

**KLASIFIKASI SENTIMEN TERHADAP DATA *TEXT*
JEJARING SOSIAL DENGAN TOPIK
PEMBELAJARAN DARING MENGGUNAKAN
*LOGISTIC REGRESSION***

TUGAS AKHIR



Oleh

**Hisyam Sanusi
09011181621028**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021**

**KLASIFIKASI SENTIMEN TERHADAP DATA *TEXT*
JEJARING SOSIAL DENGAN TOPIK
PEMBELAJARAN DARING MENGGUNAKAN
*LOGISTIC REGRESSION***

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer



Oleh:

HISYAM SANUSI

09011181621028

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN

**KLASIFIKASI SENTIMEN TERHADAP DATA TEKS
JEJARING SOSIAL DENGAN TOPIK PEMBELAJARAN
DARING MENGGUNAKAN LOGISTIC REGRESSION**

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**

Oleh

**Hisyam Sanusi
09011181621028**

Indralaya, 27 Juli 2021

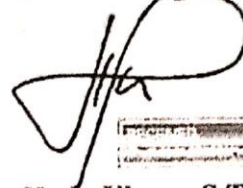
Mengetahui,

Pembimbing I



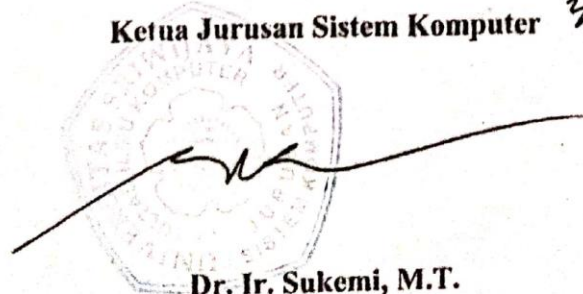
**Deris Stiawan, M.T., Ph.D
NIP. 197806172006041002**

Pembimbing II



**Huda Ubaya, S.T., M.T
NIP. 198106162012121003**

Ketua Jurusan Sistem Komputer 27/7/21



**Dr. Ir. Sukemi, M.T.
NIP. 196612032006041001**

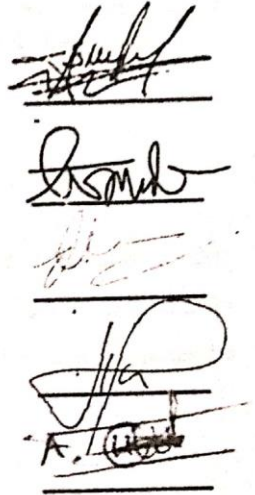
HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

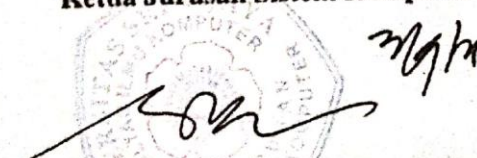
Hari : Jum'at
Tanggal : 30 Juli 2021

Tim Penguji :

1. **Ketua** : Sarmayanta Sembiring, S.SI., M.T
2. **Sekretaris** : Tri Wanda Septian, M.Sc.
3. **Pembimbing 1** : Deris Stiawan, M.T., Ph.D
4. **Pembimbing 2** : Huda Ubaya, S.T., M.T.
5. **Penguji** : Ahmad Heryanto, S.Kom., M.T.



Mengetahui,
Ketua Jurusan Sistem Komputer


Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Hisyam Sanusi

NIM : 09011181621028

Judul : Klasifikasi Sentimen Terhadap Data *Text* Jejaring Sosial dengan Topik Pembelajaran Daring Menggunakan *Logistic Regression*

Hasil Penyecekan *Software iThenticate/Turnitin* : 8 %

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Indralaya, Agustus 2021



Hisyam Sanusi
NIM. 09011181621028

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh,

Puji dan syukur penulis panjatkan kepada Allah Subhanahu Wa Ta'ala, atas segala karunia, berkat, dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Proposal Tugas Akhir ini dengan judul “Klasifikasi Sentimen Terhadap Data *Text* Jejaring Sosial Dengan Topik Pembelajaran Daring Menggunakan *Logistic Regression*”.

Dalam laporan ini penulis menjelaskan mengenai bagaimana cara mengklasifikasi *Sentimen* dengan menggunakan *Logistic Regression* disertai data-data yang diperoleh penulis saat melakukan pengujian. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan dan referensi untuk para akademisi dan peneliti lain yang juga berada atau sedang menekuni bidang *data text*, khususnya pemerintah.

Pada penyusunan laporan ini, penulis banyak mendapatkan ide dan saran serta bantuan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah Subhanahu Wa Ta'ala dan terimakasih kepada yang terhormat:

1. Orang tua beserta keluarga yang selalu mendoakan, memberikan dukungan baik moril maupun materil, serta motivasi dan semangat selama hidup penulis.
2. Bapak Jaidan Jauhari, S.Pd. M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. H.Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Deris Stiawan, M.T., Ph.D. selaku Pembimbing Akademik dan Pembimbing 1 Tugas Akhir.
5. Bapak Huda Ubaya, S.T, M.T. selaku Pembimbing 2 Tugas Akhir.
6. Nauval Faris Fadhlullah, Ilham Solehan, Jerry Pratama Saputra dan Rizki Ronaldo sebagai partner dan teman dari awal pengerjaan tugas akhir, baik dalam pencarian literatur, pemrograman, serta diskusi mengenai riset dan lainnya.

7. Teman - teman dari Jurusan Sistem Komputer yang tidak bisa disebutkan satu-persatu. Khususnya seluruh teman-teman dari kelas SK16A Indralaya.
8. Para sahabat saya yang ada di kosan, Ilham Solehan, Nauval Faris Fadhlullah, Ichwanul Hakim, Jerry Pratama, Rap Nur Muhammad, Madyus Randika & Muhammad Sultan Alif yang selalu memotifasi saya untuk cepat lulus.
9. Adithia Jovandy yang sudah banyak membantu banyak hal.
10. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa Skripsi ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi dikemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pemikiran dalam peningkatan mutu pembelajaran dan penelitian.

Indralaya, Agustus 2021

Penulis

KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL DENGAN TOPIK PEMBELAJARAN DARING MENGUNAKAN LOGISTIC REGRESSION

Hisyam Sanusi (09011181621028)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : syamesanusi@gmail.com

ABSTRAK

Wabah Corona Virus Disease (Covid-19) yang melanda lebih dari 200 Negara di Dunia, telah memberikan tantangan tersendiri bagi lembaga pendidikan, khususnya pendidikan tinggi. Mengantisipasi penularan virus tersebut pemerintah telah mengeluarkan berbagai kebijakan, seperti isolasi, *social and physical distancing* hingga pembatasan sosial berskala besar (PSBB). Kondisi ini mengharuskan warganya untuk tetap stay at home, belajar di rumah, maka klasifikasi sentimen masyarakat terhadap pembelajaran daring ini dirasa perlu. Sentimen masyarakat direkam dan dikumpulkan dari suatu jejaring sosial, yaitu jejaring sosial twitter dimana terdapat banyak sentimen masyarakat tentang topik pembelajaran daring, Klasifikasi sentimen tersebut akan menggunakan *Natural Language Processing (NLP) Logistic Regression* sebagai metode klasifikasinya. Dari percobaan yang dilakukan dengan *split* data 80% data pelatihan dengan 20% data pengujian menghasilkan nilai *Precision, Recall* dan *F1-Score* yang paling tinggi, sedangkan untuk percobaan dengan *split* data 20% data pelatihan dengan 80% data pengujian menghasilkan nilai *Precision, Recall* dan *F1-Score* yang paling rendah. Klasifikasi data dapat terlaksana dengan baik menggunakan metode *Logistic Regression*, sehingga dapat membuat sistem yang mampu untuk mengkasifikasi sentimen dengan topik pembelajaran daring. Hasil prediksi data *text* sentimen menggunakan metode *Logistic Regression* dapat dijadikan sebagai bahan pertimbangan atau acuan terhadap pengambilan keputusan pemerintah dalam mengambil kebijakan mengenai topik pembelajaran daring.

Kata Kunci : *Natural Language Processing, Logistic Regression, Pembelajaran Daring, sentimen, data text.*

CLASSIFICATION OF SENTIMENTS ON SOCIAL NETWORK TEXT DATA WITH ONLINE LEARNING TOPIC USING LOGISTIC REGRESSION

Hisyam Sanusi (09011181621028)

*Dept. of Computer Engineering, Faculty of Computer Science, Sriwijaya
University*

Email : syamesanusi@gmail.com

ABSTRACT

The outbreak of Corona Virus Disease (Covid-19) which has hit more than 200 countries in the world, has presented its own challenges for educational institutions, especially higher education. Anticipating the transmission of the virus, the government has issued various policies, such as isolation, social and physical distancing to large-scale social restrictions (PSBB). This condition requires citizens to stay at home, study at home, so the classification of public sentiment towards online learning is deemed necessary. Community sentiments are recorded and collected from a social network, namely the Twitter social network where there are many public sentiments about online learning topics. The classification of these sentiments will use Natural Language Processing (NLP) Logistic Regression as the classification method. From the experiments conducted with 80% data split of training data with 20% of test data resulted in the highest Precision, Recall and F1-Score values, while for experiments with 20% data split of training data with 80% of test data yielded Precision, Recall and The lowest F1-Score. Data classification can be carried out properly using the Logistic Regression method, so that it can create a system that is able to classify sentiments with online learning topics. The results of the prediction of sentiment text data using the Logistic Regression method can be used as consideration or reference for government decision making in making policies regarding online learning topics.

Keywords: *Natural Language Processing, Logistic Regression, Online Learning, sentiment, text data.*

DAFTAR ISI

	Halaman
Halaman Judul	i
Halaman Pengesahan	ii
Halaman Persetujuan	iii
Halaman Pernyataan	iv
Kata Pengantar	v
Abstrak	vii
<i>Abstract</i>	ix
Daftar Isi	xi
Daftar Gambar	xvi
Daftar Tabel	xviii
Daftar Lampiran	xix
Bab I. Pendahuluan	1
1.1 Latar Belakang	1
1.2. Tujuan	2
1.3. Perumusan Masalah	2
1.4 Batasan Masalah	2
1.5 Metologi Penelitian.....	3
1.6 Sistematika Penulisan.....	5
Bab II. Tinjauan Pustaka	7
2.1. Analisis Sentimen	7
2.2. <i>Text Mining</i>	7
2.3. <i>Natural Language Processing (NLP)</i>	9
2.4. <i>Text Preprocessing</i>	11

2.4.1 Pembersihan Dokumen(<i>Cleansing</i>).....	11
2.4.2. <i>Case Folding</i>	11
2.4.3. <i>Tokenizing</i>	12
2.4.4. <i>Stemming</i>	12
2.4.5. Normalisasi.....	12
2.5. <i>Feature Extraction</i>	12
2.5.1. Model Bag of Words.....	13
2.5.2. Pembobotan Term.....	13
2.6. <i>Machine Learning</i>	14
2.7. <i>Logistic Regression</i>	15
2.8. <i>K-Fold Cross Validation</i>	16
2.9. <i>Confussion Matrix</i>	17
2.10. <i>Precision, Recall, F-measure</i>	18
Bab III. Metodologi Penelitian	20
3.1. Metologi Penelitian.....	20
3.2. Kerangka Kerja	21
3.3. Pengambilan Data	22
3.4. <i>Preprocessing</i>	25
3.4.1. <i>Cleaning</i>	25
3.4.2. <i>Case Folding</i>	27
3.4.3. <i>Tokenizing</i>	28
3.4.4. <i>Stemming</i>	29
3.4.5. Pemisahan Feature dan Label.....	31

3.5. Ekstraksi fitur	31
3.6. Cross Validation	31
3.7. Confussion Matrix	32
Bab IV. Pengujian dan Analisa.....	33
4.1 Pendahuluan.....	33
4.2 <i>Dataset Text</i>	33
4.3 Proses <i>Preprocessing</i>	33
4.4 Ekstraksi Fitur Data Text	36
4.5 Percobaan dan Pengujian dan percobaan menggunakan <i>Logistic Regression</i>	37
4.5.1. Percobaan dengan data pelatihan 80% dan data pengujian 20%.....	39
4.5.1.1. Analisis Percobaan dengan data pelatihan 80% dan data pengujian 20%.....	40
4.5.2. Percobaan dengan data pelatihan 70% dan data pengujian 30%.....	43
4.5.2.1. Analisis Percobaan dengan data pelatihan 70% dan data pengujian 30%.....	44
4.5.3. Percobaan dengan data pelatihan 60% dan data pengujian 40%.....	47
4.5.3.1. Analisis Percobaan dengan data pelatihan 60% dan data pengujian 40%.....	48
4.5.4. Percobaan dengan data pelatihan 50% dan data	

pengujian 50%.....	51
4.5.4.1. Analisis Percobaan dengan data pelatihan 50% dan data pengujian 50%.....	52
4.5.5. Percobaan dengan data pelatihan 40% dan data pengujian 60%.....	55
4.5.5.1. Analisis Percobaan dengan data pelatihan 40% dan data pengujian 60%.....	56
4.5.6. Percobaan dengan data pelatihan 30% dan data pengujian 70%.....	59
4.5.6.1. Analisis Percobaan dengan data pelatihan 30% dan data pengujian 70%.....	60
4.5.7. Percobaan dengan data pelatihan 20% dan data pengujian 80%.....	63
4.5.7.1. Analisis Percobaan dengan data pelatihan 20% dan data pengujian 80%.....	64
4.6 Perbandingan hasil pengujian dengan nilai <i>Split data</i>	67
4.7 Analisa.....	69
Bab V. Kesimpulan	70
5.1 Kesimpulan	70
5.2 Saran	70

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Struktur Kalimat Bahasa Indonesia	9
Gambar 2.2 <i>Extended confusion matrix</i> 3x3.....	17
Gambar 3.1 Diagram alur penelitian	20
Gambar 3.2 Dataset yang diambil dari Drone Emprit Academy.....	21
Gambar 3.3 Wordcloud sentimen negatif	22
Gambar 3.4 Wordcloud sentimen netral	23
Gambar 3.5 Wordcloud sentimen positif	23
Gambar 3.6 Diagram alur penelitian	24
Gambar 3.7 Kerangka kerja cleanning.....	25
Gambar 3.8 Proses Cleanning.....	25
Gambar 3.9 Sentimen yang diproses cleaning.....	25
Gambar 3.10 Kerangka kerja case folding.....	26
Gambar 3.11 Proses case folding.....	26
Gambar 3.12 Kerangka kerja tokenizing.....	27
Gambar 3.13 Proses tokenizing.....	27
Gambar 3.14 Sentimen yang diproses tokenisasi.....	28
Gambar 3.15 Kerangka kerja stemming.....	29
Gambar 3.16 Proses stemming.....	29
Gambar 3.17 Sentimen yang diproses stemming.....	29
Gambar 3.18 kata-kata yang menjadi scalar.....	30
Gambar 4.1 Alur Kerja dari Prerocessing.....	33
Gambar 4.2 Perbandingan sebelum dan setelah dilakukan Cleanning.....	34
Gambar 4.3 Perbandingan sebelum dan setelah dilakukan Stemming.....	34
Gambar 4.4 Perbandingan sebelum dan setelah dilakukan <i>Tokenization</i>	35
Gambar 4.5 Contoh data sentimen dari ekstraksi fitur.....	35
Gambar 4.6 Split data Pelatihan sebanyak 80% dataset.....	38
Gambar 4.7 Split data Pengujian sebanyak 20% dataset.....	39

Gambar 4.8 Hasil percobaan dengan data pelatihan 80% dan data pengujian 20%.....	40
Gambar 4.9 Split data Pelatihan sebanyak 70% dataset.....	42
Gambar 4.10 Split data Pengujian sebanyak 30% dataset.....	43
Gambar 4.11 Hasil percobaan dengan data pelatihan 70% dan data pengujian 30% ...	44
Gambar 4.12 Split data Pelatihan sebanyak 60% dataset.....	46
Gambar 4.13 Split data Pengujian sebanyak 40% dataset.....	47
Gambar 4.14 Hasil percobaan dengan data pelatihan 60% dan data pengujian 40% ...	48
Gambar 4.15 Split data Pelatihan sebanyak 50% dataset.....	50
Gambar 4.16 Split data Pengujian sebanyak 50% dataset.....	51
Gambar 4.17 Hasil percobaan dengan data pelatihan 50% dan data pengujian 50% ...	52
Gambar 4.18 Split data Pelatihan sebanyak 40% dataset.....	54
Gambar 4.19 Split data Pengujian sebanyak 60% dataset.....	55
Gambar 4.20 Hasil percobaan dengan data pelatihan 40% dan data pengujian 60% ...	56
Gambar 4.21 Split data Pelatihan sebanyak 30% dataset.....	58
Gambar 4.22 Split data Pengujian sebanyak 70% dataset.....	59
Gambar 4.23 Hasil percobaan dengan data pelatihan 30% dan data pengujian 70% ...	60
Gambar 4.24 Split data Pelatihan sebanyak 20% dataset.....	62
Gambar 4.25 Split data Pengujian sebanyak 80% dataset.....	63
Gambar 4.26 Hasil percobaan dengan data pelatihan 20% dan data pengujian 80% ...	64
Gambar 4.27 Grafik transisi nilai dari pengujian dengan parameter Precision.....	66
Gambar 4.28 Grafik transisi nilai dari pengujian dengan parameter Recall.....	67
Gambar 4.29 Grafik transisi nilai dari pengujian dengan parameter F1-Score.....	67

DAFTAR TABEL

	Halaman
Tabel 2.1 Tabel Confusion matrix berukuran $L = 2$	16
Tabel 3.1 klasifikasi multi-class yang memiliki 3 keluaran kelas confusion matrix.....	31
Tabel 4.1 Tabel Jumlah pembagian dataset pada setiap percobaan.....	36
Tabel 4.2 Tabel klasifikasi multi-class confusion matrix 3×3	37
Tabel 4.3. Hasil perbandingan pengujian pada setiap split data yang di uji.....	66

DAFTAR LAMPIRAN

LAMPIRAN 1. Program Klasifikasi Sentimen *Logistic Regression*..... 1-

A

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Wabah Coronavirus Disease (Covid-19) yang melanda lebih dari 200 Negara di Dunia, telah memberikan tantangan tersendiri bagi lembaga pendidikan, khususnya pendidikan tinggi. Mengantisipasi penularan virus tersebut pemerintah telah mengeluarkan berbagai kebijakan, seperti isolasi, *social and physical distancing* hingga pembatasan sosial berskala besar (PSBB). Kondisi ini mengharuskan warganya untuk tetap stay at home, belajar di rumah.

Masyarakat Indonesia kini mulai menjalani aktivitas sehari-harinya seperti biasa. Hal ini kembali menimbulkan pro dan kontra karena masyarakat merasa jika pembelajaran daring sudah tidak lagi diperlukan, apalagi hal ini didukung dengan protokol COVID-19 yang sudah diberlakukan. Penyebaran vaksin juga sudah mulai dilakukan terutama bagi para petugas pendidikan. Walaupun begitu, masih banyak juga masyarakat yang berpendapat jika pembelajaran daring masih menjadi solusi yang terbaik dalam penanganan COVID-19. [1]

Untuk mengetahui tingkat keefektifitasan pembelajaran daring bagi masyarakat Indonesia, Disini akan dilakukan pengklasifikasian sentimen masyarakat dengan *Natural Language Processing* yang datanya didapat dari *Tweet* yang diunggah pada media sosial Twitter. Analisis sentimen didefinisikan sebagai ekstraksi informasi dari sepotong teks. Analisis sentimen berarti menceritakan keadaan pikiran pembicara, penulis esai, atau subjek lain tentang fakta atau ekstremitas umum yang relevan atau jawaban antusias terhadap sebuah laporan, komunikasi. Regresi logistik memiliki beberapa kelebihan dibandingkan teknik analisis lain yaitu: 1. Regresi logistik tidak memiliki asumsi normalitas dan heteroskedastisitas atas variabel bebas yang digunakan dalam model sehingga tidak diperlukan uji asumsi klasik walaupun variabel independen berjumlah lebih dari satu. 2. Variabel independen dalam regresi logistik bisa campuran dari variabel kontinu, distrik, dan dikotomis. 3. Regresi logistik tidak membutuhkan

keterbatasan dari variabel independennya. 4. Regresi logistik tidak mengharuskan variabel bebasnya dalam bentuk interval. Dr. Anand Kumar, et al melakukan penelitian yang bermaksud untuk membandingkan berbagai metode dalam melakukan klasifikasi sentimen dan salah satu sentimen yang dipakai adalah *Logistic Regression*. [2] Dari penelitian tersebut menunjukkan bahwa metode *Logistic Regression* menghasilkan akurasi yang cukup baik. Berdasar penelitian tersebut, penulis memutuskan untuk menggunakan *Logistic Regression* sebagai metode klasifikasi sentimen analisis dan menggunakan bahasa pemrograman Python. Dengan begitu klasifikasi sentimen yang berupa *data text* dapat dilaksanakan. [3]

1.2 Tujuan

Adapun tujuan dilakukannya penelitian ini adalah :

1. Tujuan pada penelitian ini adalah untuk dapat mengklasifikasikan sentimen yang tersebar di masyarakat mengenai topik pembelajaran daring menggunakan metode *Logistic Regression*.
2. Dapat dijadikan sebagai acuan untuk pihak terkait terhadap kebijakan yang diambil mengenai pembelajaran daring.

1.3 Perumusan Masalah

Dalam penelitian ini inti permasalahan adalah belum adanya klasifikasi sentimen dari masyarakat mengenai pandangan mereka terhadap pembelajaran daring yang menjadi program Pemerintah dalam mengatasi kesulitan pembelajaran bagi masyarakat karena penyebaran COVID-19.

1.4 Batasan Masalah

Berdasarkan rumusan masalah diatas, akan dilakukan pembatasan masalah agar penelitian ini tetap terarah.

1. Pada penelitian kali ini akan menggunakan dataset dari Drone Emprit Academy (DEA) yaitu dataset API Twitter dengan topik Pembelajaran Daring yang diambil mulai tanggal 16 Maret 2020 sampai 4 Juni 2021.
2. Pada penelitian kali ini akan menggunakan *Logistic Regression* untuk melakukan klasifikasi sentimen.
3. Pada penelitian kali ini bersifat simulasi menggunakan program dengan bahasa *python*.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini akan melewati beberapa tahapan sebagai berikut:

1. **Perumusan masalah**

Tahap ini adalah tahap yang menentukan permasalahan yang ada tentang bagaimana melakukan *preprocessing data text* Pembelajaran daring, ekstraksi fitur, dan cara menyelesaikan permasalahan tersebut.

2. **Studi Pustaka/literatur.**

Tahap ini adalah proses pencarian literatur yang menunjang dalam penelitian yang akan dilakukan, baik menggunakan kata kunci tertentu atau spesifik dalam bidang yang diteliti.

3. **Perancangan.**

Tahap ini adalah tahap perancangan sistem yang dibuat berdasarkan perumusan masalah yang dicari dalam penelitian. Pada tahap ini akan dibahas tentang proses pembuatan program untuk *Normalisasi data text*, segmentasi fitur, *Klasifikasi data text*.

4. **Pengujian.**

Tahap ini dilakukan dengan pengujian terhadap apa yang dibuat sesuai dengan rancangan.

5. Validasi

Tahap ini diperlukan untuk mengecek apakah sistem telah mampu melakukan klasifikasi dari dataset yang diberikan.

6. Evaluasi

Tahap ini bertujuan untuk mengetes sistem yang di buat menggunakan *data test* yang telah disediakan agar dapat mengetahui seberapa akurat sistem mengkasifikasi data sentimen.

7. Analisis

Tahap ini dilakukan dengan mengambil data dan menganalisa data yang sudah diolah dengan sistem yang dibuat. Analisa akan divisualisasikan untuk mengetahui seberapa baik program yang telah dibuat dapat menjalankan fungsi dan tujuannya.

8. Kesimpulan dan Saran

Pada tahap keenam akan dilakukan penarikan kesimpulan dari hasil analisa dan studi literatur serta saran untuk penelitian selanjutnya.

1.6 Sistematika Penulisan

Dalam penulisan skripsi ini, akan dibagi menjadi beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab pertama ini berisi tentang penjabaran secara sistematis topic yang diambil meliputi latar belakang, tujuan dan manfaat, perumusan dan batasan masalah, metodologi dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab kedua ini menjelaskan dasar teori yang menunjang pembahasan dari penelitian ini. Dasar teori ini berisi tentang Analisis Sentimen, Sosial media, *Natural Language Processing* , dan *Logistic Regression*.

BAB III METODOLOGI PENELITIAN

Pada bab ketiga ini menjelaskan bagaimana penelitian ini dilakukan yang dimulai dari pra-pemrosesan data, ekstraksi fitur pada data, klasifikasi, serta validasi.

BAB IV HASIL DAN ANALISIS

Pada bab keempat ini menjelaskan hasil dan analisa terhadap penelitian yang telah dilakukan.

BAB V KESIMPULAN

Pada bab kelima ini menarik kesimpulan berdasarkan hasil dan analisa terhadap penelitian yang telah dilakukan

Daftar Pustaka

- [1] A. Jayul and E. Irwanto, "Model Pembelajaran Daring Sebagai Alternatif Proses Kegiatan Belajar Pendidikan Jasmani di Tengah Pandemi Covid-19," *J. Pendidik. Kesehat. Rekreasi*, vol. 6, no. 2, pp. 190–199, 2020.
- [2] A. Rane and A. Kumar, "Sentiment Classification System of Twitter Data for US Airline Service Analysis," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 1, pp. 769–773, 2018, doi: 10.1109/COMPSAC.2018.00114.
- [3] A. Tyagi and N. Sharma, "Sentiment Analysis using logistic regression and effective word score heuristic," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 20–23, 2018, doi: 10.14419/ijet.v7i2.24.11991.
- [4] R. Prabowo and M. Thelwall, "Sentiment Analysis : A Combined Approach."
- [5] C. D. Manning, J. Bauer, J. Finkel, and S. J. Bethard, "The Stanford CoreNLP Natural Language Processing Toolkit," *Aclweb.Org*, pp. 55–60, 2014, [Online]. Available: <http://macopolo.cn/mkpl/products.asp>.
- [6] M. Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J. & Nithya, "Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining," vol. 5, no. October 2014, pp. 7–16, 2016.
- [7] S. Scott and S. Matwin, "Feature engineering for text classification," *Mach. Learn. Work.*, vol. 6, pp. 1–13, 1999, [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cn> <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.7033&rep=rep1&type=pdf>.
- [8] K. L. Wagstaff, "Machine learning that matters," *Proc. 29th Int. Conf. Mach. Learn. ICML 2012*, vol. 1, pp. 529–534, 2012.
- [9] C. Bonte and F. Vercauteren, "Privacy-preserving logistic regression trainin," *BMC Med. Genomics*, vol. 11, no. Suppl 4, 2018, doi: 10.1186/s12920-018-0398-y.

- [10] I. Rahman, P. S. Gilmour, L. A. Jimenez, S. K. Biswas, F. Antonicelli, and O. I. Aruoma, "Ergothioneine inhibits oxidative stress- and TNF- α -induced NF- κ B activation and interleukin-8 release in alveolar epithelial cells," *Biochem. Biophys. Res. Commun.*, vol. 302, no. 4, pp. 860–864, 2003, doi: 10.1016/S0006-291X(03)00224-9.
- [11] E. Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, "Edited by Sofia Visa, Atsushi Inoue, and Anca Ralescu," *Maics*, vol. 710, pp. 120–127, 2011.
- [12] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *Processing*, vol., pp. 1–6, 2009.
- [13] Y. Jin, "Development of Word Cloud Generator Software Based on Python," *Procedia Eng.*, vol. 174, pp. 788–792, 2017, doi: 10.1016/j.proeng.2017.01.223.
- [14] S. George K and S. Joseph, "Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature," *IOSR J. Comput. Eng.*, vol. 16, no. 1, pp. 34–38, 2014, doi: 10.9790/0661-16153438.

Program Klasifikasi Sentimen *Logistic Regression*

```
from google.colab import drive
drive.mount('/content/drive')

# Import Library Standard
import numpy as np
import pandas as pd
import re
import matplotlib.pyplot as plt
import seaborn as sns

# Import Library Sklearn
from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV, RandomizedSearchCV
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import RidgeClassifier, LogisticRegression
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, plot_confusion_matrix
from sklearn.pipeline import make_pipeline

# Import Library untuk Stemming
pip install Sastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

pip install wordcloud
from wordcloud import WordCloud

# Read Dataset
data = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/TotalDaring.xlsx')
data.head()

df = data[['Mentions', 'Sentiment']].copy()
df.rename(columns={'Mentions': 'komentar', 'Sentiment': 'label'}, inplace=True)
print('Ukuran Dataset: ', df.shape)
df.head()
```

```

import seaborn as sns
sns.countplot(df['label']);

print("CONTOH KOMENTAR HASIL SCRAPPING\n")
for i in range(1,5):
    print(f"KOMENTAR{i}:\n {df['komentar'][i]}")

def clean_tweet(tweet):
    """ cleaning tweet
    source code: https://www.machinelearningplus.com/python/py
thon-regex-tutorial-examples/
    """
    tweet = re.sub('http\S+\s*', '', tweet) # remove URLs
    tweet = re.sub('pic\.\\S+', '', tweet) # remove picture tw
itter URLs
    tweet = re.sub('RT|cc', '', tweet) # remove RT and cc
    tweet = re.sub('#\S+', '', tweet) # remove hashtags
    tweet = re.sub('@\S+', '', tweet) # remove mentions
    tweet = re.sub('[%s]' % re.escape("""!"#$%&'()*+,-
./:;<=>?@[\\]^_`{|}~"""), '', tweet) # remove punctuations
    tweet = re.sub('\s+', ' ', tweet) # remove extra whitespa
ce
    return tweet

df['komentar'] = df['komentar'].apply(lambda x:clean_tweet(x))

for i in range(9,11):
    print(f"Contoh sebelum cleaning:\n\t{data['Mentions'][i]}")
)
    print(f"Contoh setelah cleaning:\n\t{df['komentar'][i]}\n")
)

negatif_komen = df[df['label'] == 'Negative']['komentar'].valu
es
neutral_komen = df[df['label'] == 'Neutral']['komentar'].value
s
positif_komen = df[df['label'] == 'Positive']['komentar'].valu
es

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import St
opWordRemoverFactory

# Membuat stemmer
stemmerFactory = StemmerFactory()
stemmer = stemmerFactory.create_stemmer()

```

```

# Daftar Stopword
stopWordRemoverFactory = StopWordRemoverFactory()
stopwords = stopWordRemoverFactory.get_stop_words()

def word_cloud(x, stop=stopwords):
    words = WordCloud(width=3000, height=2000, stopwords=stop)
    .generate(str(x))
    plt.figure(figsize=(40,30), edgecolor='k', facecolor='k')
    plt.imshow(words, interpolation='bilinear')
    plt.axis(False)
    plt.tight_layout(pad=0)
    plt.show()

word_cloud(negatif_komen)
word_cloud(neutral_komen)
word_cloud(positif_komen)

df['komentar'] = df['komentar'].apply(lambda x: stemmer.stem(x)
))

for i in range(300, 310):
    print(f"sebelum stemming:\n {data['Mentions'][i]}")
    print(f"setelah stemming:\n {df['komentar'][i]}\n")

df_clean = pd.DataFrame()
df_clean['Komentar_scraping'] = data['Mentions']
df_clean['Komentar_clean'] = df['komentar']
df_clean['label'] = df['label']
df_clean.head(5)

df_clean.to_excel("cleaning_data.xlsx")

# pemisahan feature dan label
X = df['komentar']
y = df['label']

X_train, X_val, y_train, y_val = train_test_split(X, y, test_s
ize=0.2, stratify=y)
print(f"training size {X_train.size}")
print(f"validasi size {X_val.size}")

# encoding data label menjadi numberik karena data awal berupa
teks positif, neutral dan negatif
encoder = LabelEncoder()
y_train_code = encoder.fit_transform(y_train)
y_val_code = encoder.transform(y_val)
print(y_train_code[:10])

```

```

encoder.classes_ #negatif=0, neutral=1, positif=2

sns.countplot(y_train_code)
plt.title('Data Training')

sns.countplot(y_val_code)
plt.title('Data Validasi')

vectorizer = TfidfVectorizer(min_df=2, ngram_range=(1,2), stop
_words=stopwords) #ekstraksi fitur
#min_df = minimal kata minimal muncul 3 kali di dokumen
#stopwords = menghilangkan kata yang sering muncul di bahasa i
ndonesia

#fit dan data tranformasi
bag_of_words = vectorizer.fit_transform(X_train)
X_val_vect = vectorizer.transform(X_val)

feature_names = vectorizer.get_feature_names()
print(feature_names[200:215])
print(feature_names[1000:1015])

#model selection untuk memilih model machine learning mana yan
g memberikan hasil yg baik
models = { 'LogisticRegression': LogisticRegression()}

score={}
for n, model in models.items():
    score[n] = np.mean(cross_val_score(model, bag_of_words, y_
train_code, cv=2, scoring='accuracy'))

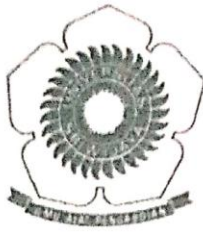
sorted(score.items(), key=lambda x:x[1], reverse=True)

pipe_v1 = make_pipeline(TfidfVectorizer(min_df=2, ngram_range=
(1,2), stop_words=stopwords),
                        LogisticRegression())
pipe_v1.fit(X_train, y_train_code)

print('training report')
plot_confusion_matrix(pipe_v1, X_train, y_train_code)
print(classification_report(y_train_code, pipe_v1.predict(X_tr
ain)))

print('testing report')
plot_confusion_matrix(pipe_v1, X_val, y_val_code)
print(classification_report(y_val_code, pipe_v1.predict(X_val)

```



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI
UNIVERSITAS SRIWIJAYA
FAKULTAS ILMU KOMPUTER
JURUSAN SISTEM KOMPUTER

Jalan Palembang – Prabumulih Km. 32 Indralaya Kabupaten Ogan Ilir Kode Pos 30662
Telepon (0711)7072729, 379249, 581700 Faksimili (0711) 379248, 581710
Pos-el : info@ilkom.unsri.ac.id

FORM PERBAIKAN UJIAN SKRIPSI (TUGAS AKHIR II)

Nama Mahasiswa : Hisyam Sanusi
NIM : 09011181621028.
Jurusan : Sistem Komputer
Hari / Tanggal : Jumat / 30 Juli 2021
Waktu : 15.01 s.d 15.01 WIB
Judul Tugas Akhir : Klasifikasi Sentimen Terhadap Data Text Jejaring Sosial
dengan Topik Pembelajaran Daring Menggunakan
Logistic Regression
Pembimbing : Deris Stiawan, M.T., Ph.D
Huda Ubaya, M.T.
Perbaikan/Saran :

Jangka Waktu Perbaikan :

Telah diperbaiki sesuai dengan saran dan koreksi tim penguji ujian komprehensif.

No.	Nama Penguji	Status Penguji	Tanda Tangan
1.	Deris Stiawan, M.T., Ph.D	Pendamping (Pembela) I	

Palembang, ~~30 Juli 2021~~
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T.
NIP 196612032006041001



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI
UNIVERSITAS SRIWIJAYA
FAKULTAS ILMU KOMPUTER
JURUSAN SISTEM KOMPUTER

Jalan Palembang – Prabumulih Km. 32 Indralaya Kabupaten Ogan Ilir Kode Pos 30662
Telepon (0711)7072729, 379249, 581700 Faksimili (0711) 379248, 581710
Pos-el : info@ilkom.unsri.ac.id

FORM PERBAIKAN UJIAN SKRIPSI (TUGAS AKHIR II)

Nama Mahasiswa : Hisyam Sanusi
NIM :
09011181621028.
Jurusan : Sistem Komputer
Hari / Tanggal : Jumat / 30 Juli 2021
Waktu : 15.01 s.d 15.01 WIB
Judul Tugas Akhir : Klasifikasi Sentimen Terhadap Data Text Jejaring Sosial
dengan Topik Pembelajaran Daring
Menggunakan Logistic Regression
Pembimbing : Deris Stiawan, M.T.,
Ph.DHuda Ubaya, M.T.
Perbaikan/Saran :

Jangka Waktu Perbaikan :

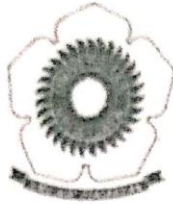
Telah diperbaiki sesuai dengan saran dan koreksi tim penguji ujian komprehensif.

No.	Nama Penguji	Status Penguji	Tanda Tangan
1.	Huda Ubaya, M.T.	Pendamping (Pembela) II	

Palembang, ~~30 Juli 2021~~
Ketua Jurusan Sistem Komputer

30/7/21

Dr. Ir. H. Sukemi, M.T.
NID 105612032006041001



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI
UNIVERSITAS SRIWIJAYA
FAKULTAS ILMU KOMPUTER
JURUSAN SISTEM KOMPUTER

Jalan Palembang – Prabumulih Km. 32 Indralaya Kabupaten Ogan Ilir Kode Pos 30662
Telepon (0711)7072729, 379249, 581700 Faksimili (0711) 379248, 581710
Pos-el : info@ilkom.unsri.ac.id

FORM PERBAIKAN UJIAN SKRIPSI (TUGAS AKHIR II)

Nama Mahasiswa : Hisyam Sanusi
NIM : 09011181621028.
Jurusan : Sistem Komputer
Hari / Tanggal : Jumat / 30 Juli 2021
Waktu : 15.01 s.d 15.01 WIB
Judul Tugas Akhir : Klasifikasi Sentimen Terhadap Data Text Jejaring Sosial dengan Topik Pembelajaran Daring Menggunakan Logistic Regression
Pembimbing : Deris Stiawan, M.T., Ph.D
Huda Ubaya, M.T.
Perbaikan/Saran :

Jangka Waktu Perbaikan :

Telah diperbaiki sesuai dengan saran dan koreksi tim penguji ujian komprehensif.

No.	Nama Penguji	Status Penguji	Tanda Tangan
1.	Ahmad Heryanto, M.T	Penguji	

Palembang, ~~30 Juli 2021~~
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T.
NIP 196612032006041001

Klasifikasi Sentimen Terhadap Data Text Jejaring Sosial Dengan Topik Pembelajaran Daring Menggunakan Logistic Regression

ORIGINALITY REPORT

8% SIMILARITY INDEX	5% INTERNET SOURCES	1% PUBLICATIONS	5% STUDENT PAPERS
-------------------------------	-------------------------------	---------------------------	-----------------------------

PRIMARY SOURCES

1	Submitted to Sriwijaya University Student Paper	4%
2	pt.scribd.com Internet Source	1%
3	id.scribd.com Internet Source	1%
4	core.ac.uk Internet Source	1%
5	journal.ipb.ac.id Internet Source	1%
6	repository.dinamika.ac.id Internet Source	1%

Exclude quotes On

Exclude bibliography On

Exclude matches < 1%