confusion matrix* hasil pengujian akurasi, presisi dan recall akan diketahui.

Pengujian akan dilakukan pada 1.252 data latih (*training*). Kemudian 20% dari total data *training* akan dijadikan sebagai data uji. Untuk mendapatkan objektivitas dari komentar pada data *training* maka dilakukan pelabelan untuk pengklasifikasian kelas sentimen pada data *training*.

4.2.1 Pengujian Confusion Matrix

Pengujian pada data uji dilakukan secara manual dengan mengambil data sampel sebanyak 250 data sentimen dari komentar-komentar yang ada di *twitter* yang sudah di input kedalam sistem.

Dari hasil analisis data uji tersebut maka dibentuk sebuah tabel *confusion matrix* yang dapat dilihat pada tabel 4.1 berikut

Tabel 4.1. Tabel *Confusion Matrix*

Prediksi	AKTUAL		
	Positif	Negatif	Netral
Positif	168	60	3
Negatif	10	3	1
Netral	3	1	1

Tabel *confusion matrix* digunakan sebagai acuan dalam melakukan pengujian akurasi, presisi dan recall. Melalui *confusion matrix* dapat diketahui *True Positif*, *False Negative*, *False Positive* dan *True Negative* yang dihasilkan oleh sistem dalam melakukan analisis sentimen.

Perhitungan dari nilai akurasi dengan menggunakan rumus 2.10 yang ada pada bab II adalah sebagai berikut :

$$\text{Akurasi} = \frac{168+3+1}{168+60+3+10+3+1+3+1+1} = 0,68$$

Perhitungan dari nilai *precision* dengan menggunakan rumus 2.7 pada bab II adalah sebagai berikut :

$$\text{Precision} = \frac{168}{168+60+3} = 0,72$$

Perhitungan dari nilai *recall* dengan menggunakan rumus 2.8 pada bab II adalah sebagai berikut :

$$\text{Recall} = \frac{168}{168+10+3} = 0,92$$

Hasil dari perhitungan pengujian dengan menggunakan *confusion matrix* dapat dilihat tabel 4.2 yang merupakan tabel hasil dari pengujian akurasi, presisi dan *recall* pada metode *maximum entropy* dengan seleksi fitur *information gain*.

Tabel 4.2. Hasil Pengujian

Sentimen Uji	Akurasi	Presisi	Recall
250	68%	72%	92%

4.3. Pembahasan

Nilai prediksi yang benar akan dibandingkan dengan seluruh nilai pada data yang ada, dimana hasil dari perbandingan nilai tersebut akan menghasilkan nilai akurasi. Berdasarkan hasil pengujian manual diatas memiliki akurasi yang cukup baik yaitu 68%

Nilai *precision* (presisi) merupakan jumlah dokumen pada teks yang relevan untuk dikenali diantara semua dokumen teks yang terpilih oleh sistem. Nilai presisi merupakan

nilai benar positif yang dibagi seluruh nilai yang diperkirakan bernilai positif. Berdasarkan pengujian nilai presisi manual diatas, sistem mendapatkan nilai presisi yang cukup baik yaitu 72%

Recall menunjukkan perbandingan diantara nilai yang benar positif dengan seluruh data yang sebenarnya. Berdasarkan pengujian pada nilai *recall* diatas, sistem mendapatkan nilai 92%.

Pengujian pada sistem yang telah dilakukan menghasilkan bahwa metode *Maximum Entropy* dengan seleksi fitur *Information Gain* memberikan nilai akurasi sebesar 68%, presisi sebesar 72% dan *recall* sebesar 92%. Tingkat keakurasian pada klasifikasi yang ada ada pada sistem sangat dipengaruhi oleh proses pembobotan pada kata dan proses *training* atau data latih.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut :

1. Analisis sentimen dengan menggunakan metode *Maximum Entropy* dengan fitur *Information Gain* dapat menghasilkan klasifikasi komentar sebanyak 3 kelas yaitu positif, negatif dan netral dengan nilai yang cukup baik pada data komentar.
2. Hasil pengujian pada sistem metode *Maximum Entropy* dengan seleksi fitur *Information Gain* memberikan akurasi sebesar 68%, presisi 72% dan *recall* 92%

5.2 Saran

Pada penelitian ini, terdapat beberapa keterbatasan dan kekurangan. Keterbatasan dan kekurangan pada penelitian ini bisa dijadikan acuan dan pertimbangan untuk melakukan penelitian selanjutnya. adapun beberapa keterbatasan dan kekurangan pada penelitian ini sebagai berikut :

1. Data komentar diperluas tidak hanya menggunakan data twitter saja.
2. Pengambilan data seharusnya bisa dilakukan secara *real time* dengan menggunakan *API key*.
3. Data komentar dapat dikembangkan dengan menggunakan komentar berbahasa inggris.

DAFTAR PUSTAKA

- Aditya, B. R. (2015). *Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining*. 7(2), 8.
- Alroy, A. B., Adikara, P. P., & Adinugroho, S. (2019). Klasifikasi Hoaks Menggunakan Metode Maximum Entropy Dengan Seleksi Fitur Information Gain. *9 September 2019*, 3, 8.
- Anggraeni, D. (2008). *KLASIFIKASI TOPIK MENGGUNAKAN METODE NAÏVE BAYES DAN MAXIMUM ENTROPY PADA ARTIKEL MEDIA MASSA DAN ABSTRAK TULISAN*. 12.
- Attabi, A. W., Muflikhah, L., & Fauzi, M. A. (2018). *Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain*. 2, 7.
- Bridge. (2011). *Unstructured Data and the 80 Percent Rule*. <https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>.
- Cindo, M., Dian Palupi Rini, & Ermatita. (2019). Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan Support Vector Machine dan Maximum Entropy. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(3), 402–407. <https://doi.org/10.29207/resti.v3i3.1159>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Gunawan, F., Fauzi, M. A., & Adikara, P. P. (2017). Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). *Systemic: Information System and Informatics Journal*, 3(2), 1–6. <https://doi.org/10.29080/systemic.v3i2.234>

- Han, J., & Kamber, M. (2000). *Data Mining: Concepts and Techniques*. In *San Francisco: Morgan kaufmann*;
- Hearst. (2003). *What Is Text Mining?* <http://www.sims.berkeley.edu/~hearst/text.mining.html>
- Hidayat, R. (2019). *SENTIMENT ANALYS PADA MEDIA SOSIAL DENGAN METODE NAÏVE BAYES CLASSIFER (Studi Kasus: Twitter Komisi Pemilihan Umum Republik Indonesia)*. UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” YOGYAKARTA.
- Howard, P. N., & Parks, M. R. (2012). Social Media and Political Change: Capacity, Constraint, and Consequence. *Journal of Communication*, 62(2), 359–362. <https://doi.org/10.1111/j.1460-2466.2012.01626.x>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*.
- Maulida, I., Suyatno, A., & Hatta, H. R. (2016). *Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain*. 17(2), 10.
- Mejova, Y. (2009). *Sentiment Analysis: An Overview*. 34.
- Mujilahwati, S. (2016). *PRE-PROCESSING TEXT MINING PADA DATA TWITTER*. 8.
- Negara, A. B. P., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), 599. <https://doi.org/10.25126/jtiik.2020711947>
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). *Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet*. 1, 9.

- Pamungkas, D. S., Setiyanto, N. A., & Dolphina, E. (2015). *ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAIVE BAYES CLASSIFIER TERHADAP KATA KUNCI "KURIKULUM 2013."* 14(4), 16.
- Powers, D. & Ailab. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol*, 2, 2229–3981. <https://doi.org/10.9735/2229-3981>
- Rasywir, E., & Purwarianti, A. (2015). *Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin*. 3(2), 8.
- Rini, M. P., Adikara, P. P., & Adinugroho, S. (2019). *Klasifikasi Teks Pengaduan Suara Warga Kabupaten Pasuruan menggunakan Metode Maximum Entropy*. 3, 5.
- Rofqoh, U., Perdana, R. S., & Fauzi, M. A. (n.d.). *Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features*. 8.
- Sabily, A. F., Adikara, P. P., & Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. *5 Mei 2019*, 3, 6.
- Susilo, T. H. (2013). *PENGLASIFIKASIAN TOPIK DAN ANALISIS SENTIMEN DALAM MEDIA SOSIAL*. 10.
- Syah, A. P., Adiwijaya, & Faraby, S. A. (2017). *ANALISIS SENTIMEN PADA DATA ULASAN PRODUK TOKO ONLINE DENGAN METODE MAXIMUM ENTROPY*. 4, 9.